Crash Test Dummies

Contents:

Abhisek Gautam (13679042)

Bruno Quevedo (13500069)

Ganesh Arunagiri (13662388)

Priscilla Soenario (11685169)

Priyanka Srinivasa (13684182)

Sam Knowles (13762467)

Front cover source: https://besthqwallpapers.com/landscapes/asphalt-road-forest-highway-usa-mountain-9549

1. Executive Summary

As at 30 September 2019, the number of lives lost on Victorian roads in 2019 reached 205 (TAC 2019). This figure is above the targeted 200 and working its way towards the 2018 fatality level of 213. From an investigation of historical data since 2006, road accidents in Victoria are most common amongst young adults.

The causes of the increasing number of accidents may be multifactorial, it is however important to treat and eliminate some of these factors so the accident wouldn't occur to begin with. As the number of vehicles increases, the number of accidents go up as well. This calls for effective measures to be taken to reduce the mortality rate.

Given the nature of the datasets employed in this investigation and the question sought to be answered, a binomial logistic regression model was employed to ascertain the probability of death of severe injury, given the host of variables modelled.

While the exploratory data analysis looks at the total population of crashes, the regression model itself focuses on 2018, uncovering highly significant results in the majority of variables included.

Most notably it finds wet weather has a weak and negative relationship with the injury outcome of individuals in road crashes. Moreover, it reaffirms the significance and weight variables associated with current Towards Zero initiatives have, with the speed zone, seat belt usage and road quality aspects all strong indicators.

Thus, it's recommended that Towards Zero continue their policy trajectory.

2. Introduction

Road accidents claim many lives each year and hospitalise many others. In 2016, 5.34 road deaths occurred in Australia per 100,000 people (Bureau of Infrastructure, Transport and Regional Economics 2018). The number of vehicles registered each year in Australia continues to increase with a 1.7% increase from Jan 2018 to Jan 2019, inline with a 1.6% increase in the country's population from Dec 2017 to Dec 2018 (Australian Bureau of Statistics 2019, Australian Bureau of Statistics 2018). These increases indicate the potential for a greater number of road accidents as a result of more vehicles on the road combined with a greater population.

The Towards Zero initiative is the road safety strategy for the state of Victoria, aiming to reduce the number of road fatalities to nil. The initiative's approach encompasses a shift in philosophy where the responsibility for reducing road deaths and serious injuries lie with the system designers and not the individual road user. Hence designing a safe road system is integral to the success of the initiative.

The short term objective of Towards Zero is to reduce annual road fatalities to below 200 by 2020 and reduce serious injuries, which require hospital admission, by 15% over the 5 year period between 2016-2020 (Towards Zero 2019b).

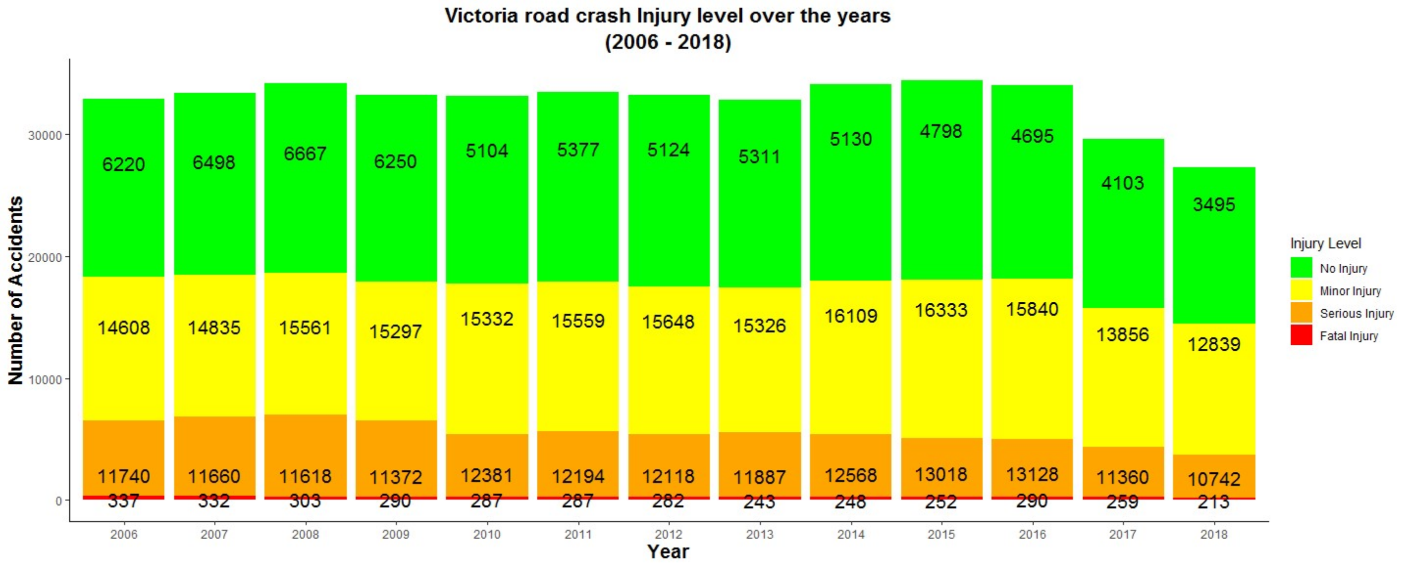


Figure 2.1 - Number of Persons in accidents over

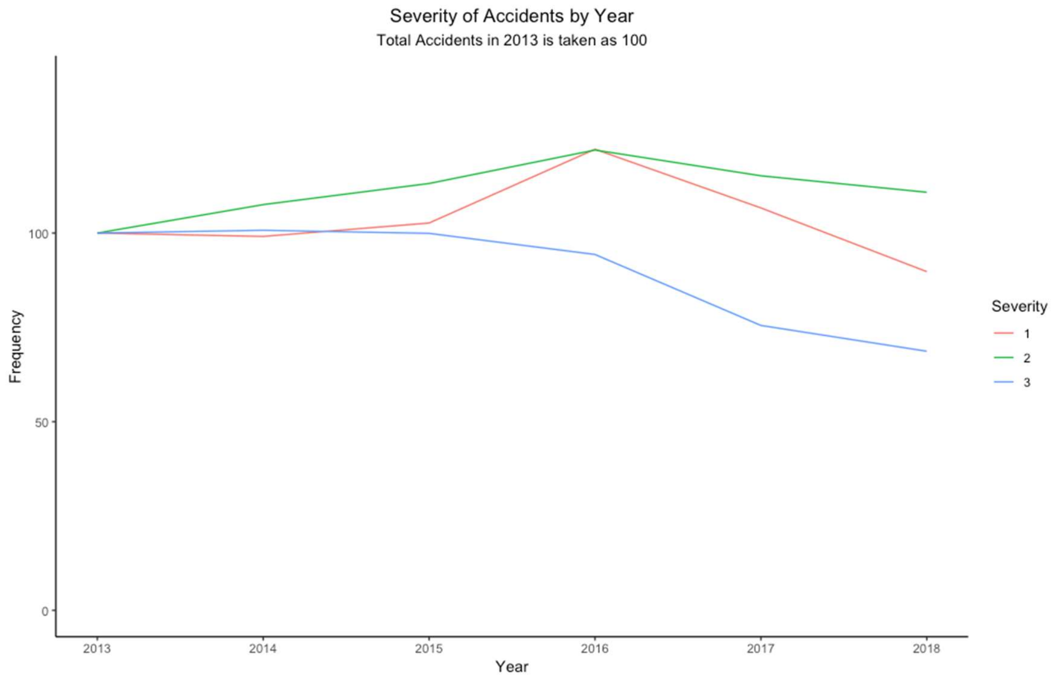


Figure 2.2 – Number and severity of accidents relative to the 2013 accident count.

Number of accidents at 2013 attributed a value of 100. The change in accident frequency of successive years is plotted to obtain relative frequency-

Since the implementation of the Towards Zero strategy in 2016, progress has been made to reduce both road fatalities and serious injuries as seen in Figure 2.1 where declines in number of people involved in accidents, and severity are evident.

Using the number of accidents in 2013 as a base (3 years before Towards Zero) Figure 2.2 indicates that the relative frequency of accidents followed an upward trend to 2016, before declining since the implementation of the Towards Zero initiative. Even though the frequency of accidents involving a serious injury are higher in 2018 relative to 2013, the number of fatal injuries has decreased indicating a journey towards zero.

