Will you Survive the Drive?

UK Road Safety Data 1979 - 2016

COMP 30780 Data Science in Practice

Ciara O' Reilly (15496208), Orlaith Joyce (15386636)
School of Computer Science
University College Dublin

May 11, 2018

Abstract.
This research project involves the analysis of UK Road Safety data over 37 years. Our research into this area will question how factors such as location, time, and purpose of journey contribute to road accidents. Road Safety affects each of us, whether you drive, take public transport, cycle or even walk. With a large volume of road users around the world, it is essential that we understand all attributes and causes of road accidents. We will also analyse unique road accident attributes which have been overlooked in past research. These include factors such as weather, driver IMB decile and purpose of journey. The overall aim of this project is to present the major causes of road accidents, to promote safer road usage and to minimise the number of fatal outcomes.

Declaration. We (Ciara O' Reilly, 15496208 and Orlaith Joyce, 15386636) declare that this assignment is our own work and that we have correctly acknowledged the work of others. This assignment is in accordance with University and School guidance1 on good academic conduct in this regard.

[1] See https://www.cs.ucd.ie/sites/default/files/cs-plagiarism-policy_august2017.pdf

1. Introduction

Within this project we will analyse road safety data in the UK from 1979 to 2016. Initially we chose this topic as we were interested in the car insurance market, particularly the quotes which are generated for drivers. We observed that insurance companies were quoting significantly varied prices for individuals, which was based on previous road safety research. It became apparent that there was a gap in the findings which were being referenced by these insurance companies. Upon closer inspection into the area of road accidents, we questioned how factors such as road accident location, time, and purpose of journey have been overlooked in past research. We are particularly interested in discovering the key factors which contribute most to road accidents. Our main focus is to promote safer road usage, not only in the UK but in Ireland also.

The population of the UK is currently 66.5 million, of which 79% drive. This highlights the importance of road safety as a research topic. Each year, there are around 200,000 road accidents, of which 26,000 results in serious injuries and fatalities (Department for Transport UK, (2017)). Although most people use the roads every day in some capacity, many of these become complacent of just how dangerous our roads can be. With such a large volume of road users, it is essential to understand and share any knowledge of road safety that we find. We placed a significant importance on obtaining UK road safety data over a long period of time. We collected a large data set with information covering 37 years of road safety accident reports. Details contained in this dataset include attributes such as location, vehicle type, age band of driver, accident severity and weather conditions for each recorded road accident. Working with such a large dataset enables us to obtain more accurate results, while covering a wide range of road accident features.

As road safety is a topic of international relevance, there are constant efforts made to introduce up-to-date road safety regulations for all capacities of road users. These include penalty point systems, stricter rules for learner drivers and improved road infrastructure. In recent years we have also seen massive improvements put into the safety features and components of modern vehicles. These include automatic emergency breaking, sensors, GPS systems, forward collision warning, reversing cameras, lane-keeping assist and tyre pressure monitors. However, with a current high number of annual road accidents, we question how effective these modern technologies and regulations really are.

Existing research on road safety data is widely available. However, we found that previous research is repeatedly focused on factors such as the gender of drivers (McKenna, F. P., Waylen, A. E., & Burkes, M. E. (1998)). Our research project focuses on previously unexplored factors of road accidents such as vehicle type, time, location and purpose of journey. Our aim is to identify and analyse these unique factors and understand how they may contribute to road accidents. This approach revealed many interesting results. For example, we discovered that location and time of year does impact road accidents.

The remainder of this report is organised as follows. In the next section we will discuss our motivations and objectives, to give a more detailed account of our project's background and outline our main research questions. We will then move on to discuss the Data Wrangling process of our project, outlining our datasets and how we both obtained and prepared them. Next, we will discuss our research questions in greater detail, outlining our analysis and results. This will be followed by a discussion piece, focusing on the ethical and reproducibility considerations we undertook for this project, and finishing with any limitations that we faced. We will then move on to draw up our final conclusions, and discuss any future work which could potentially be carried out on road safety data. Our report will finish with a bibliography and appendix of extra results which we believe are not to be missed.

2. Motivations & Objectives

A further discussion into the background of our project and our main research questions/objectives.

2.1. Background & Motivations

We chose the topic of road safety upon discovering the methods used by insurance companies to quote individual drivers. We were surprised to discover that quotes were calculated based on factors such as gender, age, vehicle age and occupation of driver. These calculations are based on previous road safety research. As a result, we were interested in taking a further look at this existing research.

Research on the topic of road safety is widely available, due to its international relevance. We found existing research papers on the topic of novice drivers (McKnight, A. J., & McKnight, A. S. (2003)), mobile phone use (Hislop, D. (2012)), speeding definitions for male and female drivers (Lewis, I. M., Watson, B. C., White, K. M., Elliott, B., Thompson, J., & Cockfield, S. (2012)), trends in fatal car accidents (Broughton, J., & Walter, L. (2007)) and casualty rates by type of car in road accidents (Broughton, J. (2008)). Upon analysing these research papers, we found a gap in existing findings for factors such as location, IMD Decile and vehicle types for road accidents. Existing research on road safety within Ireland is extremely limited, and as such we were unable to obtain datasets for Irish road accidents. This resulted in our decision to use UK road safety data as it was openly available and we believe the socio-economic similarities between the UK and Ireland would yield results indicative of Irish road safety data.

Our main objective is to identify the most influential factors in road accidents. Once we have identified these, we hope to raise awareness of the importance of road safety both in Ireland and the UK. We will present these findings in a way which can be easily read and comprehended by all, irrespective of an individual's prior knowledge on the topic of road safety.

Most importantly, we aim to encourage the use of road safety analysis within Ireland. This would require publication of raw data for road accidents by Irish governing bodies such as the Road Safety Authority. Particularly important features of Irish road safety data which should be publicly available include the location, vehicle type, age of driver, weather conditions and severity of road accidents. We hope that our research on UK road safety data will highlight the need for open research opportunities in Ireland.

2.2. Research Questions

Our project consists of three main research questions, each broken down into subsections as follows.

2.2.1. RQ1: Is There a Safest Time for Travel?

For research question one, we will analyse whether different time periods (e.g. season, day) play a role in road accidents and accident severity. We have broken this research question down into sub-questions as follows.

A) Is it Safer to Drive in the Summer than in the Winter?
B) Is it Safer to Drive During the Day-time than Night-time?
C) Are Occupational Drivers More Dangerous than Non-Occupational Drivers?

2.2.2. RQ2: What Vehicles are Involved in Most Road Accidents?

For research question two, we will analyse all vehicle types involved in road accidents, including the age of vehicles at the time of the accident. We aim to explore whether/how these factors impact road accidents. We have broken this research question down into sub-questions as follows.

A) Are Newer Vehicles Safer than Older Vehicles?
B) Is it Safer to Travel by Car than by Motorcycle?

2.2.3. RQ3: Does Location Impact Road Safety?

For our third research question, we will analyse certain locational aspects of recorded road accidents. These include urban/rural classification of accident location, and the 'IMD Decile' of drivers. We have broken this question down into the following sub-questions.

A) Is it Safer to Drive in Urban or Rural Areas?
B) Are Drivers from 'most Deprived' IMD Decile Regions Involved in the Most Road Accidents?

3. Data Wrangling

Further insight into each of the key aspects of getting and preparing our data.

3.1. Data Acquisition

We obtained our data from the UK government website (Department for Transport UK, (2017)). Each of our datasets are open-sourced and available to download as zip/csv files. For our project, we obtained 8 datasets, in total spanning the years 1979 – 2016. Of these 8 datasets, 4 contained information specific to road accidents which occurred in the UK for certain time periods. This information included Latitude, Longitude, Date, Number of Casualties, Accident Severity and Time. The remaining 4 datasets contained information specific to vehicles involved in UK road accidents for certain time periods. This information included Vehicle Type, Journey Purpose of Driver, Gender of Driver, Age of Vehicle, IMD Decile of Driver.

3.2. Data Cleaning & Preparation

Once we obtained our eight datasets, we merged each 'Accident' dataset with its corresponding 'Vehicle' dataset. To ensure that our resulting merged datasets contained matching rows of data, we merged them based on our 'Accident Index' column. Each recorded accident is given its own unique accident index, which can be used to retrieve accident specific data. This resulted in four datasets which summed to over 13 million rows of data, with each row representing a recorded UK road accident. We then used our knowledge of MapReduce scalability to divide our four large datasets into smaller datasets of roughly 3 million rows. This allowed us to easily process our data and to run our analysis notebooks in parallel.

At this stage in our preparation, our datasets were filled with integer values. Each of these integer values corresponded with a string value within our 'Road Accident Safety Data Guide'. We decided to use the pandas mapping method to assign each integer value within our datasets to its corresponding string value within our guide. This allowed us to better analyse our data and focus on string values e.g. car, to obtain our results.

We then defined relevant, non-null columns for each of our research questions, and saved our data into multiple question-specific datasets. For example, to answer research question 1 (a) regarding seasonal data, we required our 'date' column, and did not require our 'vehicle type' column. This produced smaller datasets for us to process easily, which improved our efficiency and accuracy. It was then important to identify and remove null values within each of these datasets. Our data included multiple representations of null values. They are as follows; '-1' represents data which is missing or out of range, 'Unknown' and 'Other/Not known (2005-10)' representing values which were not recorded at the time of the accident. We decided to replace each of these values with NaN values. Unfortunately, we found that columns within some of our datasets contained all NaN values. These columns included Longitude and Latitude information, which is essential in answering research question three. We decided to remove all NaN values from columns within our datasets which had less than 30% NaN values, and dropped columns with entirely null values or over 30% NaN values. Once this preparation was complete, our individual datasets summed to over 12 million rows, however each of our research question specific datasets contain road accident records for differing time periods. This was a direct result of our previous issue with columns containing all null values for certain year groups. This meant we were unable to use those year groups to answer some of our questions.

Once this preparation was complete, we used our date column to obtain the year in which accidents occurred. We then sorted our datasets in ascending order based on the year they occurred. This was done to analyse the range of years being used to answer each individual research question.

4. Data Analysis & Results

Further insight into our methods of data analysis and our key results.

4.1. RQ1 - Is There a Safest Time for Travel?

- **(A)** Is it safer to drive in the Summer than in the Winter?
- **(B)** Is it Safer to Drive During the Day-time than Night-time?
- **(C)** Are Occupational Drivers More Dangerous than Non-Occupational Drivers?

4.1.1. Datasets

Please note, as mentioned in section 3.2, question one parts (a) and (b), will have multiple input files. Each of these files contains data of UK road accident records over the years 1979 – 2016. As such, all analysis is carried out on each individual dataset, and our results are aggregated.

(A)

Our datasets for research question one (A) included UK road accident data from 1979 – 2016. To answer this question, we required information on the season in which each road accident occurred. This was obtained by adding a new column to our datasets, filled using data from our 'month' column. Additional data which was relevant to this question was data describing the accident severity, weather and road conditions for all recorded road accidents. This data was available in our datasets.

(B)

Our datasets for research question one (B) included UK road accident data from 1979 – 2016. To answer this question, we required data relevant to the time of day that each accident occurred. As a result, we added a new column to our datasets which stated whether an accident occurred during the day-time or night-time, which we filled using our 'time' column. Additional data which was required for answering this question included the accident severity of each accident, and the day of the week on which each road accident occurred. This data was available in our datasets.

(C)

Our datasets for research question one (C) included UK road accident data from 2005 – 2014. To answer this question, we required specific information on the journey purpose of drivers involved in all recorded road accidents. As a result, we added a new column, 'occupational/non', to our datasets which listed whether the journey purpose of drivers involved in each accident was occupational or non-occupational. This column was filled using the values from our 'journey purpose of driver' column, where journey purposes which were listed as anything other than 'occupational', were assigned the value 'non-occupational' in our new column. Additionally, we required data on the age band and gender of drivers for each road accident. This data was available in our datasets.

4.1.2. Approach

(A)

To answer research question one (A), we first grouped all road accidents by our 'season' column. We then counted the total number of accidents which were recorded for each season, to analyse which season had the most/least amount of recorded road accidents from 1979 – 2016. Once this was complete, we then calculated the total number of slight, serious and fatal road accidents which were recorded for each season. This was done using our 'accident severity' column. We then normalised this resulting data to get a percentage value of fatal, serious and slight accidents for each season. This allowed us to analyse whether any one season had more/less severe accident outcomes. We then took a closer look at the weather and road conditions (See Appendix Fig 29) which were recorded for each accident. We then counted the total number of accidents which occurred for each weather and road condition, and grouped them by the season in which they occurred. Finally, we then normalised these results based on all accidents within each season. This allowed us to analyse how road and weather conditions may contribute to road accidents, and in turn may affect road safety.

(B)

To analyse our data for research question one (B), we first grouped all road accidents by our 'day_night' column. We then counted the total number of accidents which were recorded as having occurred in the day time and the night time. We then normalised this data to get the percentage of road accidents for both day and night time. This enabled us to analyse whether more accidents occurred in the day time of the night time. Once this was complete, we grouped all road accidents by our 'accident severity' column. We then counted the total number of accidents which had fatal, serious and slight outcomes and grouped this result again by our 'day_night' column. We normalised this data to obtain the breakdown of accident severities during both the day and night, which allowed us to analyse whether accidents which occurred at night had more/less severe outcomes than those that occurred during the day. Finally, we grouped accidents by our 'day of week' column, and counted the total number of accidents which occurred for each day of the week. We then grouped this data by our 'day_night' column and normalised our results. This gave us an insight into which day of the week had the highest/lowest number of day/night time accidents from 1979 – 2016.