Although the number of lives lost on Victorian Roads has decreased since Towards Zero with 213 fatalities in 2018, the September 2019 year to date fatality count of 205 may present a break in this trend and a shift back upwards (Towards Zero 2019a).

To continue the Towards Zero journey, this investigation aims to identify the key factors influencing crash severity and hence focus the State Government's resources and policies towards creating an efficient and effective safe road system.

Weather Conditions as a factor of Road Safety

Image Source: https://wall.alphacoders.com/tags.php?tid=23526

The Towards Zero action plan to 2020 encompasses multiple elements of safer roads, safer vehicles, safer people and safer speeds. Weather conditions are a factor which may contribute to each of these elements of the initiative and hence should be considered.

Inclement weather conditions are widely regarded as a contributing factor to increased risk of road accidents. Conditions such as precipitation, high winds and temperature extremes may affect driver capabilities, traffic flow and vehicle performance via traction, stability and manoeuvrability (Baguley 2019; Goodwin 2019).

It's surprising, then, to see the Towards Zero campaign omit any explicit consideration in their policy positions towards the mitigation of risk when roads are adversely affected by rainfall. Therefore, the purpose of this investigation is to ascertain how the likelihood of severe injuries and death change with the consideration of rainfall at the crash-site, in addition to observing the effects of those phenomena already identified as problematic.

3.1 Data Acquisition

To assess the effect of numerous variables on road accident severity, multiple datasets are required. The datasets selected satisfied the criteria of a credible source, sufficient number of records, granularity at an individual level and relevant variables to address the proposed questions. The datasets identified and used for the investigation are as follows:

1. Road Accident Data

Provided by VicRoads, the state road and traffic authority, the road accident dataset contains up to 130 variables for over 180 thousand accidents occurring between 2006 and 2019 in Victoria. The dataset contains ten files, each with specific descriptive information regarding the event, location, conditions, vehicles and persons involved in the accident. Figure 3.1 below provides a brief description of each file. The complete list of variable details can be seen in Appendix A - Datasets and Merging.

File Name	Description
ACCIDENT.CSV	Basic accident details, date, time, severity, injuries, etc.
ACCIDENT_LOCATION.CSV	Road accident details (road name, road type, etc.)
PERSON.CSV	Person based details, age, sex, etc.
VEHICLE.CSV	Vehicle based data, vehicle type, make, etc.
ACCIDENT_EVENT.CSV	Sequence of events e.g.: left road, rollover, caught fire, etc.
ROAD_SURFACE_COND.CSV	Whether the road was wet, dry, icy, etc.
ATMOSPHERIC_COND.CSV	Rain, Wind, etc.
SUB_DCA.CSV	Detailed codes describing accident, etc
NODE.CSV	Master location table (NB subset of accident table), latitude, longitude, etc.
ACCIDENT_CHAINAGE.CSV	Has detailed route and chainage data, etc.

Figure 3.1 – Road Accident Dataset

2. Rainfall Dataset

The rainfall dataset is sourced from the SILO database, hosted by the Queensland Department of Environment and Science (DES). This dataset provides gridded daily rainfall data across Australia, with a resolution of 0.05° latitude by 0.05° longitude (approximately 5 km × 5 km). The daily rainfall data was derived by interpolating observational data from the Bureau of Meteorology (BOM), using the ordinary kriging method (Jeffrey, Carter, Moodie & Beswick 2001). Although a detailed dataset, there are some limitations which are described in detail in section 6. Limitations.

Although the original rainfall dataset is not in a data frame format,

Field	Description
Date	Date in daily format (01/01/2019)
Longitude	Latitude ranging from 10°S to 44°S with a resolution of 0.05°
Latitude	Longitude ranging from 112°E to 154°E with a resolution of 0.05°
Rainfall (mm)	Daily rainfall (millimetre)

Figure 3.2 – Rainfall dataset

3. Victoria Population Dataset

The population dataset provides demographic information of Victoria by Local Government Area (LGA) between 2013 and 2018, including the number of total people and the population density. Provided by ABS, this dataset enables contextual layer to identify areas of anomalies with respect to frequency of crashes. Figure 3.2 demonstrates the variables used in this research.

Field	Description
Code	Local Government Area ID
Label	Local Government Area (i.g.: Melbourne)
Year	The referred year
Estimated Resident Population – Total	Number of people living by LGA
Population Density	The estimated population density by LGA

Figure 3.3 – Rainfall Dataset

4. Additional Data

In addition to the datasets above, additional GIS files were utilised to aid spatial data visualisations and exploratory data analysis. These files are outlined in Figure 3.4 below

GIS	Source	Description
Victoria State Outline	Australian Government – Open Data	Spatial Lines including all roads within the state of Victoria, from major highways to local council roads.
Victorian Local Government Areas	Australian Government – Open Data	Spatial Polygon outlining the state of Victoria
Victorian Roads	Victoria State Government – Spatial data	Spatial Polygon of all 93 Local Government Areas within the state of Victoria

Figure 3.4 – GIS Data

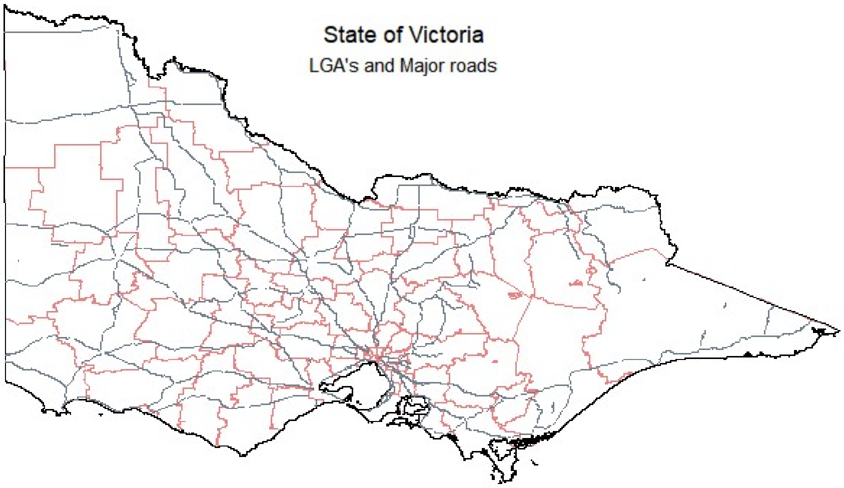


Figure 3.4 – Spatial map of Victorian LGAs and Major roads

3.2 Data Merging

The vehicle table was merged with the accident table to obtain generic information about the accident, such as date, time, light condition, severity. The vehicle table was then merged with the person table in order to get information about the driver's age and sex, as well as the number of people in the vehicle during the accident and the number of serious injured or deaths per vehicle after the accident. Furthermore, to obtain the road type where the accident occurred, the accident location table was also combined. Last, the accident node table, which contains latitude, longitude and LGA, was merged. All tables were joined using a common identifier, the Accident Number. The person and vehicle tables were merged using another common identifier as well, the Vehicle ID.

The population and the vehicle dataset were merged using two keys, year and LGA. Some manual transformations were performed to treat inconsistent naming conventions.

Lastly, the rainfall dataset was merged with the vehicle dataset using the latitude, longitude and date keys. This merge was complex due to the nature of the rainfall dataset, an NC file format, containing millions of values spread across a considerable amount of time and space. Having converted the dataset to an R data frame, filtering was conducted on latitudinal and longitudinal data with the help of a shapefile which was necessary to remove unused values and reduce computational resources. The last step done was rounding the accident latitude and longitude values to two decimal places, similar to the rainfall dataset, enabling the merge between both files.

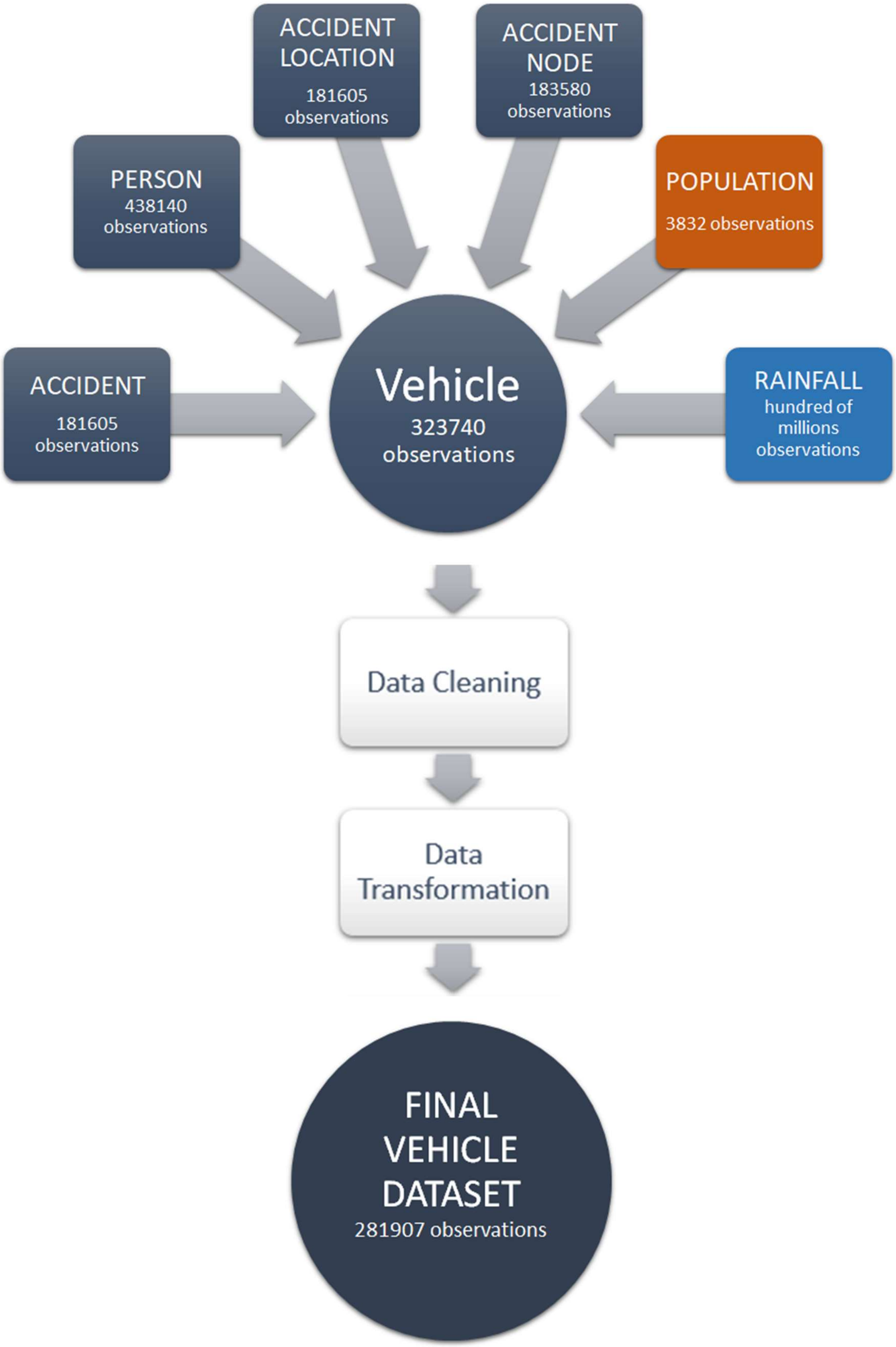


Figure 3.5 – Flowchart of the merging, cleaning and transformation of data

3.3 Data Cleaning & Transformation

Data cleaning and transformation are important steps to execute before modelling to ensure accuracy of results. Unsurprisingly, the merged dataset contained missing values that needed to be resolved, as can be seen in Figure 3.6. As a result, the following were deleted:

- Where speed limit was unknown or missing
- Where driver's age or sex was missing;
- Where lighting conditions were unknown.

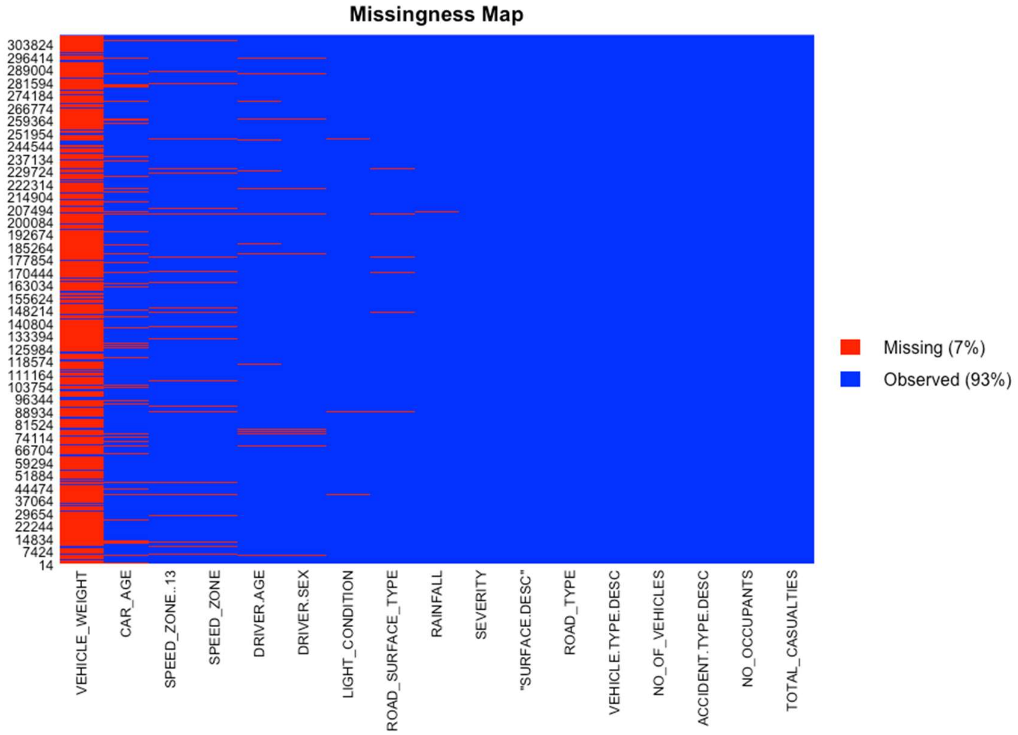


Figure 3.6 – Map of missing observations within the final dataset

Moreover, some variables and values needed to be transformed before exploration:

- A count of casualties (individuals who were seriously injured or dead) per vehicle was created to be used in the regression model
- A count of individuals who weren’t wearing security equipment per vehicle was created
- Although the vehicle dataset provides the age of manufacture of the car, a categorical variable was created to group the ages
- A variable was created to show how old was the car at the moment of the accident, by subtracting the accident year and age of manufacture of the car
- The light condition variable had two values for dark roads: roads with lights turned off, and roads without any lights. These two values were merged into one
- During the EDA, we have realised that people aged 70 and over had substantial difference comparing to other ages regarding their injury level after an accident. Hence, a variable was created to count all the individuals in the vehicle aged 70 and over.

Source: https://www.australia.com/en-gb/trips-and-itineraries/perth-and-surrounds/crossing-the-nullarbor.html

4. Exploratory Data Analysis

Population Density

Densely populated areas are expected to have a higher frequency and severity of accidents, given a greater number of people, vehicles and roads. Hence, it is unusual to find that the greatest proportion of fatalities occur in sparsely populated areas as identified in Figure 4.1 below.

As per Figure 4.2, rural LGA's have a lower population density. Therefore, the higher accident count in rural areas are a result of factors other than population density, such as road quality, light conditions and higher speed zones.