(C)

To answer research question one (C), we began by grouping road accidents by our ‘occupational/non’ column. We totalled these grouped results to analyse whether there were more accidents which listed occupational journey purpose, compared to non-occupational. Once this was complete, we then grouped all road accidents by our ‘sex_of_driver’ column. We then counted the number of accidents recorded for both sexes, and grouped our results by our ‘occupational/non’ column. This allowed us to analyse whether there were more male or female drivers involved in road accidents, who’s purpose of journey was occupational or non-occupational. This would also allow us to analyse the ratio of male to female drivers within our dataset. Finally, we calculated the total number of road accidents recorded and grouped these by our ‘age_band_of_driver’ column. We then grouped these results again by our ‘occupational/non’ column. We normalised this result to account for the uneven distribution of drivers by age (e.g. due to the legal age of driving in the UK being over 17). This enabled us to analyse whether certain age groups have a higher number of recorded accidents, where journey purpose is listed as occupational. Results

4.1.3. Results

(A)

The highest number of road accidents occurred in Autumn, with Summer in second place. The least amount of road accidents occurred in Winter (Fig 1)

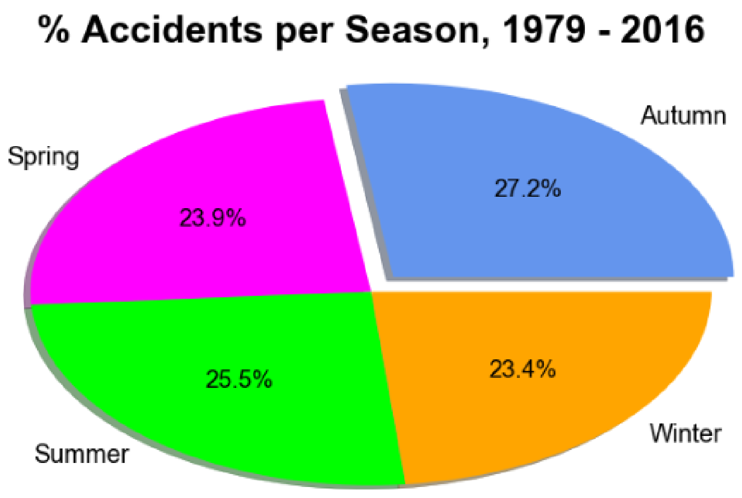

Figure 1, A graph showing % of road accidents which occurred in each season from 1979 - 2016

The highest number of both fatal and slight accidents occurred in Winter, while the highest number of serious accidents occured in Summer. (Fig 2)

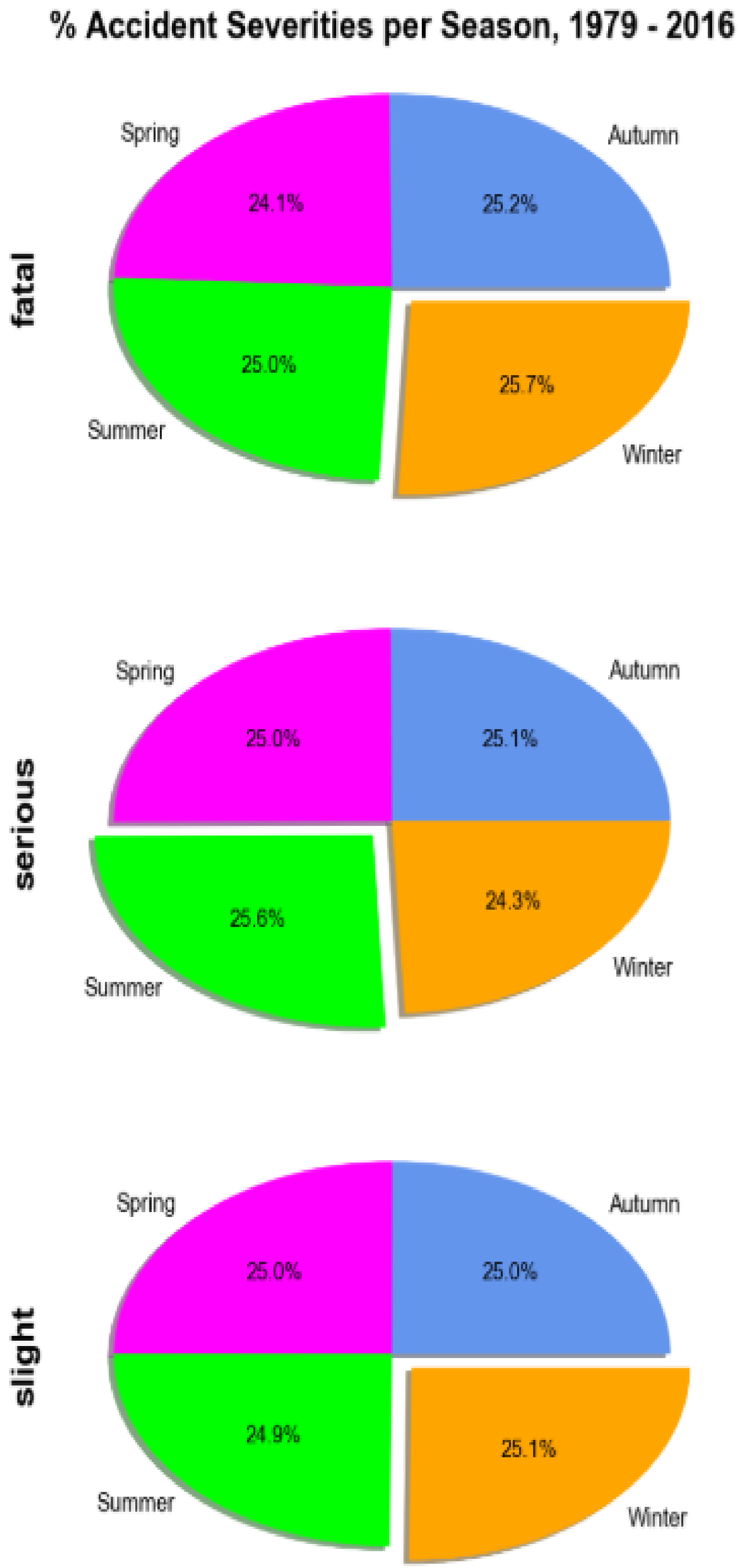

Figure 2, showing % of fatal, serious and slight accidents for each season

Throughout all seasons, most road accidents occurred during ‘fine no high winds’ weather, while the second highest number of accidents occurred in ‘raining no high winds’weather conditions. (Fig 3)

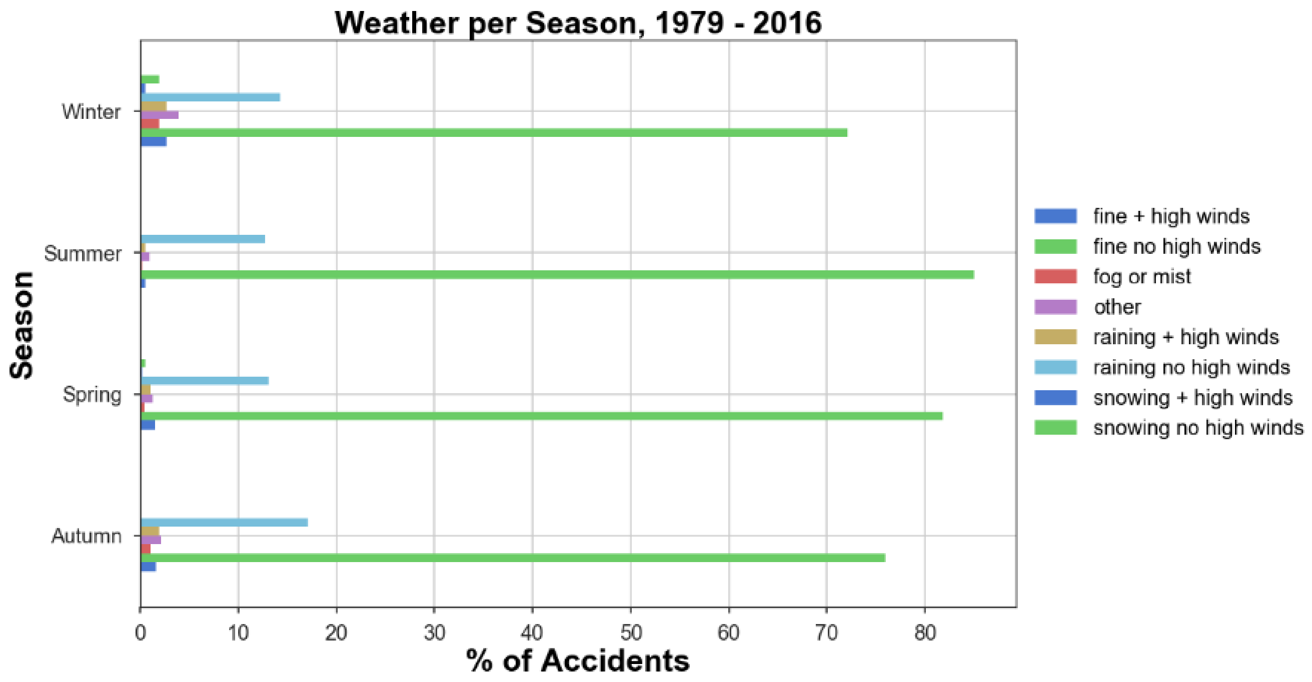

Figure 3, showing the weather conditions at the time of accidents for all season

(B)

There were more road accidents during the day time (Fig 4)

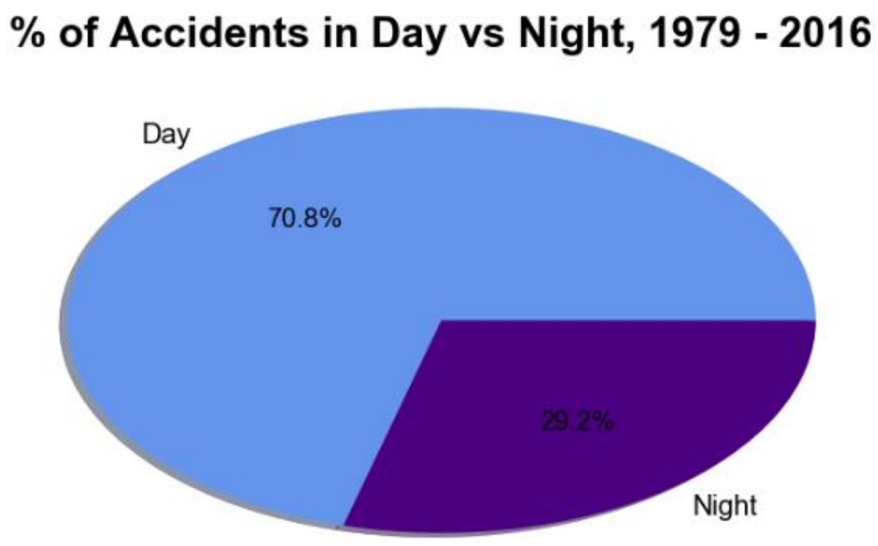

Figure 4, Showing the % of Road Accidents which occurred at Day time and Night time

Accidents which occurred during the night time were of higher fatal and serious severity. Accidents which occurred during the daytime had mostly slight severity. (Fig 5).

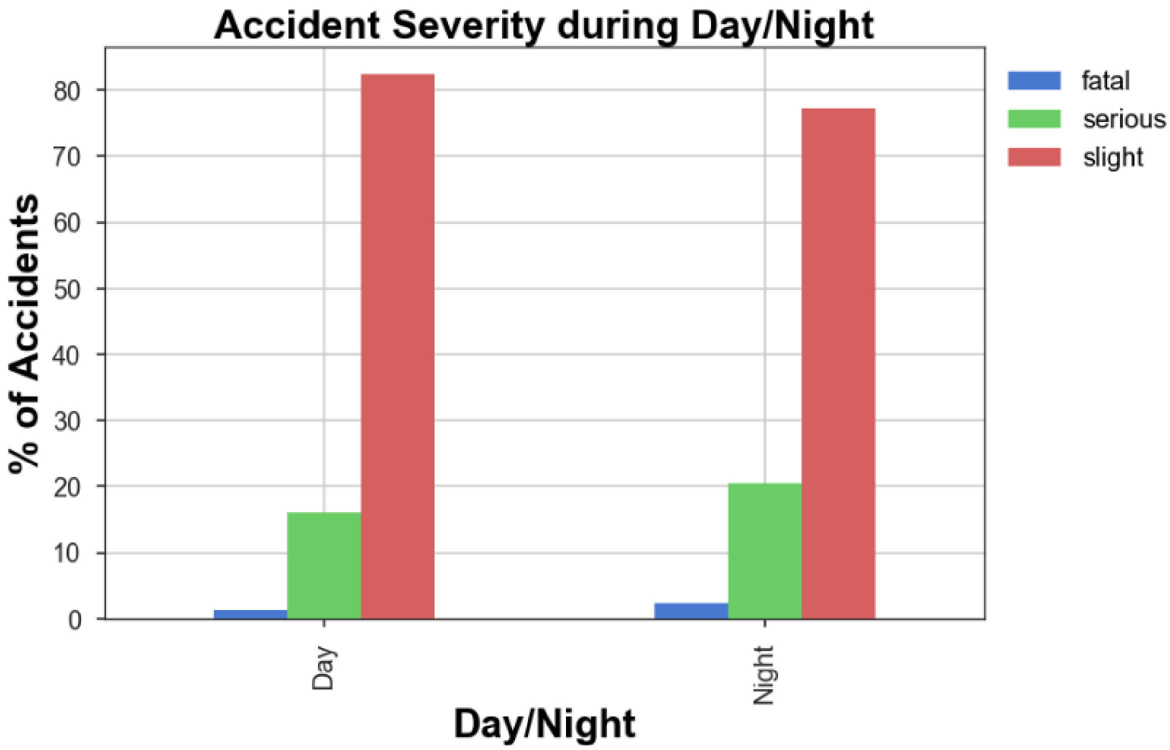

Figure 5, Showing total % of accidents during day/night which were fatal/serious/slight.