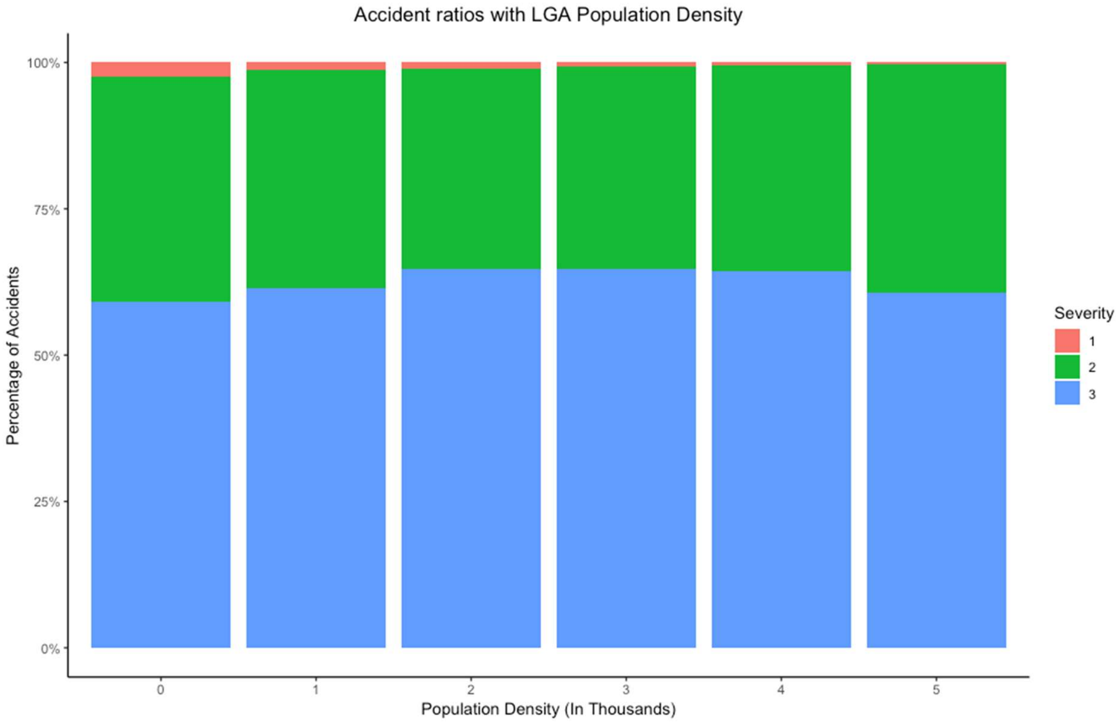
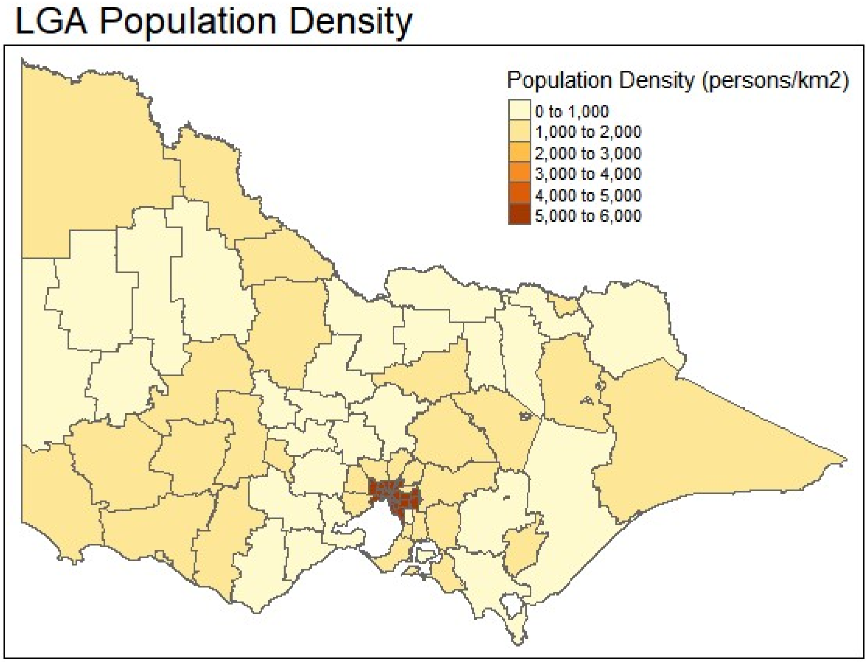


Figure 4.1 –Accident severity ratios by population density where fatalities are attributed a severity of 1

Figure 4.2 - Victorian Local Government Area Population Density



Speed Zones

The speed at which a vehicle is travelling impacts the severity of an accident. An indicator of vehicle speed is the speed zone. Figure 4.3 indicates that the majority of fatalities occur within the 100km/hr speed zone, followed by the 60km/hr and 80km/hr zones respectively.

Fewer fatalities are present in speed zones below 60km/hr and at the 110km/hr, indicating that driver behaviour may vary at different speed limits.

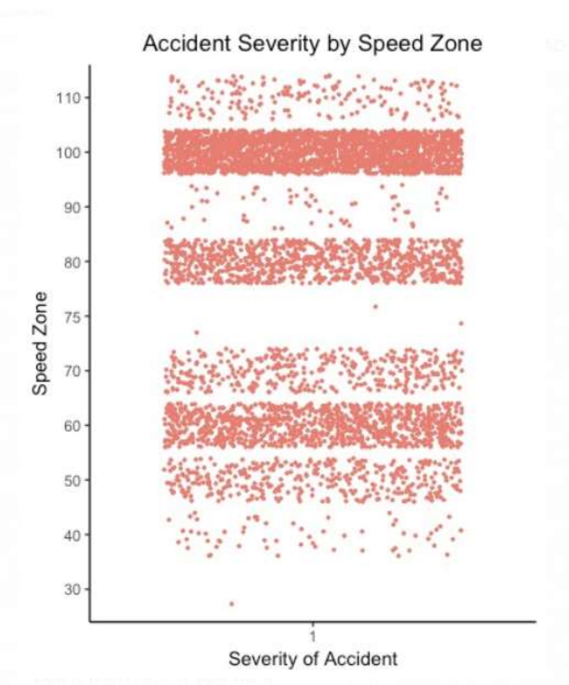
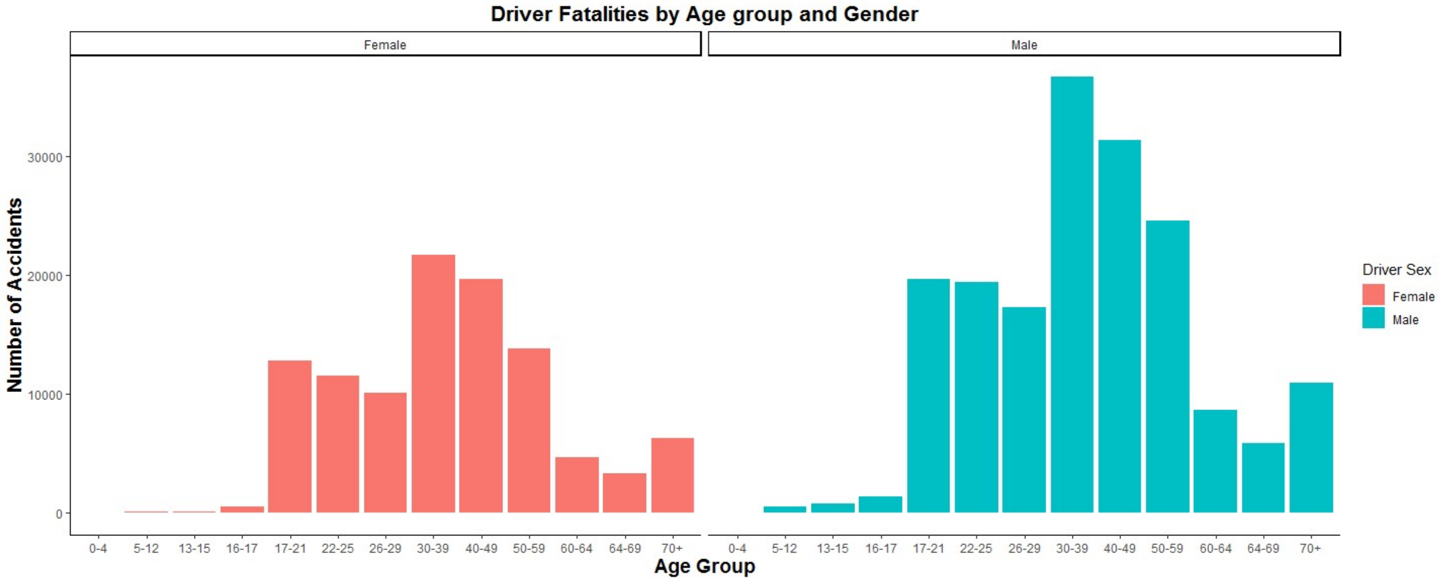


Figure 4.3 - Fatalities by speed zone

Driver Characteristics:
Gender

Figure 4.4 – driver fatality by age group and gender



Overall, male driver fatalities are higher than female fatalities, although this may be a result of a greater number of male drivers in comparison to female drivers. Drivers in the older age groups are generally more fragile and hence susceptible to injuries with a higher severity. Across both genders, greater number of fatalities are observed in the 70+ age group, clustered around the 30-50 year olds and the 17-21 age group. Although it may be a factor of the number of drivers on the road of each age group, the distribution consistency across driver gender illustrates a potential for differences in driver behaviour between these age groups.

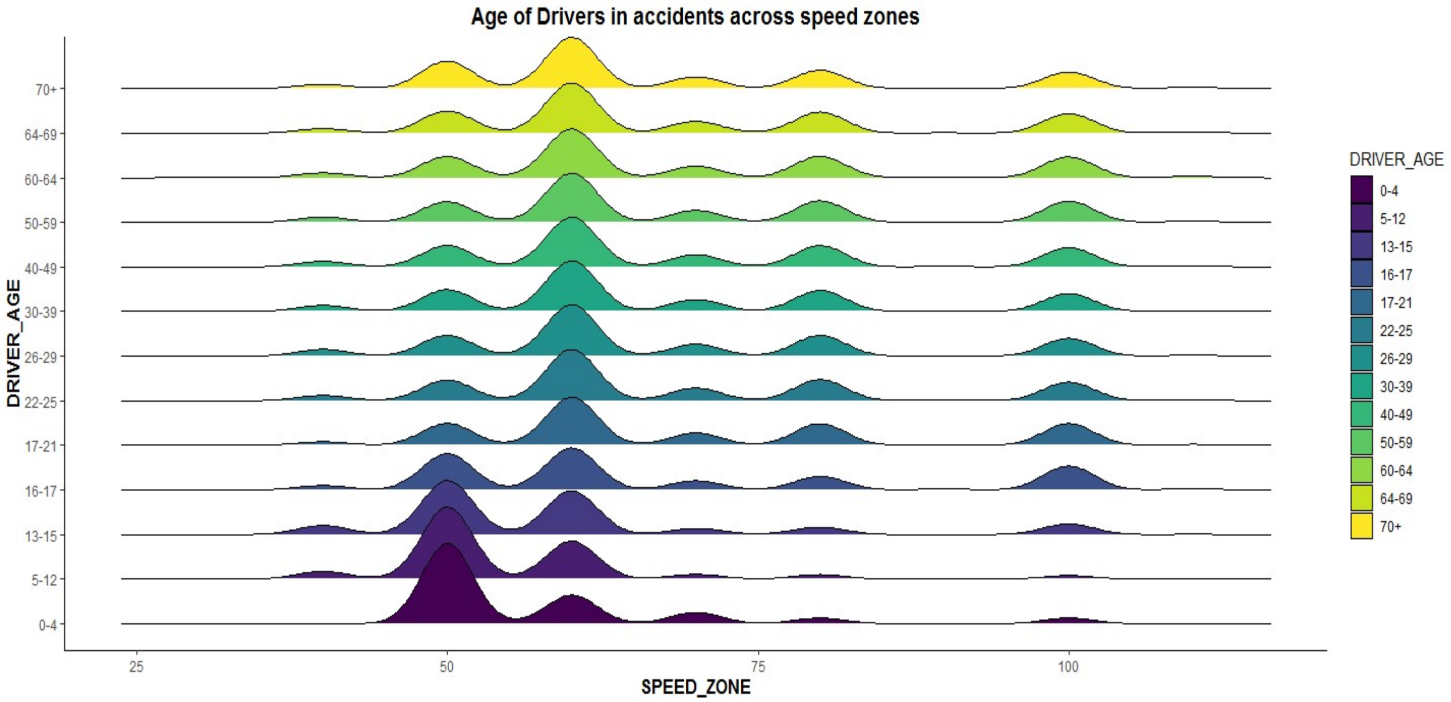
Driver Age

In the state of Victoria, a drivers licence can be legally issued at the age of 16 with no upper restriction or regulations on older drivers (VicRoads 2019). Although there is no legal obligation, medical examinations are recommended for older drivers to verify their ability to safely operate a vehicle. In Victoria, there is also no speed restriction for any licences types (VicRoads 2019).

The ridgeline plot in Figure 4.5 indicates a similar distribution of accidents in age groups above 16, with some small deviations in the 16-17 age group.

Interestingly, drivers under the age of 16 have a higher distribution of accidents within the 50km/hr speedzone. This might seem, questionable at first, but after further investigation these are identified as drivers of bicycles. Therefore, vehicle type should be explored next.

Figure 4.5 Distribution of accidents by driver age in multiple speed zones-



Vehicle Type

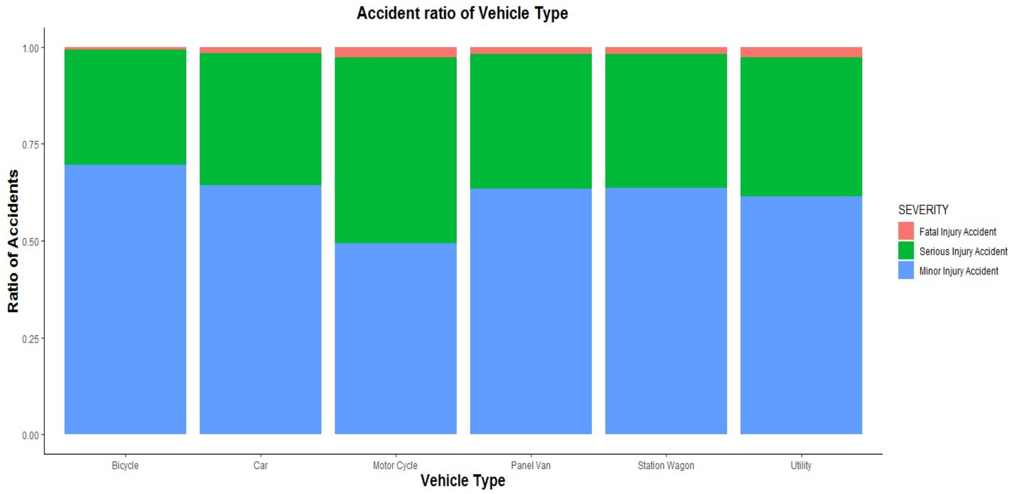


Figure 4.6 Driver fatality by gender and age group

As can be observed, motorcycles are exposed to a higher rate of injury severity in crashes, which is logical. Perhaps the low figures for bicycle drivers is reflective of the higher proportion of children riding them, likely in residential areas, and the slower speeds they tend to travel.

As this category seems to be quite significant, we will look to explore all those numerous categories omitted from the visual in our regression efforts later.

Light Condition

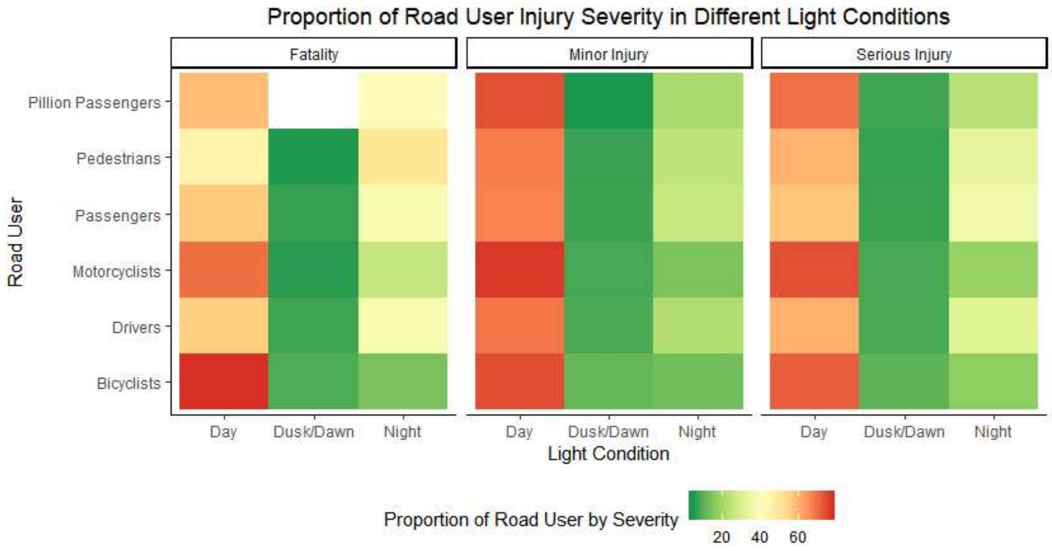


Figure 4.7 Road user injury severity at different times of the day

Figure 4.7 is calculated through the formula below in Figure 4.8, focusing on the impact of light conditions on injury severity for each road user.

$$Proportion\ of\ Road\ User\ by\ Severity = \frac{Count\ by\ Road\ User, Severity\ and\ Time\ of\ Day}{Count\ by\ Road\ User\ and\ Severity}$$

Figure 4.8 Proportion of Road user by Severity formula

Across all injury level, a greater proportion of injuries occur during the day. As injury severity increases, injury proportions during dusk/dawn remain low. On the other hand, as injury severity increases, there is a shift in proportion of injuries from day towards the night column. This change illustrates that light condition is a factor which influences the injury severity an accident.