The highest number of both night-time and day-time accidents occurred on a Saturday. (Fig 6)

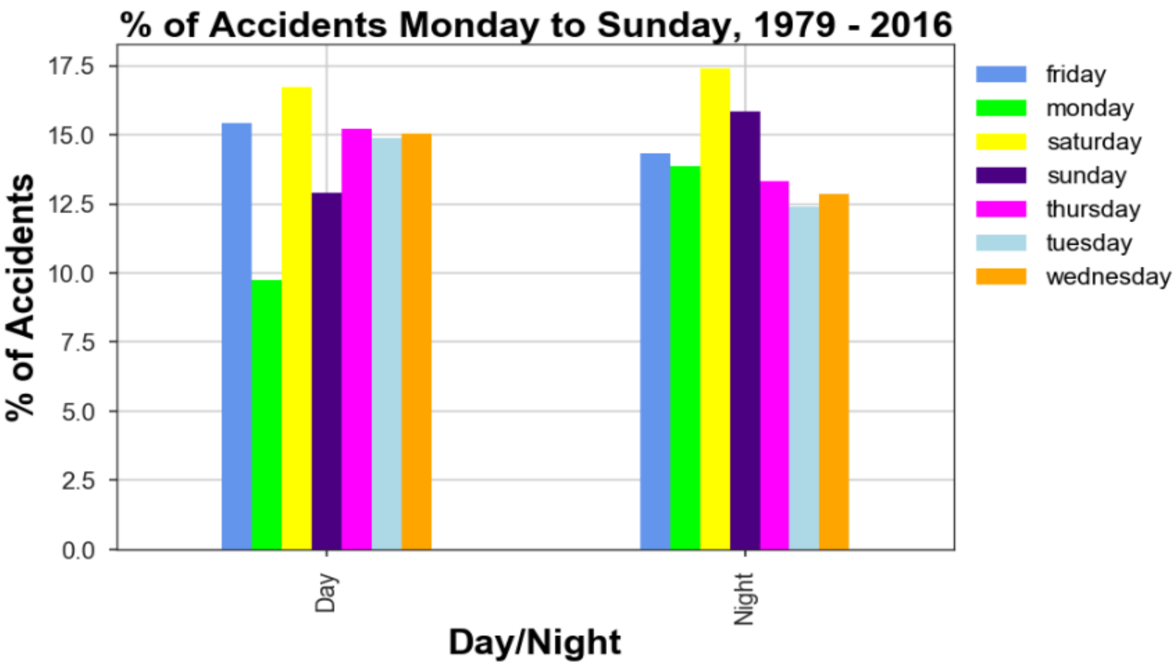

Figure 6, showing total number of accidents which occurred on each day of the week.

(C)

There were a higher number of male drivers than female drivers, for accidents in which journey purpose of driver was both occupational and non-occupational. (Fig 7)

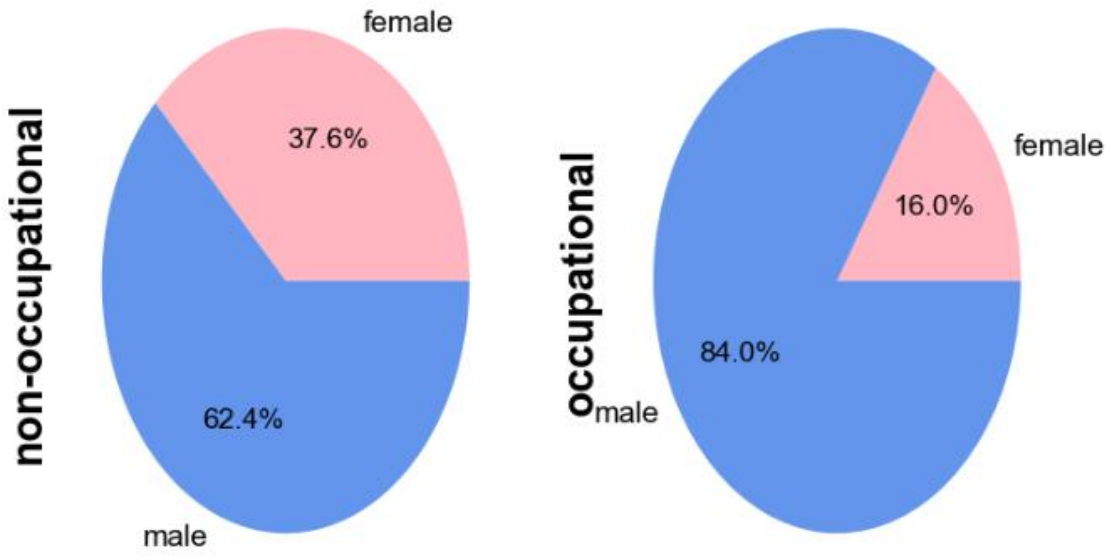

Figure 7, showing total % of male and female occupational/non-occupational drivers involved in road accidents

Road accidents which involved drivers who's journey purpose was occupational, were at their highest for drivers aged between 26 and 75 years old. (Fig 8)

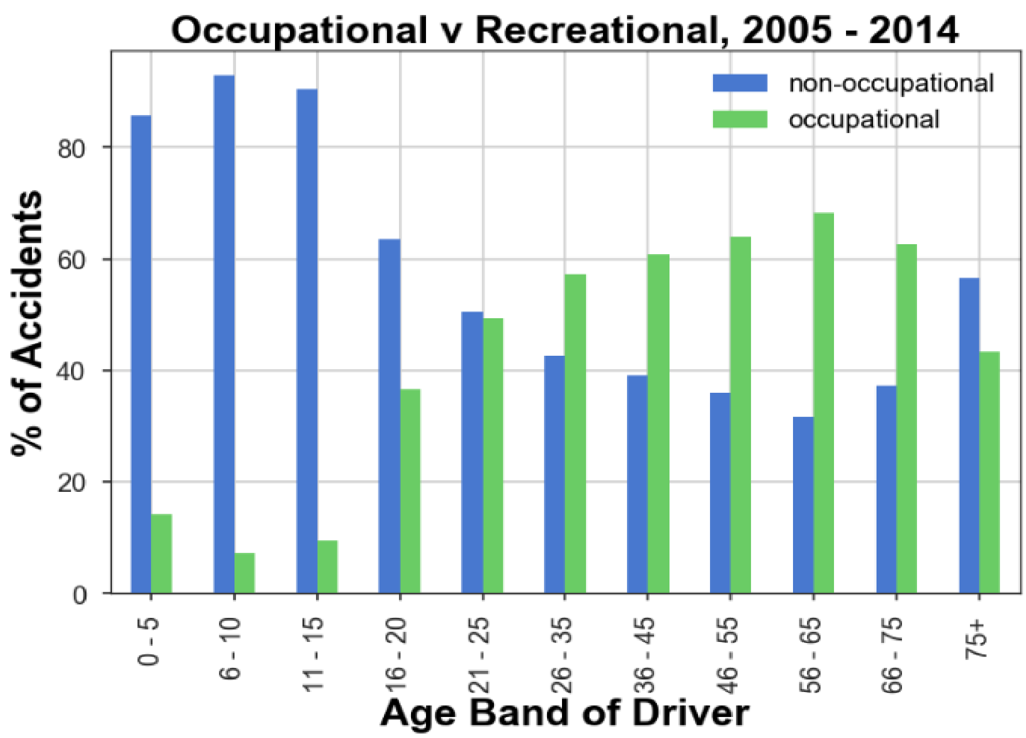

Figure 8, showing % of occupational/non-occupational drivers for each age group, involved in road accidents

4.1.4. Discussion

(A)

For research question one (A), we found that although most accidents between 1979 to 2016 occurred in Autumn, the accident severity of road accidents within each season gives us a more accurate indication of which season is most dangerous for divers. From Fig 1 and Fig 2, we found that Winter is the most dangerous season to drive. This is due to Winter having the highest percentage of fatal accidents, when compared to all other seasons. However, we would like to acknowledge that further work could be carried out to examine the number of wet/snowy/fine weather days which occurred in the UK for each season/year, to further normalize our results, and gain a more accurate prediction of road safety within each season. We were unable to obtain this data.

(B)

For research question one (B), we found that most road accidents occurred during the day time rather than night time. Fig 5 shows us however that night time accidents had a higher percentage of fatal and serious accident severities, and fewer slightly severe accidents when compared to accidents which occurred during the day time. This indicates perhaps that although more accidents occur during the day, they are likely to result in only slight injury. Fig 6 also shows us that the total number of day and night time road accidents is at its highest on Saturdays. There is potential for further work to be carried out on these results, whereby data could be normalised based on the total number of drivers on the road for each day of the week, and at both day and night time. This was data which we were unable to obtain.

(C)

For research question one (C), we found that overall, there are more male than female drivers involved in road accidents between 2005 - 2014 for both occupational journey purposes and non-occupational. This result was to be expected as there are a higher number of male drivers than female drivers in England (Department for transport, England (2014)). Fig 8 shows that occupational journey purposes were more dominant in road accidents for drivers aged between 26 – 75. This is perhaps due to the legal age of driving in the UK set as 17 years old. Further analysis could be done on this result to normalise based on the number of people both on the roads, and whom drive for work e.g. Delivery drivers. Unfortunately, we were unable to obtain this data.

4.2. RQ2 – What vehicles are involved in the most road accidents?

(A) Are newer vehicles safer than older vehicles?
(B) Is it safer to travel by car than by motorcycle?

4.2.1. Datasets

(A)

Our data for part A of our second research question uses UK road safety data for the years 1993-2016. To be able to answer this question we also used the data provided on the age of the vehicle at the time of the accident to compare old and new vehicles. For our comparisons we will use data on the number of casualties that resulted in each recorded accident. The accident severity data tells us whether the severity of the accident was fatal, serious or slight. This was an important piece of information which we used in multiple sections of our analysis. For example, this allowed us to explore the severity of accidents for old and new vehicles as well as determining the influence of weather conditions at the time of the accident and modern road safety features. Each row in the 'Accident_Index' column represents an accident which was very important to count the number of accidents to compare old and new vehicles.

(B)

Our data for part B of our second research question uses UK road safety data for the years 1993-2016. Using information on the type of vehicle involved in each accident we will use the following data to compare which is the safest. Each row in the 'Accident_Index' column represents an accident which was very important to count the number of accidents to compare vehicle types. Data provided on the accident severity data tells us whether the severity of the accident was fatal, serious or slight. This was an important piece of information which we used in multiple sections of our analysis. For example, this allowed us to explore the severity of accidents for each vehicle type as well as look at which vehicle type is safer in rainy weather conditions at the time of the accident.

4.2.2. Approach

(A)

Q: "Are older vehicles in more accidents?"
Since we are focusing on the age of vehicle here we grouped the accidents based on the age of the vehicle at the time of the accident and looked at the count of these accidents between 1993 and 2016 using the accident index. However, after getting surprising we broke the data into two timeframes, 1993-2004 and 2005-2016, to be able to have a closer look at the data. This was repeated for each of the three data frames and the results were aggregated accordingly.

Q: "Are there more casualties in older vehicles?"
Again, for each of the three data frames the data was grouped by the age of vehicle at the time of the accident and the average number of casualties in an accident for each age was calculated. The results were then aggregated by summing the means for each vehicle age in each of the three data frames and dividing them by 3 to get the "mean of means" for each giving us the overall mean.

Q: "Are accidents more severe in older vehicles?"
To analyse the severity of accidents for each vehicle age the data was grouped by the age of the vehicle at the time of the accident and the accident severity. The count of the total number of accidents for each group was then calculated using the accident index. This was repeated for each of the three data frames and the results were aggregated accordingly. Following this the data was normalised to get a further breakdown of the accident severity for each vehicle age.

Q: "Are newer vehicles safer in rainy weather conditions?"
To compare old and new vehicles in rainy weather conditions the data was grouped by the age of the vehicle and the data provided on the weather conditions at the time of the accident. The value_counts was then calculated for the accident severity in each resulting group to return the number of slight, serious and fatal accidents for each. This was repeated for each of the three data frames and the results were aggregated accordingly. Following this the data was normalised to get a further breakdown of the accident severity for each vehicle age. Using these results, all the data related to "rainy" weather conditions was extracted.

Q: "How did the introduction of modern technologies influence road safety?"
To see if modern technologies influenced road safety we decided to look at the number of accidents in each year by grouping the data y the year and get the total count of accidents using the accident index. We decided to look specifically at airbags and AEB brakes. Using information on when they became compulsory in vehicles by law/were introduced we looked at the four years before and after to see how it influenced the number of accidents, if at all.

(B)

Q: "Which type of vehicle has the highest number of accidents?"
Since we are focusing on the type of vehicle here we grouped the accidents based on the vehicle type involved in the accident and looked at the count of these accidents between 1993 and 2016 using the accident index. This was repeated for each of the three data frames and the results were aggregated accordingly. Following this the data was normalised by the total number of accidents to be able to make comparisons.

Q: "On average, which vehicle type has the most casualties?"
Again, for each of the three data frames the data was grouped by the type of vehicle involved in the accident and the average number of casualties in an accident for each vehicle type was calculated. The results were then aggregated by summing the means for each vehicle age in each of the three data frames and dividing them by 3 to get the "mean of means" for each giving us the overall mean.

Q: "Which type of vehicle is involved in the most fatally severe accidents?"
To analyse the severity of accidents for each vehicle type the data was grouped by the type of vehicle involved in the accident and the accident severity. The count of the total number of accidents for each group was then calculated using the accident index. This was repeated for each of the three data frames and the results were aggregated accordingly. Following this the data was normalised to get a further breakdown of the accident severity for each vehicle type.

Q: "Is one vehicle type safer than another in rainy weather conditions?"
To compare the different types of vehicles in rainy weather conditions the data was grouped by the vehicle type and the data provided on the weather conditions at the time of the accident. The value_counts was then calculated for the accident severity in each resulting group to return the number of slight, serious and fatal accidents for each. This was repeated for each of the three data frames and the results were aggregated accordingly. Following this the data was normalised to get a further breakdown of the accident severity for each vehicle type. Using this all the data related to "rainy" weather conditions was extracted. The results were then normalised again to further analyse the severity in each rainy weather condition specified.

4.2.3. Results

(A)

Our data shows that newer vehicles are involved in more accidents (Fig 9). However, we were very surprised by this result and therefore decided to look at the number of accident for new and old vehicles between 1993-2004 (Fig 10) and 2005-2016 (Fig 11). While the highest number of accidents was 200,000 fewer in the later time period they both still followed the same trend with new vehicles having more accidents.

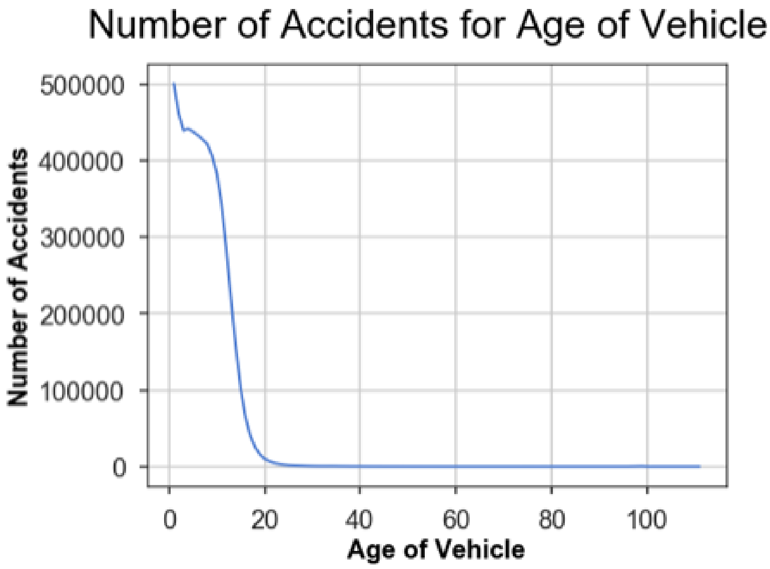

Figure 9, showing the total number of accidents for each vehicle age involved in accidents between 1993 and 2016.

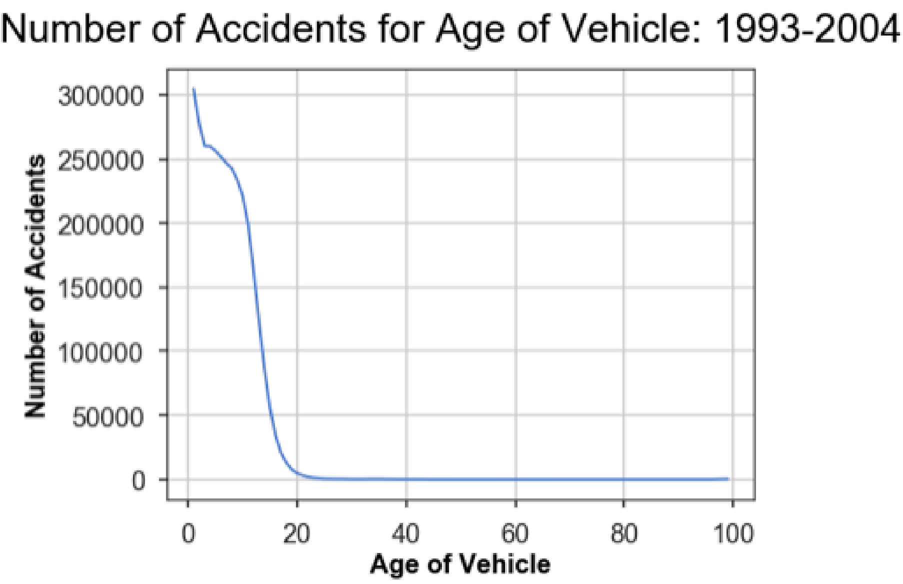

Figure 10, showing the total number of accidents for each vehicle age involved in accidents between 1993 and 2004.

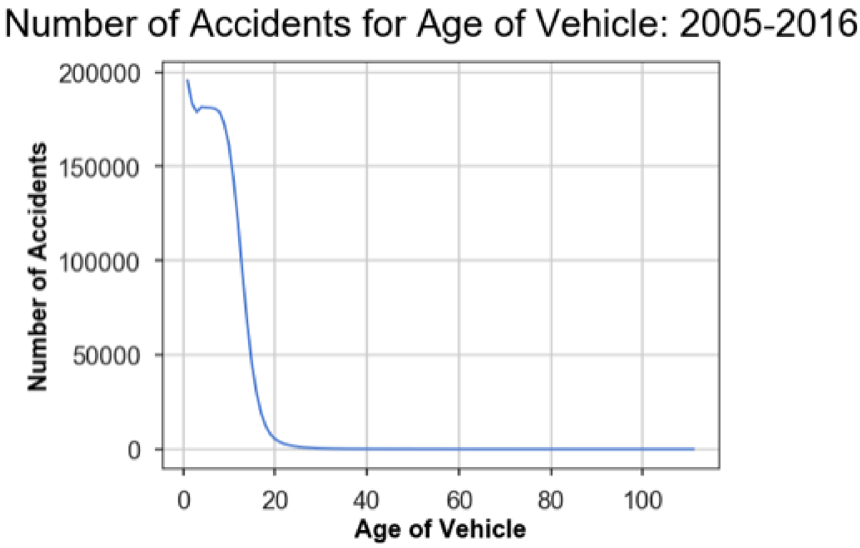

Figure 11, showing the total number of accidents for each vehicle age involved in accidents between 2005 and 2016.

On average, there are more casualties in accidents in older vehicles but there was still a higher average than expected in newer vehicles. (Fig 12)

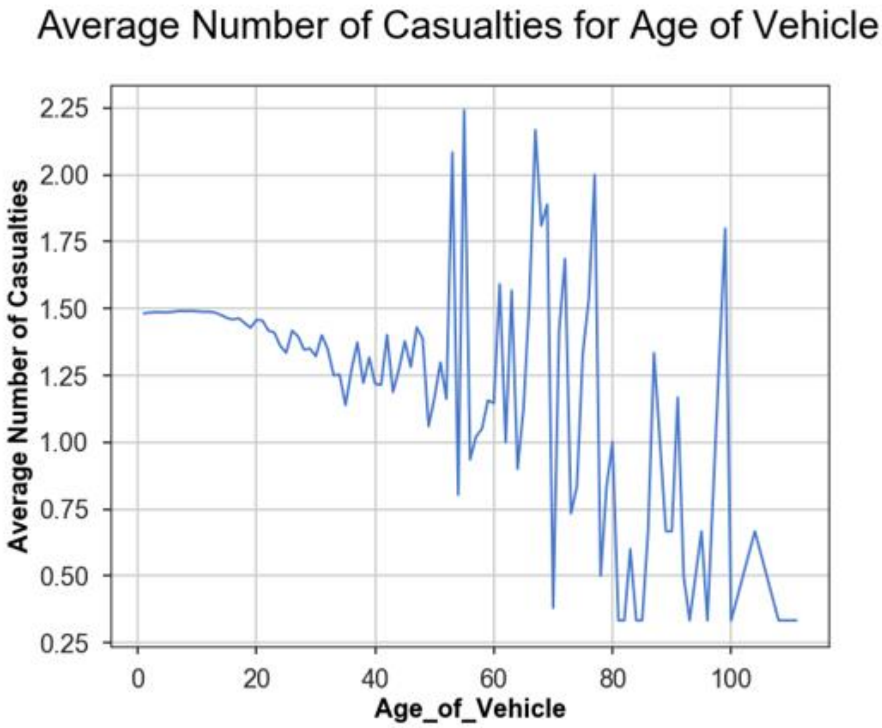

Figure 12, showing the average number of casualties in accidents for each vehicle age involved in accidents between 1993 and 2016.

Older vehicles have a lot more fatal and serious accidents than newer vehicles. However, both old and new have a high % of slight accidents. (Fig 13)

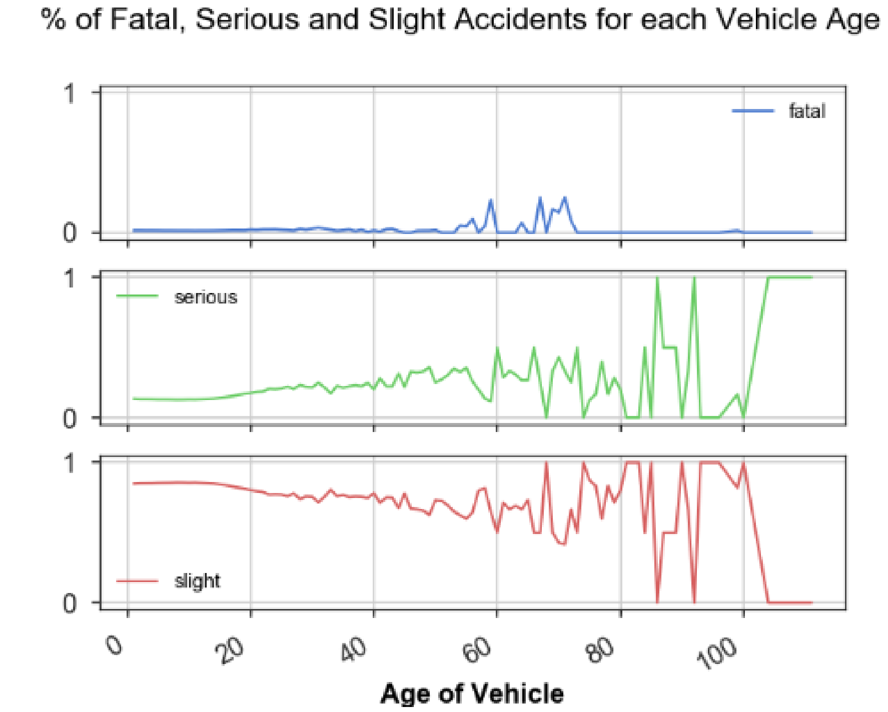

Figure 13, showing the % of fatal, serious and slight accidents for each vehicle age involved in accidents between 1993 and 2016.

In rainy weather conditions older vehicles have a lot more fatal and serious accidents than newer vehicles. However, both old and new have a high % of slight accidents. (Fig 14)

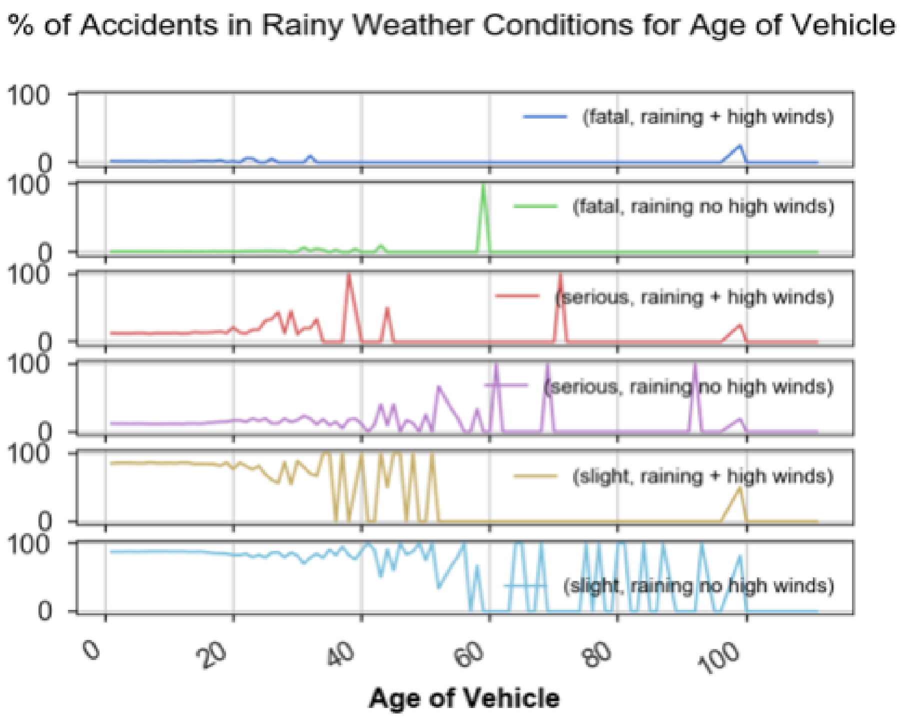

Figure 14, showing the % of fatal, serious and slight accidents for each vehicle age involved in accidents in rainy weather conditionsbetween 1993 and 2016.

The number of road accidents overall in the UK declines after 2004. (Fig 15)

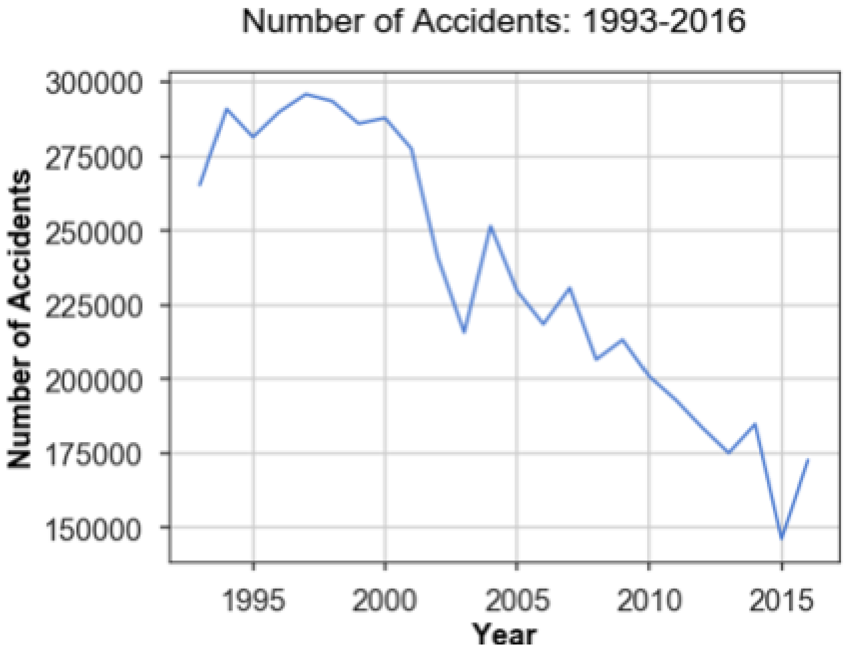

Figure 15, showing the total number of accidents that occurred between 1993 and 2016.