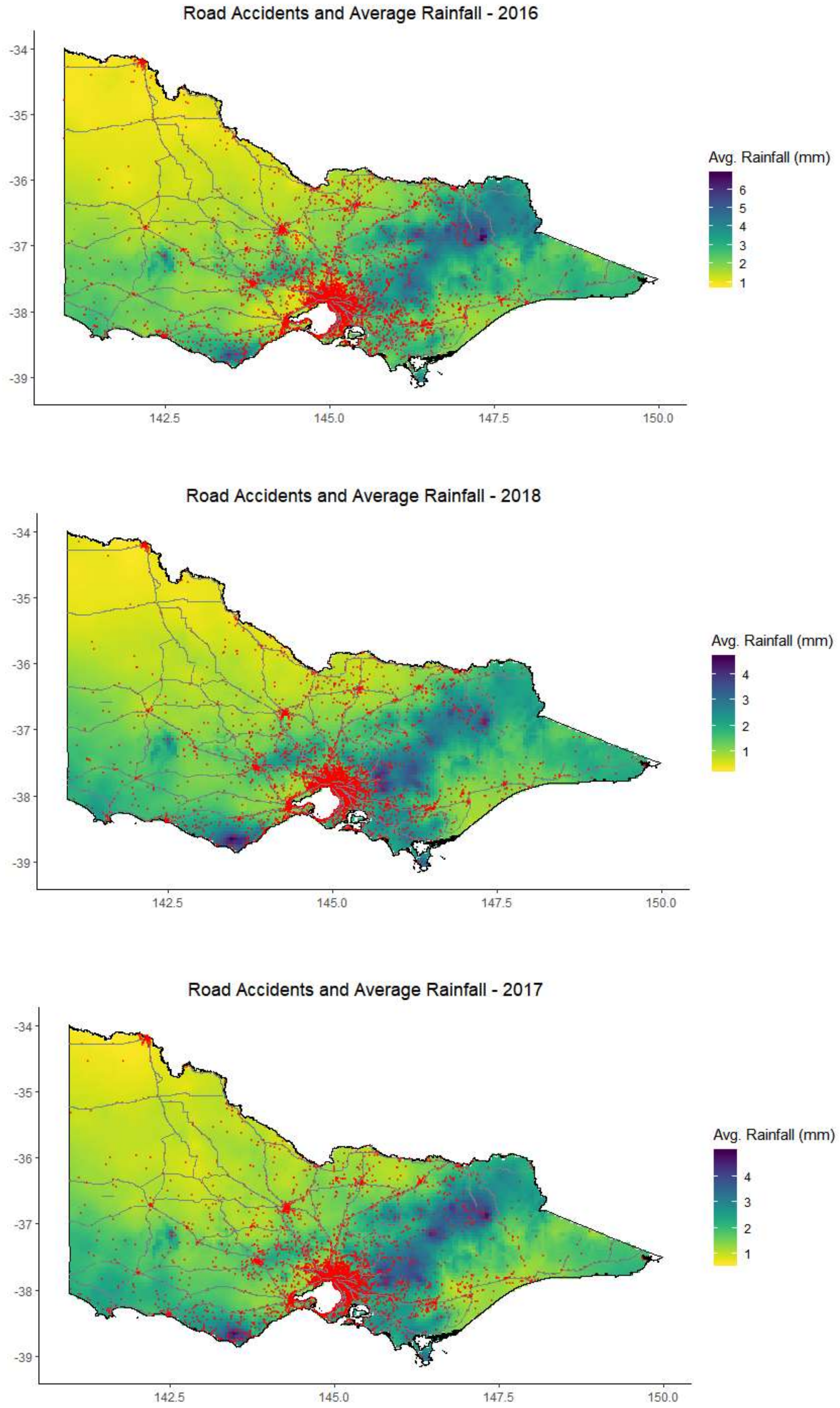
This shift is most evident in the pedestrian, driver and passenger categories and hence indicate these road users are more susceptible to lighting conditions.

Road Type and Rainfall

In Figure 4.9, the spatial distribution of fatalities and serious injuries are highly concentrated around the greater Melbourne area. As we move away from the city areas, accidents are observed along the major highways and freeways throughout the state, clustering around the intersections of multiple major roads.

Figure 4.9 also represents the average annual rainfall across Victoria, with particular high rainfall along the mountain ranges of the Australian and Victorian Alps in the north east and along the southern coast of the state.

The spread of accidents and average rainfall remain fairly consistent year on year


Road Accidents and Average Rainfall - 2016

Road Accidents and Average Rainfall - 2018

Road Accidents and Average Rainfall - 2017

Figure 4.9 Geographical spread of road accidents and rainfall over the past 3

5. Regression Model

Data Set Selection

For the purposes of this model, we will only be analysing those individuals utilising a vehicle. Pedestrians caught in accidents, while embodying their own interesting subset of data and strong relationships therein, deserve their own investigation.

With this investigation aiming to inform present decision making, the dataset is narrowed to crashes occurring during 2018.

Model Justification

As the primary objective of the Towards Zero campaign is reducing the number of deaths to zero with the implied secondary objective of reducing severe injuries, we sought to adopt a binomial logistic regression model to predict the impact of explanatory variables on the probability of an individual dying or incurring a severe injury during a crash.

The binomial logistic regression model is ideal as it works well with binary response variables. Additionally, the output will allow the user to predict the number of casualties given a crash scenario as defined by corresponding values of the modelled inputs.

Assumptions

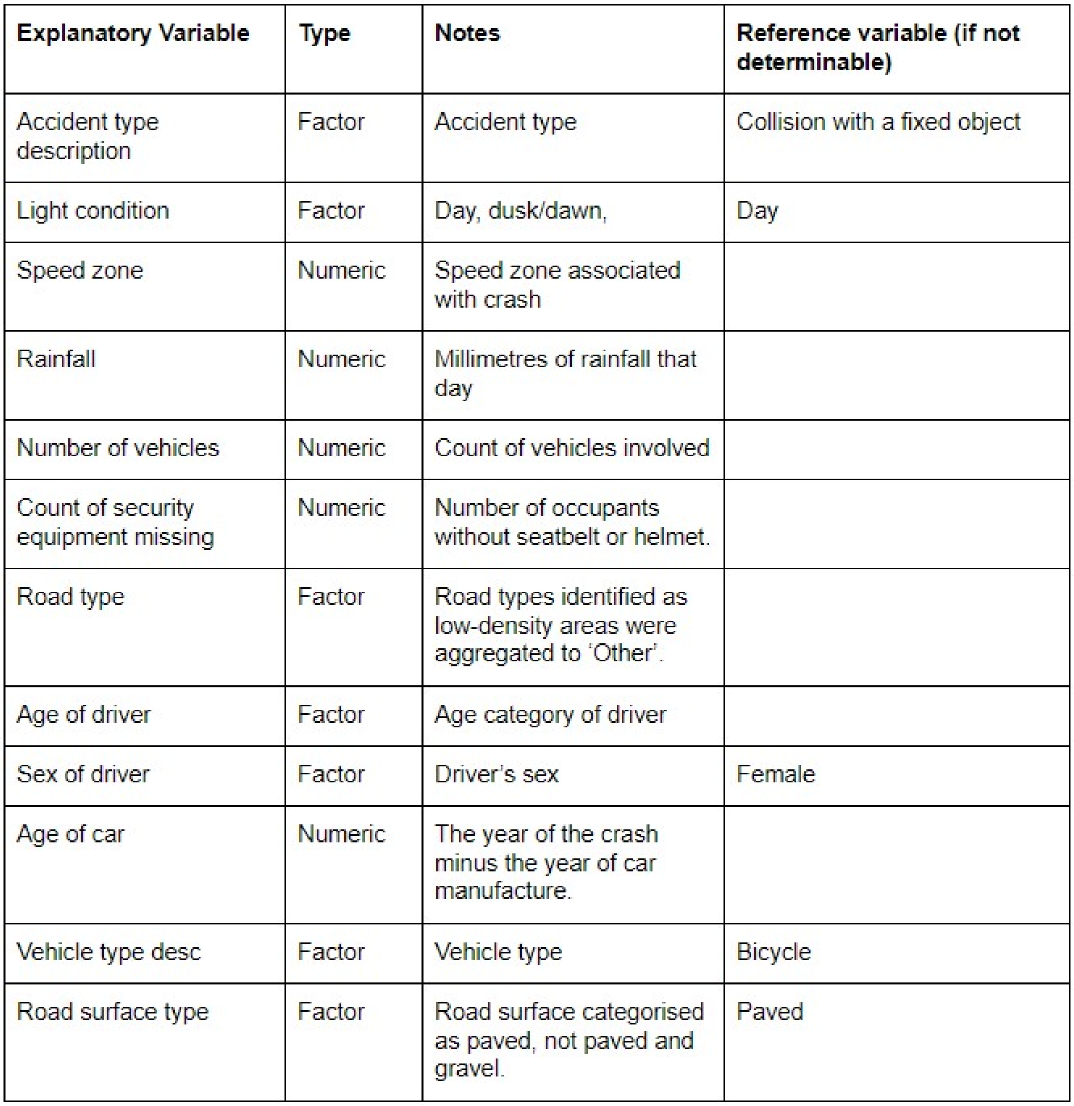
The binomial logistic regression embodies a number of assumptions, including:

- The response variable is binary
- The probability that y = 1 is consistent with the outcome whose probability is sought
- Error terms need to be independent, along with minimal multicollinearity between explanatory variables
- A linear relationship exists between the logit of the response and the explanatory variables
- Large sample sizes are used, usually denoted by a minimum of 10 observations per variable

The response variable in this investigation is someone either dying or becoming severely injured whilst in a vehicle during a crash. As the dataset is aggregated at the vehicle-level, the response variable will be the number of occurrences of the event in the respective vehicle, weighted by the total number of occupants of said vehicle.

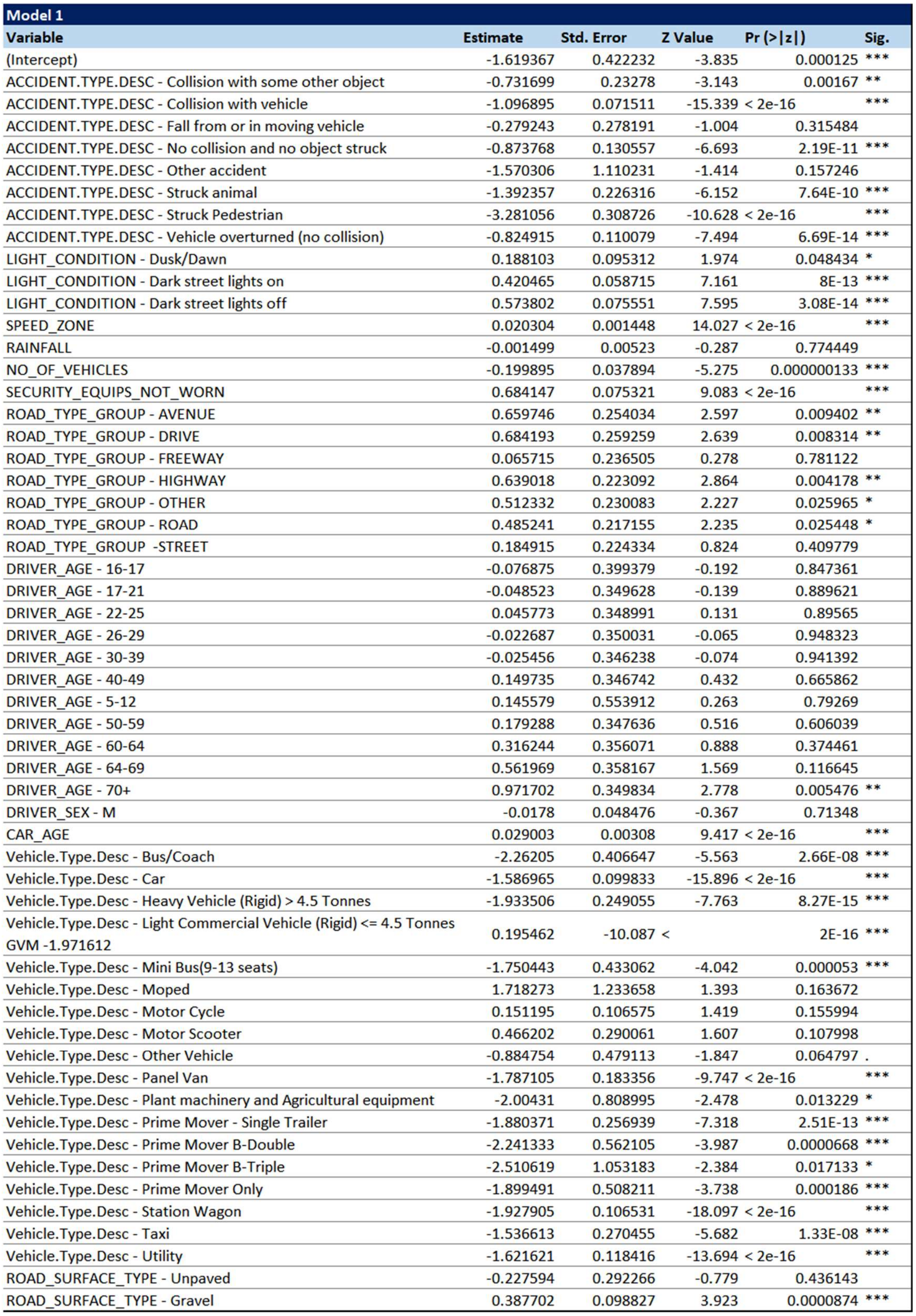
Chosen Explanatory Variables

The difficult job in regression modelling is selecting the appropriate number and choice of variables to construct the model. Working with intuition and the findings from EDA, we construct our first list of variables:



Explanatory Variable	**Type**	**Notes**	**Reference variable (if not determinable)**
Accident type description	Factor	Accident type	Collision with a fixed object
Light condition	Factor	Day, dusk/dawn,	Day
Speed zone	Numeric	Speed zone associated with crash	
Rainfall	Numeric	Millimetres of rainfall that day	
Number of vehicles	Numeric	Count of vehicles involved	
Count of security equipment missing	Numeric	Number of occupants without seatbelt or helmet.	
Road type	Factor	Road types identified as low-density areas were aggregated to 'Other'.	
Age of driver	Factor	Age category of driver	
Sex of driver	Factor	Driver's sex	Female
Age of car	Numeric	The year of the crash minus the year of car manufacture.	
Vehicle type desc	Factor	Vehicle type	Bicycle
Road surface type	Factor	Road surface categorised as paved, not paved and gravel.	Paved