After airbags in 2004 there is hardly any decline in the number of fatal accidents and an extremely slight decline in the number of serious accidents. However, there is a noticeable decline in the number of slight accidents. (Fig 16)

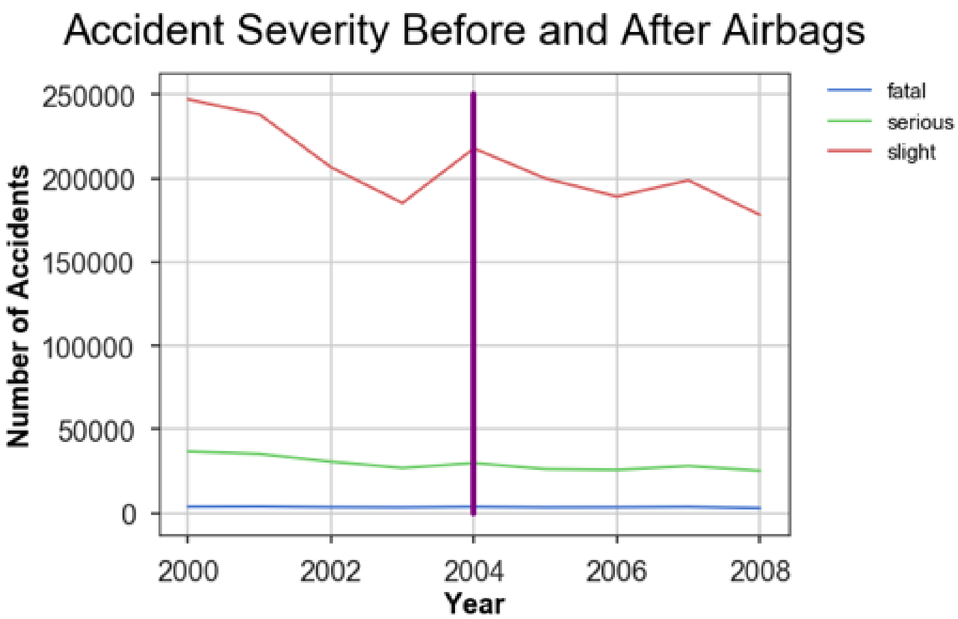

Figure 16, comparing the number of fatal, serious and slight accidents that occurred before and after airbags became a compulsory feature in the EU un 2004.

After AEB brakes in 2010 there is a decline in the number of slight accidents but nearly no change in the number of fatal and serious accidents. (Fig 17)

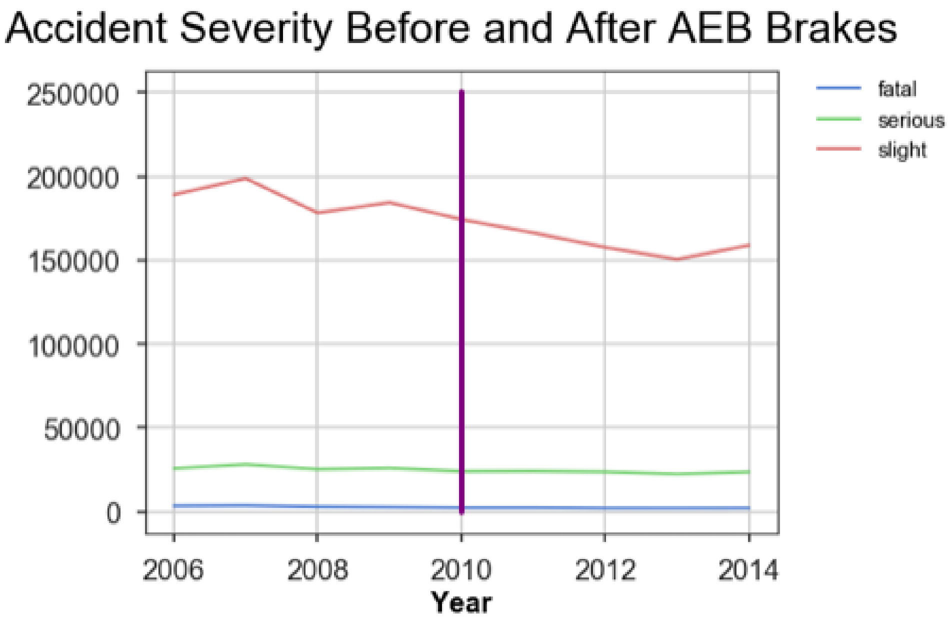

Figure 17, comparing the number of fatal, serious and slight accidents that occurred before and after AEB brakes were introduced in 2010.

(B)

Cars have a significantly higher % of accidents in comparison to the other vehicle types with pedal cycle and motorcycle being the second highest. (Fig 18)

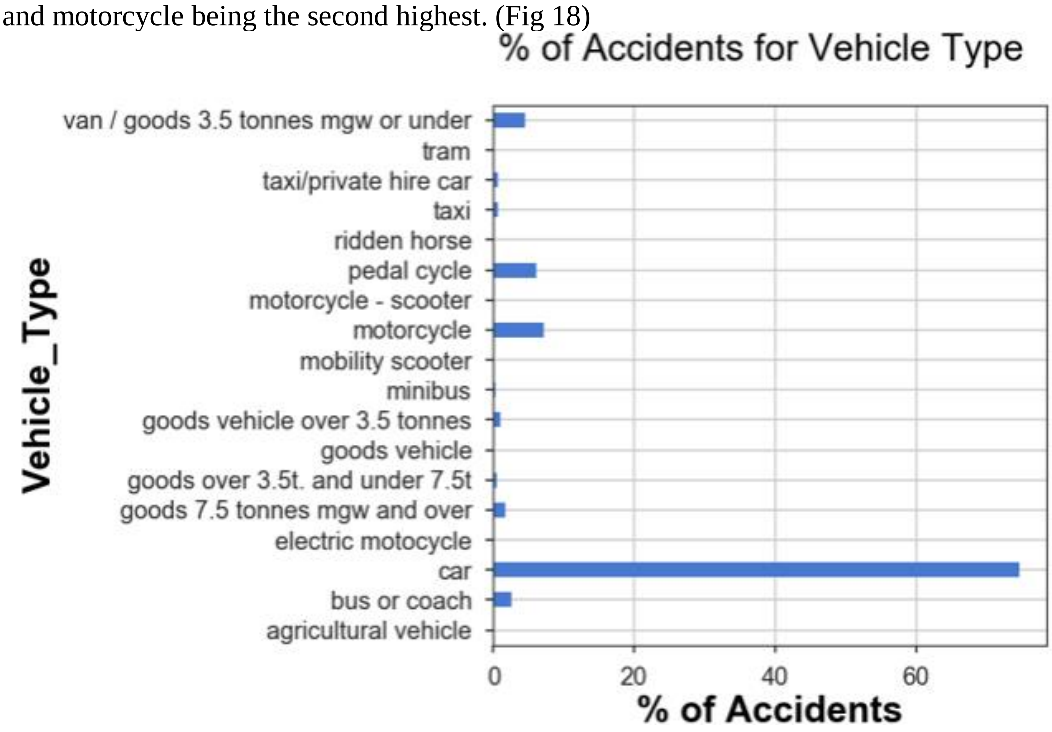

Figure 18, showing the total number of accidents that each vehicle type was involved in.

Cars are involved in roughly 67% more accidents than motorcycles. (Fig 19)

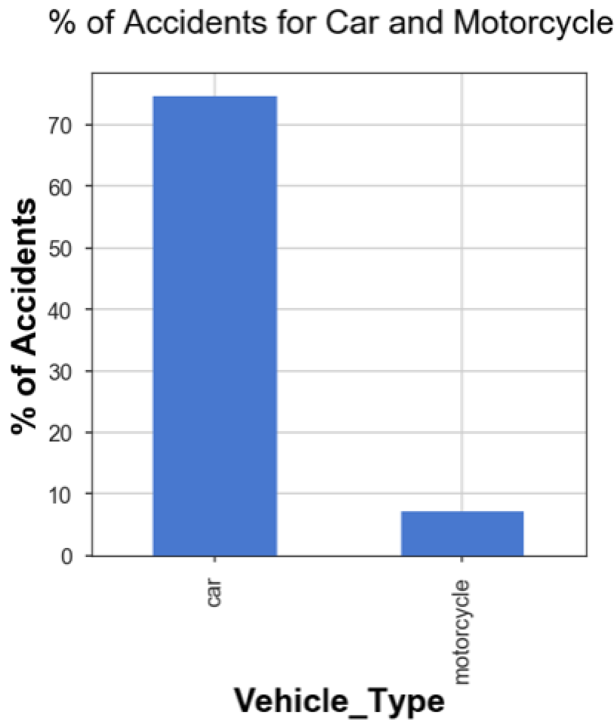

Figure 19, comparing the total number of accidents that both cats and motorcycles weres involved in.

On average, accidents involving cars have in more casualties than accidents involving motorcycles. (Fig 20)

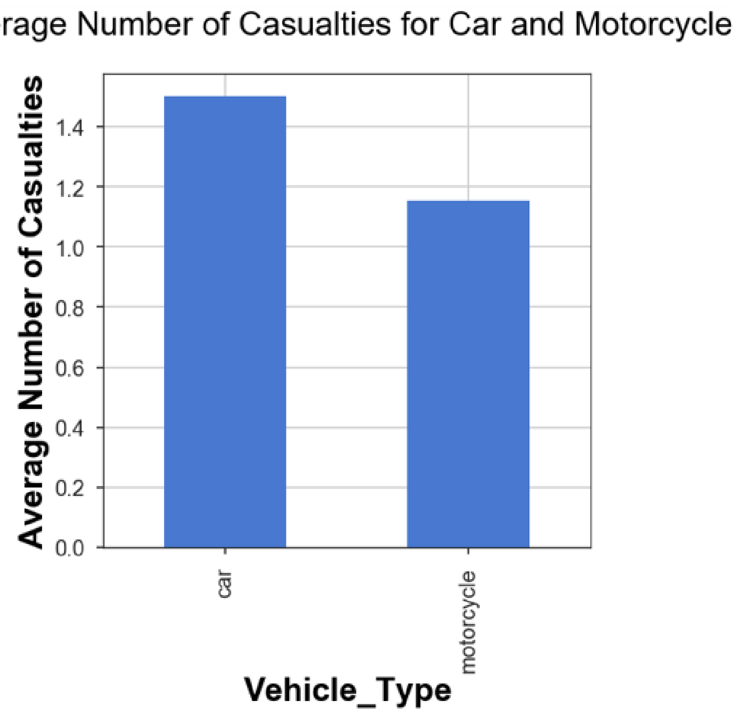

Figure 20, comparing the average number of casualties in accidents for both cars and motorcycles.

Both cars and motorcycles follow the same distribution in regards to the % of fatal, serious and slight accidents. However, motorcycles are involved in more serious and fatal accidents than cars. (Fig 21)

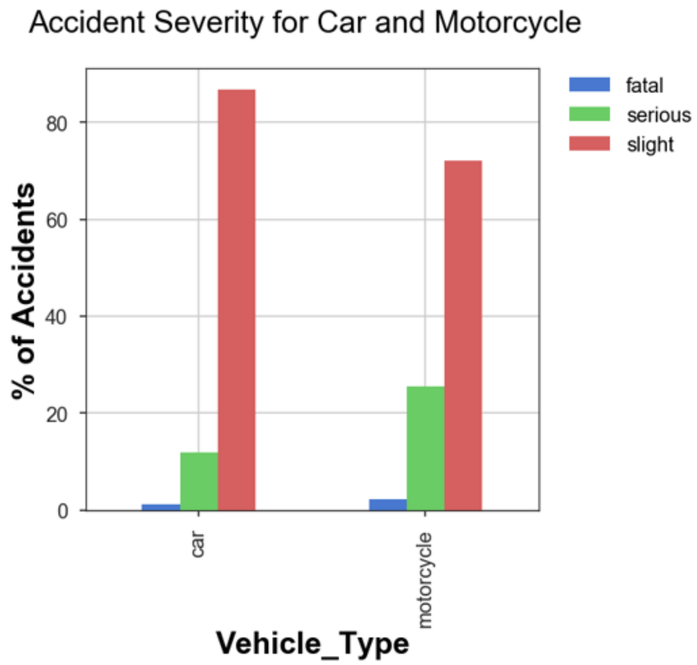

Figure 21, comparing the % of fatal, serious and slight accidents that both cars and motorcyles are involved in.

Both cars and motorcycles have a very small % of accidents in rainy weather conditions overall. However, motorcycles have a higher % of serious accidents in comparison to cars which have a higher % of slight accidents. (Fig 22)

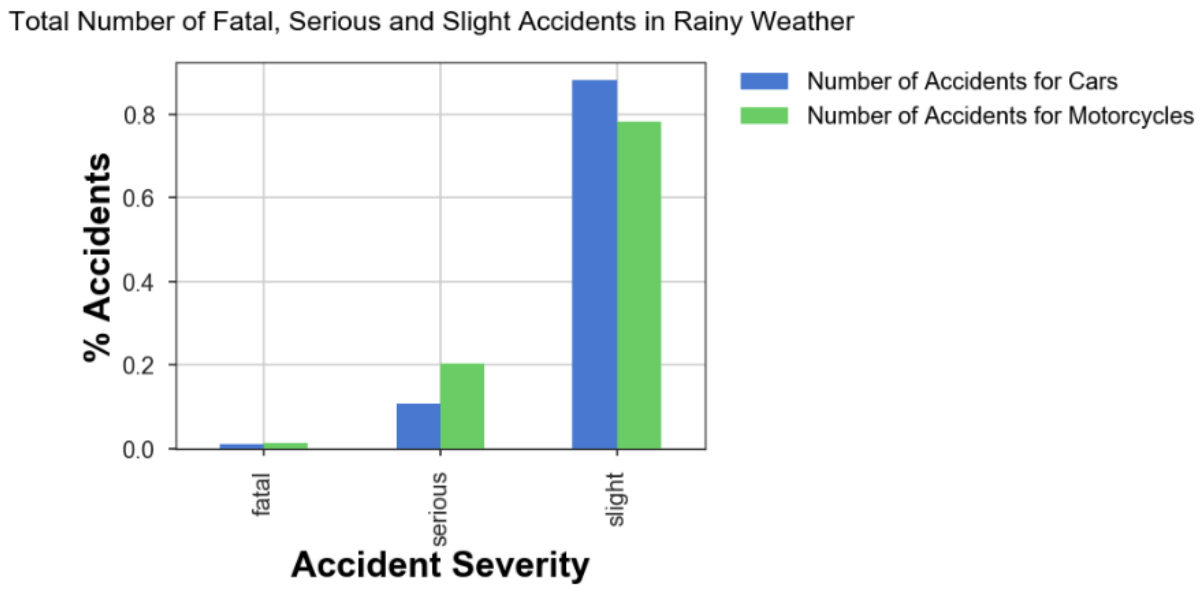

Figure 22, comparing the % of total fatal, serious and slight accidents in rainy weather conditions for cars and motocyles.

4.2.4. Discussion

(A)

Our data tells us that newer vehicles are involved in more accidents when compared to older vehicles. However, we found this to be a very surprising result and therefore split the data into two time periods to analyse the data more closely. Despite this we found that both followed the same trend as before. We were unable to obtain data defining the number of “new” and “old” vehicles on UK roads for any given year which we could use to normalise the data. Therefore, this is a question for possible further research. When looking at the number of casualties we found that newer vehicles have fewer casualties on average. Additionally, older vehicles are involved in more fatal and serious accidents when compared to new vehicles. This suggests that perhaps new technologies have made new cars safer in terms of casualties and accident severity. This also corresponds with our findings that the number of road accidents has declined in recent years. Therefore, we decided to have a closer look at this by looking at airbags and AEB brakes in relation to accident severity. However, we failed to see a decline in the number of serious and fatal accidents for both significant enough to conclude that they had a strong influence on road safety. Therefore, further analysis could be carried out on this by looking closer into the influence of modern technologies on factors other than accident severity.

(B)

From our analysis we found that cars are the vehicle type involved in the highest percentage of road accidents by far (Fig 18) which we believe is because there are more cars on the road than any other type of vehicle. From this we decided to compare cars with the next highest vehicle type which was motorcycles. We found that on average there are more casualties in accidents involving cars. We believe the reason for this is that cars can carry more passengers than motorcycles which can carry at most two. However, we were interested to discover that motorcycles were involved in serious and fatal accidents when compared to cars but cars were involved in more serious and fatal accidents in rainy weather conditions. This may be because fewer people would travel by motorcycle in rainy weather conditions. Therefore, we cannot conclusively say that it is safer to travel by car than by motorcycle.

4.3. RQ3 - Does Location Impact Road Safety?

(A) Is it Safer to Drive in Urban or Rural Areas?
(B) Are Drivers from 'most Deprived' IMD Decile Regions Involved in the Most Road Accidents

4.3.1. Datasets

(A)

The data for part A of our third research question uses UK road safety data between 1999 and 2014. Using data on whether an accident occurred in an urban or rural area we looked at the number of accidents by using the ‘Accident_Index’ column where each row represents an accident. Additionally, we used details on the accident severity which tells us whether the accident was fatal, serious or slight. In our dataset we have information on the number of casualties in each accident which we used to compare casualties in urban and rural areas. From this we then used the longitude and latitude coordinates to map the areas with the highest average number of casualties.

(B)

To answer our research question three (B), we required road accident data from 1999 – 2014. For this question, we had 2 input files which we performed our analysis on. Results were then aggregated for these datasets for each step of our analysis. Within these datasets, we specifically required information regarding the IMD Decile region of drivers involved in each road accident. We also required data listing the latitude, longitude and local district of each recorded road accident. This data was available within our datasets.

4.3.2. Approach

(A)

Q: "Are there more accidents in rural or urban areas?"
Since we are specifically focusing on comparing the number of accidents in urban and rural areas the data was grouped by the 'Urban_or_Rural_Area' column and got the total count of accidents within both groups. (*Note: The 'Urban_or_Rural_Area' indicates if the accident occurred in an urban or rural area.)* This was repeated for both data frames being used to answer this question and the results were aggregated accordingly. The results were then normalised by the total number of accidents to get a total % of accidents in rural and urban areas for a more direct comparison.

Q: "Are accidents more or less severe in urban or rural areas?"
To analyse the severity of accidents in urban and rural areas the data was grouped by the 'Urban_or_Rural_Area' and the accident severity. The count of the total number of accidents for each group was then calculated using the accident index. This was repeated for both data frames and the results were aggregated accordingly. Following this the data was normalised to get a further breakdown of the accident severity for both areas.

Q: "On average, are there more casualties in accidents in urban or rural areas?"
Again, for both data frames the data was grouped by the 'Urban_or_Rural_Area' column and the average number of casualties in an accident for both areas was calculated. This was repeated for both data frames question and the results were aggregated accordingly.

(B)

To analyse our data for question three (B), we first grouped each road accident by our 'IMD Decile of driver' column. We then calculated the total recorded road accidents for each IMD Decile. This allowed us to analyse whether specific IMD Decile regions were involved in a higher or lower number of road accidents than others. Once this was done, we grouped all road accidents by our 'local district' column. We then counted the total number of accidents which occurred in each district. After this, we calculated the latitude and longitude for each local district which had one or more recorded road accidents. This was calculated by getting the average latitude and longitude of all road accidents which had the same local district. This allowed us to analyse the average location in which road accidents occurred from 1999 – 2014, and to examine whether these locations showed an even distribution pattern.

4.3.3. Results

(A)

There are more accidents in urban areas than rural areas. (Fig 23)

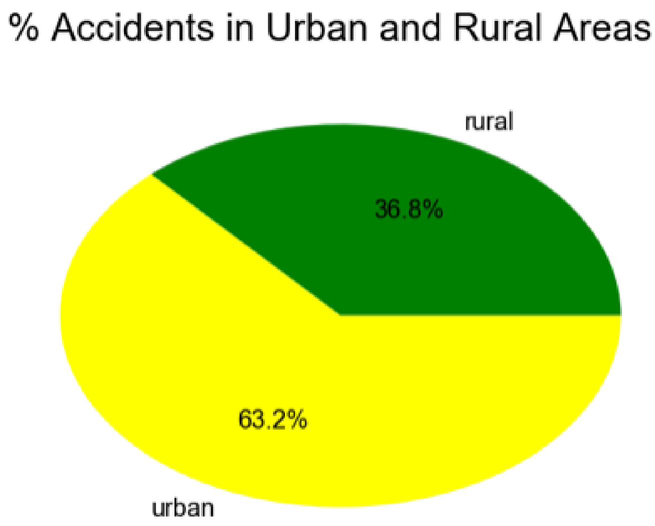

Figure 23, showing the % of accidents in urban and rural areas.

There are more serious and fatal accidents in rural areas. (Fig 24)

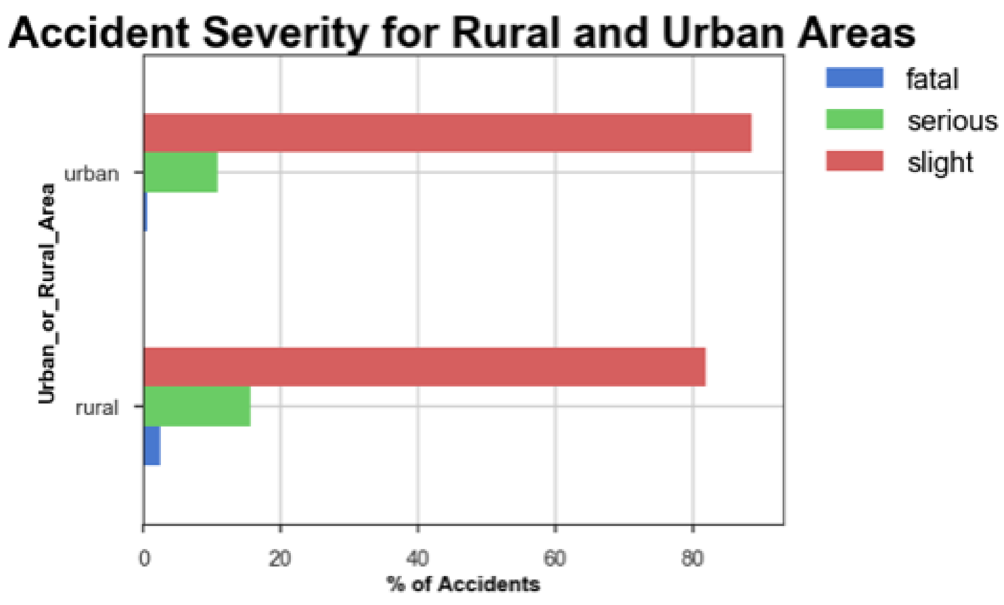

Figure 24, showing the % of fatal, serious and slight accidentsin urban and rural areas.

On average, there are more casualties in accidents in rural areas than urban areas. (Fig 25)

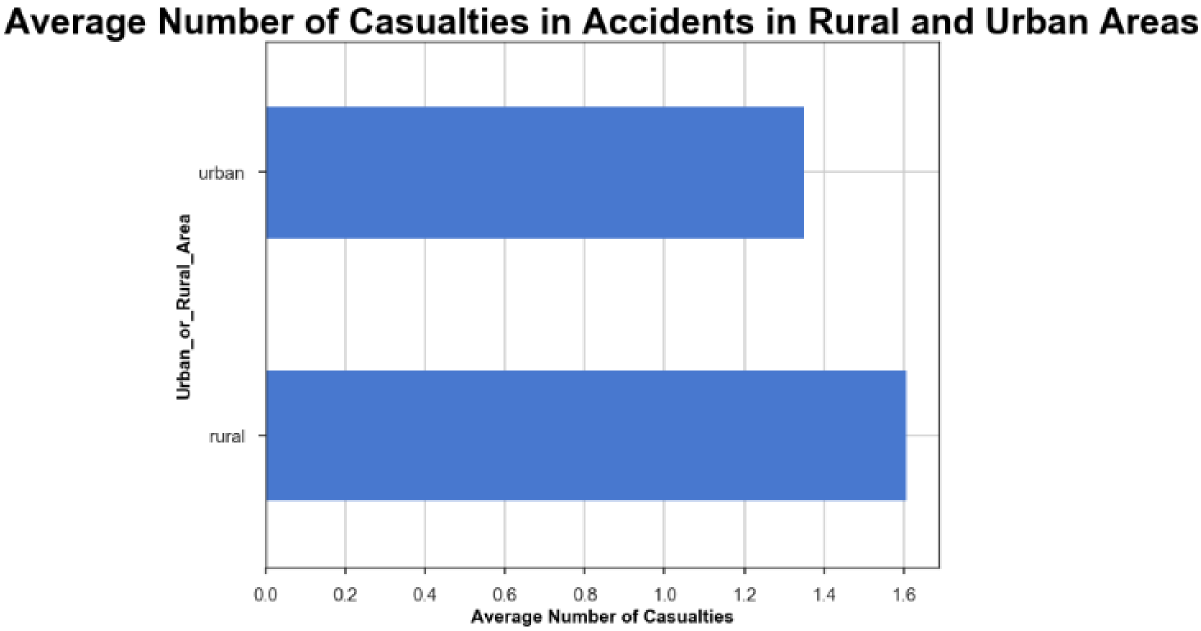

Figure 25, showing the average number of casualties in accidents in urban and rural areas.

Accidents in rural areas which are represented by the pink points incur more casualties than urban areas which are represented by the purple dots. (Fig 26)

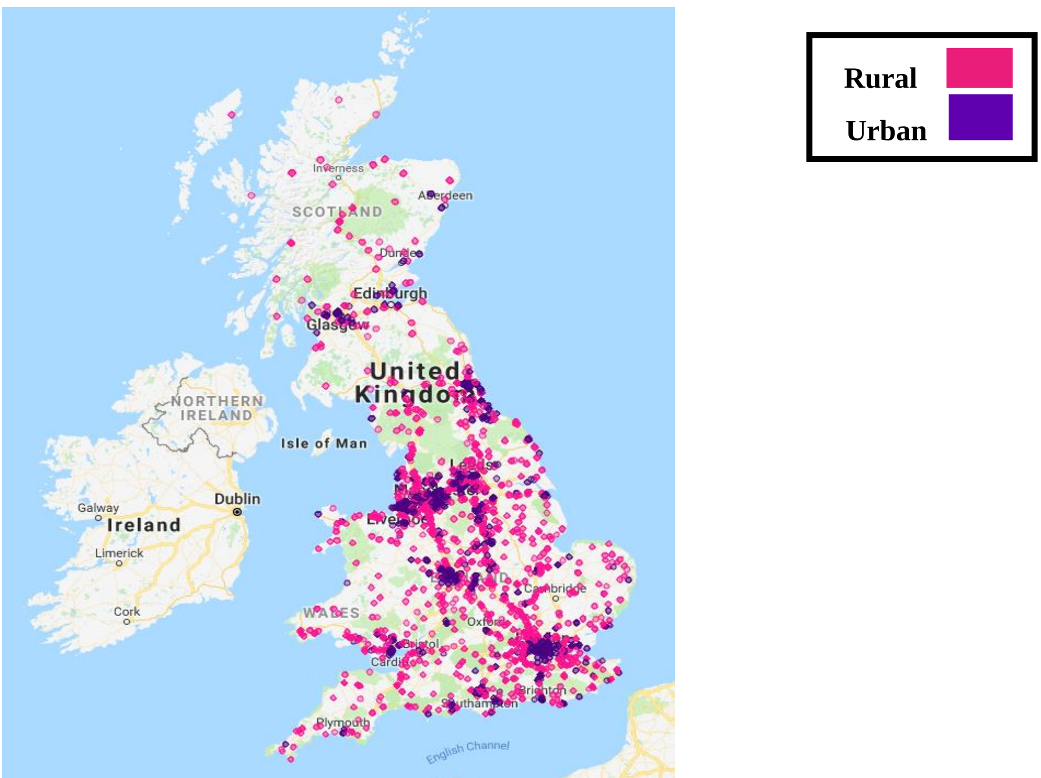

Figure 26, showing the average number of casualties in urban and rural Areas
(pink-rural, purple-urban)

(B)

Drivers from 'least deprived 10%' IMD Decile regions have been in the least amount of recorded accidents. Drivers from 'most deprived 10-20%' IMD Decile regions have been in the most amount of recorded accidents. (Fig 27)

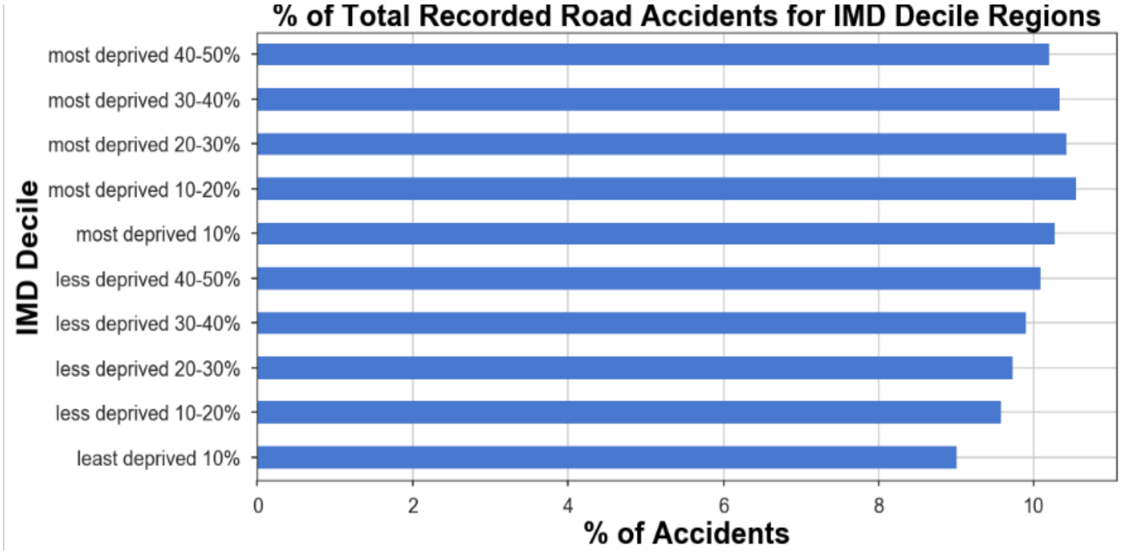

Fig 27, showing the % of road accidents involving drivers from each 'IMD Decile' region.

Local districts which have one or more recorded road accidents are concentrated in urban/high population areas such as London and Manchester. (Fig 28)

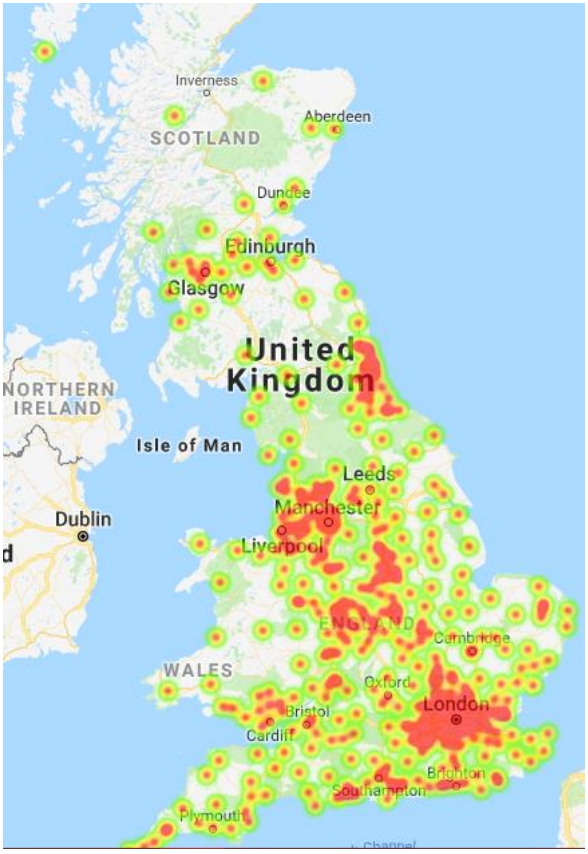

Fig 28, showing local districts with recorded road accidents

4.3.4. Discussion

(A)

There are more accidents in urban areas than rural areas with urban areas accounting for ~63% of road accidents in comparison to ~36% in rural areas. We believe that this large difference in due to the large population density in urban areas in the UK. After learning this we were interested to compare the severity of accidents and found that in fact there are a higher percentage of serious and fatal accidents in rural areas. Also, road accidents in rural areas have a higher number of casualties on average. Therefore, from our data we can suggest that it is safer to drive in urban areas.

(B)

Fig 27 shows that 'least deprived 10%' Driver IMD Decile has been in the least amount of recorded accidents between 1999 – 2014. This graph also shows that 'most deprived 10-20%' driver IMD Decile regions have been in the most amount of recorded accidents between 1999 - 2014. This indicates that perhaps 'most deprived' driver IMD Decile regions are involved in more road accidents than 'least deprived' driver IMD Deciles. However, there may be population imbalances within 'IMD Decile' regions and as a result we cannot conclusively say that this is the truth.

We also found from Fig 28, that local districts with recorded road accidents were unevenly distributed though-out the UK, and of these are concentrated in urban areas such as London and Manchester. This may display how population spread is uneven throughout the UK, and perhaps within IMD Decile regions. Unfortunately, we could not find information recording the populations within each 'IMD Decile' region. However, we believe that future work into the differing population densities within each 'IMD Decile' would provide a clearer idea of whether drivers from 'most Deprived' IMD Decile regions are involved in the most road accidents.

5. Discussion

A further discussion into important aspects of our research project, including ethical considerations taken, ensuring reproducibility of our work and limitations to our project.

5.1. Ethical Considerations

We acknowledge that the data we are analysing contains records of real life fatalities. We understand that our data is not merely facts and figures, as each of these numbers represents an individual with family and friends. Therefore, we will make a conscious effort to be respectful of the individuals represented in our data, and mindful of their families. We will also ensure the use of humanizing and dignifying terms throughout our work.

Further ethical considerations for this project are as follows. Firstly, we responsibly obtained openly sourced datasets, which ensured the reproducibility of our work. We also examined our datasets to identify any personal information which may have been present. Due to the nature of our data, it was important to exclude any sensitive information which could be used to identify individuals. It was imperative to ensure that our project would not enable the identification of those who have tragically passed away in a road accident. Likewise, we must prevent the possible identification of drivers involved in road accidents. Subsequently, our datasets did not contain this sensitive information, therefore no anonymisation was necessary.

Another ethical consideration for this project was to identify how our results could be used by third parties. For example, we do not wish to contribute to any incriminating bias which could result from our findings. As we are focusing on attributes of drivers involved in road accidents, we must ensure that we do not place a generalisation upon any group of individuals. As such, upon analysing road accidents based on driver 'IMD Decile', we chose to follow the IMD Decile Metric established by the UK government (Department for communities and local government. (2015)). We were unsure of the use of the word 'Deprivation' when describing regions within each IMD Decile. As a result, we will refer only to regions as they are defined within the metric. For example, we will use the term 'lower IMD Decile areas' rather than 'more deprived areas'.

5.2. Reproducibility

Throughout our project we took any necessary steps to ensure the reproducibility of our work. Maintaining reproducibility is essential to enable future work to be carried out on road safety. The following measures were taken to guarantee reproducibility within our project.

We included detailed markdown cells in all coding work, describing exactly what actions we took and why. Where appropriate we also included links to websites which we referenced when making important decisions within our analysis. For example, we used UK Met Office information (Met office, UK (2018)) to define the seasons of the year for first research question.

We applied a standardised naming system for each of our files and commented our code throughout to ensure that anyone reading our work would be able to comprehend it. Additionally, we have created a README file listing details on how the project is organised, what each notebook contains and how the code files are named. We also created a file which provides relevant details of the environments we used to conduct our project, named 'requirements.txt'. All the raw datasets which we used are openly available on the UK Government website, and are readily downloadable. Our code can also be found publicly on GitHub (Joyce O., O'Reilly C., 2018).

5.3. Limitations

Limitations to our work include the following. Firstly, we had hoped to further analyse the impact of technological advancements on road safety. However, we could not obtain the relevant data such as the year that certain technologies became compulsory by law, which would enable us to conduct this comparative research. We were also unable to obtain figures indicating the number of total road users within the UK, as well as the average number of 'new' and 'old' cars on the road for each year of our datasets. This limited our research and prevented us from normalising our data based on these figures.

Lastly, we had aimed to produce a predictive model which could be used to calculate whether you would 'survive the drive'. This predictive model would take a set of parameters such as time, weather and location of a proposed journey, and subsequently calculate a percentage safety score for this trip. Unfortunately, we were working within a limited time frame and were unable to produce this.

6. Conclusions & Future Work

A final analysis of our work and results, including possible areas for future work.

The intention of our research project was to investigate the key factors of road accidents, and to analyse whether and how they play a role in road accidents. Our overall aim was then to share this information, to contribute to improved road safety awareness both in the UK and Ireland. We analysed UK road accidents between 1979 to 2016, focusing on attributes such as weather, time, vehicle type, age of vehicle and location. We found a varying degree of influence for each of these attributes on the severity and volume of road accidents.

We found that there is a safest time for travel, though this varies by season and time of day. Through our research we found that it is safer to drive in the Winter rather than the Summer, as accidents are less severe. We also found that it is safer to drive during the day time than at night, as although there are fewer accidents at night time, they result in more severe outcomes by comparison. Our research has shown us that there is no clear evidence within our dataset that indicated that occupational drivers are more or less dangerous than non-occupational drivers.

To examine what vehicles are involved in the most road accidents, we first analysed the age of vehicles involved in road accidents. Our data showed that newer vehicles (under 10 years old) were involved in the most road accidents, however these accidents resulted in less severe outcomes than accidents involving older vehicles. We found that this may have been due to improved technology and vehicle regulations, such as the introduction of air bags and Autonomous Emergency Breaking (ABS). However, we found that newer vehicles were safer to drive in rainy weather conditions both in terms of accident severity and total number of accidents.

We found that location impacts road safety. Our results have shown that most accidents occur in urban areas. However, we found that the average number of casualties in urban areas was roughly 1.3, compared to roughly 1.6 for rural areas. We also found that accidents which occurred in rural areas resulted in a higher percentage of fatal and serious outcomes when compared to accidents in urban areas. This perhaps indicates that it is safer to drive in urban areas than in rural areas. Through examining the ‘IMD Decile’ region of drivers involved in road accidents, we could not conclusively declare whether this plays a major role in road safety.

There were many areas of future work which became apparent to us throughout the course of this research project. These include the following. The safety of cyclists on our roads, focusing specifically on features of road accidents involving cyclists. An analysis of Irish Research Data, a similar analysis of road accidents which have been recorded by the Road Safety Authority, to promote safer road usage in Ireland. Another area of potential future work is a further analysis of how the IMD Decile region of drivers may influence road safety, focusing on the population of each region and popular vehicle types, ages, and make/model. Finally, an interesting area of future work on road safety would be the creation of a prediction model, whereby users can specify journey information (e.g. time, location, weather conditions etc.) and a prediction will be generated of how likely you are to ‘Survive the Drive’.

7. Bibliography

Relevant Citations used throughout our research projects

Department for Transport, UK. (2017). *Road Safety Data.* UK: Department for Transport. Retrieved March 12th, 2018, from https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

Department for transport, U.K. (2017). Reported road Casualties in Great Britain: 2016 annual report. Retrieved 10 May, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/648081/rrcgb2016-01.pdf

Department for transport, England. (2014). National Travel Survey. Retrieved 10 May, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/458414/how-people-travel-car.pdf

McKnight, A. J., & McKnight, A. S. (2003). Young novice drivers: careless or clueless? *Accident Analysis & Prevention*, *35*(6), 921-925.

Hislop, D. (2012). Hanging on the telephone: Mobile phone use patterns among UK-based business travellers on work-related journeys. *Transportation research part F: traffic psychology and behaviour*, *15*(2), 101-110.

Lewis, I. M., Watson, B. C., White, K. M., Elliott, B., Thompson, J., & Cockfield, S. (2012). How males and females define speeding and how they'd feel getting caught for it: Some implications for anti-speeding message development.

Broughton, J., & Walter, L. (2007). *Trends in fatal car accidents: Analysis of CCIS Data* (No. PPR172).

Broughton, J. (2008). Car driver casualty rates in Great Britain by type of car. *Accident Analysis & Prevention*, *40*(4), 1543-1552.

Department for communities and local government. (2015). The English Index of Multiple Deprivation (IMD) 2015 – Guidance. Retrieved 10 May, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/464430/English_Index_of_Multiple_Deprivation_2015_-_Guidance.pdf

Met office, UK (2018). Seasons. Retrieved 10 May 2018, from https://www.metoffice.gov.uk/learning/seasons/

McKenna, F. P., Waylen, A. E., & Burkes, M. E. (1998). Male and female drivers: how different are they?

Joyce O., O' Reilly C., *Will you survive the Drive*, (2018), GitHub repository, https://github.com/ciaraaoreilly/C-O

8. Appendix

Additional findings and graphs.

8.1. An Appendix Subsection

Most road accidents occurred on ‘dry’ road conditions in the seasons of Summer, Spring and Autumn. The highest number of road accidents in the Winter occurred on ‘wet or damp’ road conditions. (Fig 29)

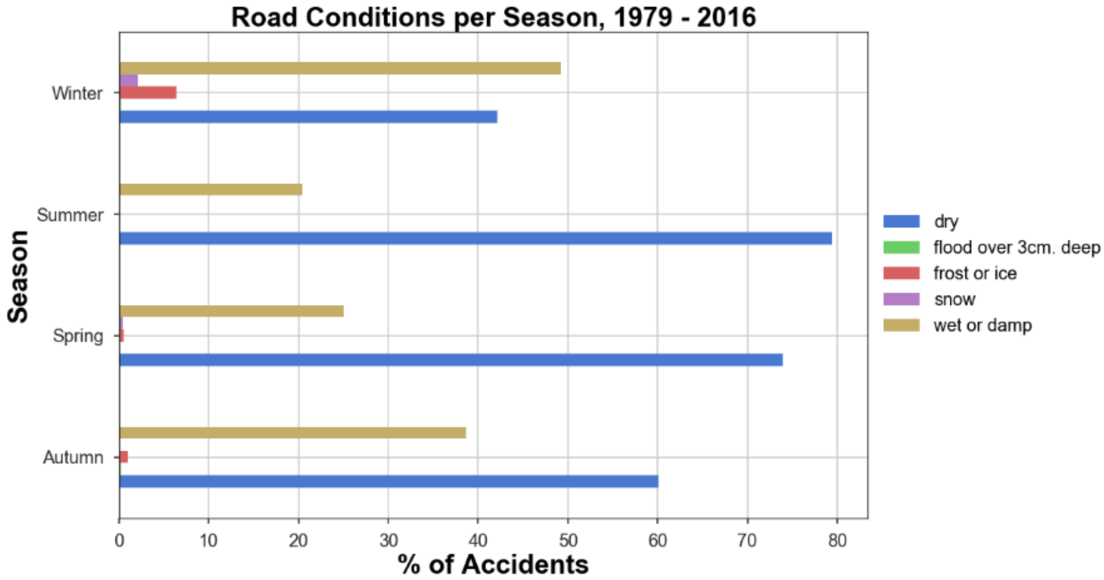

Figure 29, showing the road conditions at the time of accidents for each season

Minibuses have the most number of casualties on average. There is a close comparison between the 7 vehicles types after minibus with them all having between ~1.4 and 1.5 casualties. Mobility scooters have the least number of casualties. (Fig 30)

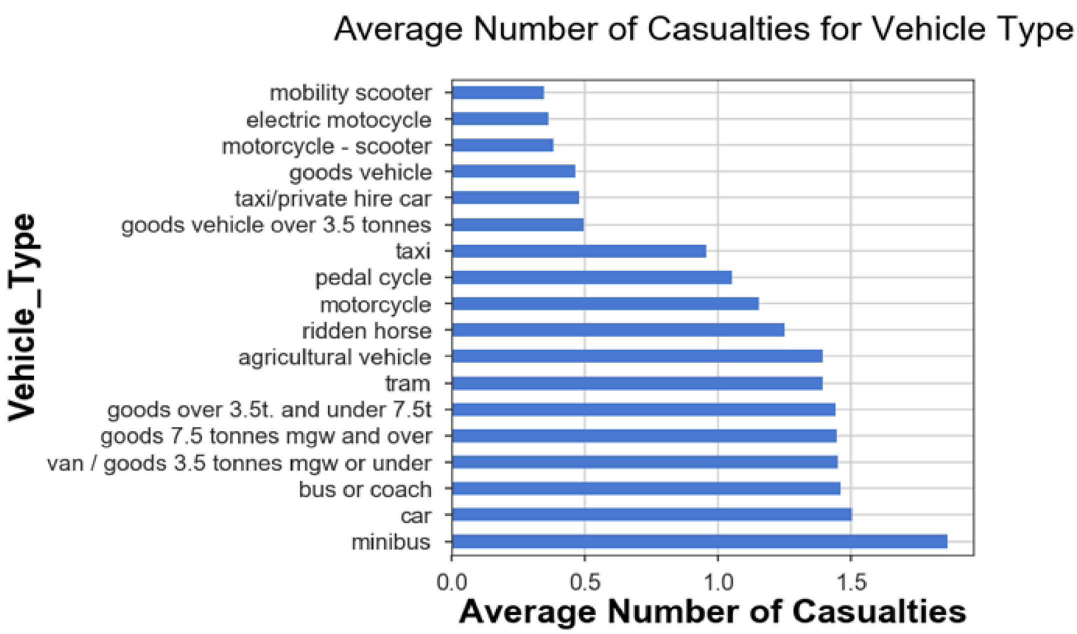

Figure 30, showing the average number of casualties in accidents for each vehicle type.

Overall, all vehicle types have a much higher % of slight accidents when compared to the number of serious and fatal accidents. Also, all vehicle types have fewer fatal accidents than serious accidents. (Fig 31)

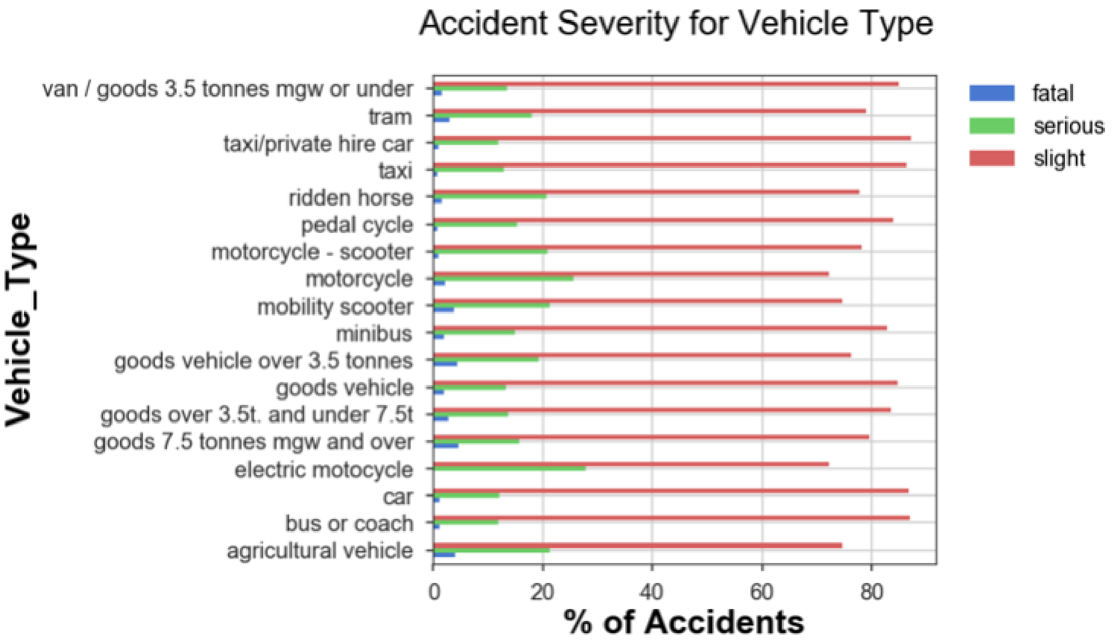

Figure 31, comparing the % of fatal, serious and slight accidents that all vehicle types are involved in.

Both cars and motorcycles have a low % of fatal accidents in all rany weather conditions. cars are involved in more slight accidents in 'raining no high winds' but motorcycles are involved in more serious accidents in 'raining no high winds'. (Fig 32)

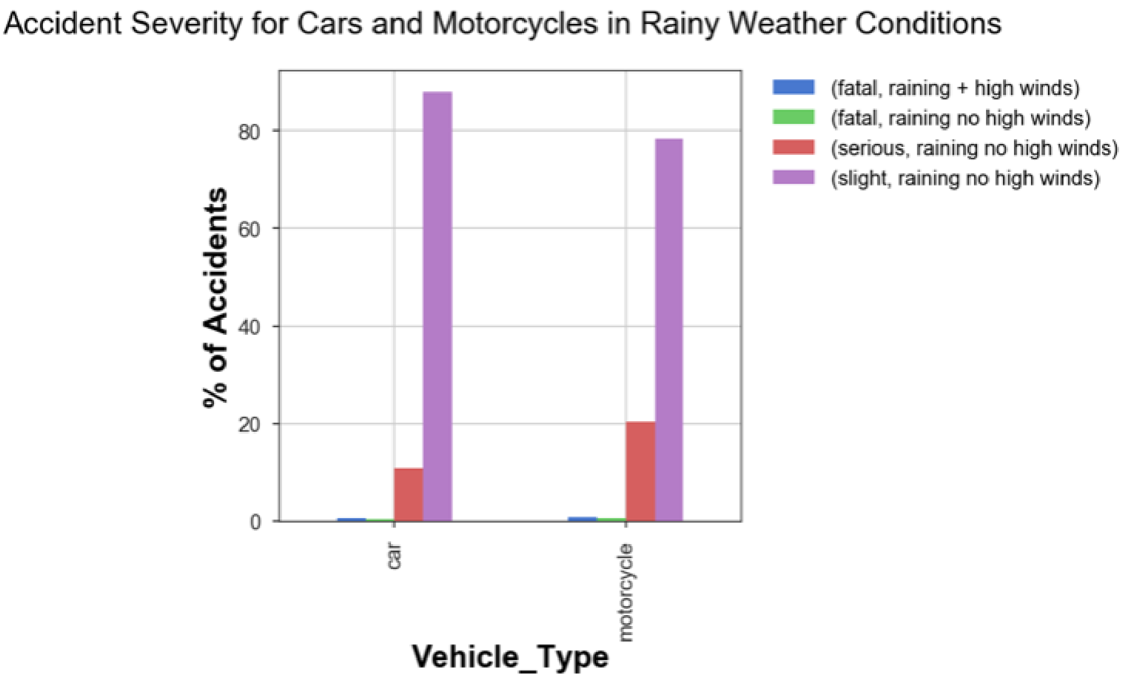

Figure 32, comparing the % of fatal, serious and slight accidents in each rainy weather condition for both cars and motocyles.

In certain years there is a decline in the average number of casualties in both urban and rural areas. However, there is not a big enough difference to say that there was a significant improvement. (Fig 33)

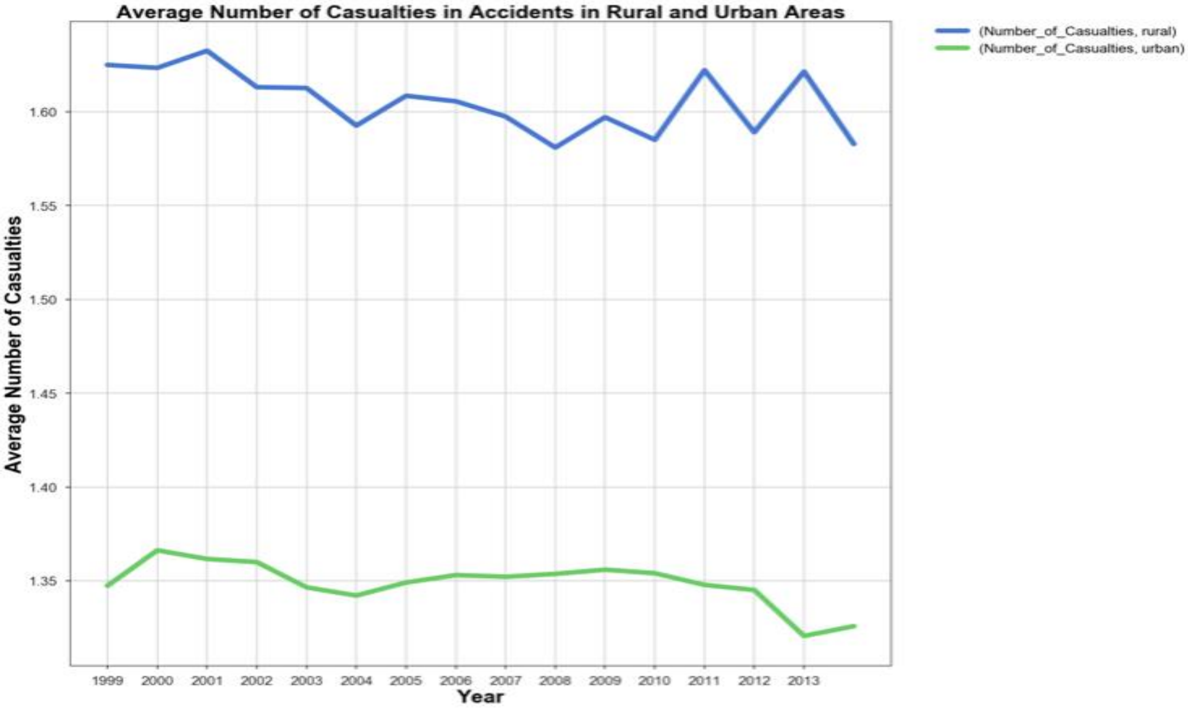

Figure 33, showing the average number of casualties in accidents in urban and rural areas between 1999 and 2014.