Model 1 Results



Model 1

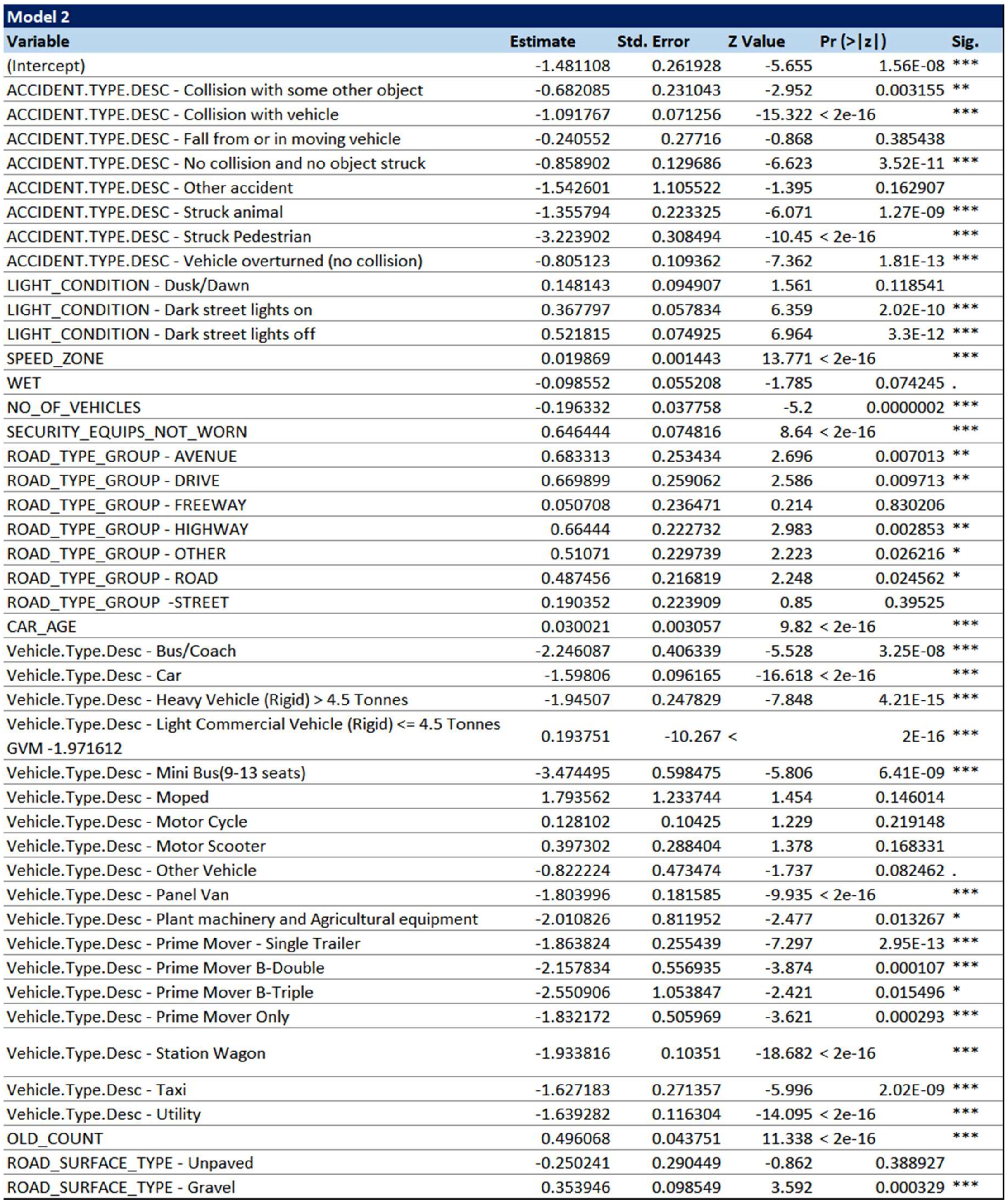
Variable	Estimate	Std. Error	Z Value	Pr (>|z|)	Sig.
(Intercept)	-1.619367	0.422232	-3.835	0.000125	***
ACCIDENT.TYPE.DESC - Collision with some other object	-0.731699	0.23278	-3.143	0.00167	**
ACCIDENT.TYPE.DESC - Collision with vehicle	-1.096895	0.071511	-15.339	< 2e-16	***
ACCIDENT.TYPE.DESC - Fall from or in moving vehicle	-0.279243	0.278191	-1.004	0.315484	
ACCIDENT.TYPE.DESC - No collision and no object struck	-0.873768	0.130557	-6.693	2.19E-11	***
ACCIDENT.TYPE.DESC - Other accident	-1.570306	1.110231	-1.414	0.157246	
ACCIDENT.TYPE.DESC - Struck animal	-1.392357	0.226316	-6.152	7.64E-10	***
ACCIDENT.TYPE.DESC - Struck Pedestrian	-3.281056	0.308726	-10.628	< 2e-16	***
ACCIDENT.TYPE.DESC - Vehicle overturned (no collision)	-0.824915	0.110079	-7.494	6.69E-14	***
LIGHT_CONDITION - Dusk/Dawn	0.188103	0.095312	1.974	0.048434	*
LIGHT_CONDITION - Dark street lights on	0.420465	0.058715	7.161	8E-13	***
LIGHT_CONDITION - Dark street lights off	0.573802	0.075551	7.595	3.08E-14	***
SPEED_ZONE	0.020304	0.001448	14.027	< 2e-16	***
RAINFALL	-0.001499	0.00523	-0.287	0.774449	
NO_OF_VEHICLES	-0.199895	0.037894	-5.275	0.000000133	***
SECURITY_EQUIPS_NOT_WORN	0.684147	0.075321	9.083	< 2e-16	***
ROAD_TYPE_GROUP - AVENUE	0.659746	0.254034	2.597	0.009402	**
ROAD_TYPE_GROUP - DRIVE	0.684193	0.259259	2.639	0.008314	**
ROAD_TYPE_GROUP - FREEWAY	0.065715	0.236505	0.278	0.781122	
ROAD_TYPE_GROUP - HIGHWAY	0.639018	0.223092	2.864	0.004178	**
ROAD_TYPE_GROUP - OTHER	0.512332	0.230083	2.227	0.025965	*
ROAD_TYPE_GROUP - ROAD	0.485241	0.217155	2.235	0.025448	*
ROAD_TYPE_GROUP -STREET	0.184915	0.224334	0.824	0.409779	
DRIVER_AGE - 16-17	-0.076875	0.399379	-0.192	0.847361	
DRIVER_AGE - 17-21	-0.048523	0.349628	-0.139	0.889621	
DRIVER_AGE - 22-25	0.045773	0.348991	0.131	0.89565	
DRIVER_AGE - 26-29	-0.022687	0.350031	-0.065	0.948323	
DRIVER_AGE - 30-39	-0.025456	0.346238	-0.074	0.941392	
DRIVER_AGE - 40-49	0.149735	0.346742	0.432	0.665862	
DRIVER_AGE - 5-12	0.145579	0.553912	0.263	0.79269	
DRIVER_AGE - 50-59	0.179288	0.347636	0.516	0.606039	
DRIVER_AGE - 60-64	0.316244	0.356071	0.888	0.374461	
DRIVER_AGE - 64-69	0.561969	0.358167	1.569	0.116645	
DRIVER_AGE - 70+	0.971702	0.349834	2.778	0.005476	**
DRIVER_SEX - M	-0.0178	0.048476	-0.367	0.71348	
CAR_AGE	0.029003	0.00308	9.417	< 2e-16	***
Vehicle.Type.Desc - Bus/Coach	-2.26205	0.406647	-5.563	2.66E-08	***
Vehicle.Type.Desc - Car	-1.586965	0.099833	-15.896	< 2e-16	***
Vehicle.Type.Desc - Heavy Vehicle (Rigid) > 4.5 Tonnes	-1.933506	0.249055	-7.763	8.27E-15	***
Vehicle.Type.Desc - Light Commercial Vehicle (Rigid) <= 4.5 Tonnes GVM -1.971612	0.195462	-10.087 <		2E-16	***
Vehicle.Type.Desc - Mini Bus(9-13 seats)	-1.750443	0.433062	-4.042	0.000053	***
Vehicle.Type.Desc - Moped	1.718273	1.233658	1.393	0.163672	
Vehicle.Type.Desc - Motor Cycle	0.151195	0.106575	1.419	0.155994	
Vehicle.Type.Desc - Motor Scooter	0.466202	0.290061	1.607	0.107998	
Vehicle.Type.Desc - Other Vehicle	-0.884754	0.479113	-1.847	0.064797	.
Vehicle.Type.Desc - Panel Van	-1.787105	0.183356	-9.747	< 2e-16	***
Vehicle.Type.Desc - Plant machinery and Agricultural equipment	-2.00431	0.808995	-2.478	0.013229	*
Vehicle.Type.Desc - Prime Mover - Single Trailer	-1.880371	0.256939	-7.318	2.51E-13	***
Vehicle.Type.Desc - Prime Mover B-Double	-2.241333	0.562105	-3.987	0.0000668	***
Vehicle.Type.Desc - Prime Mover B-Triple	-2.510619	1.053183	-2.384	0.017133	*
Vehicle.Type.Desc - Prime Mover Only	-1.899491	0.508211	-3.738	0.000186	***
Vehicle.Type.Desc - Station Wagon	-1.927905	0.106531	-18.097	< 2e-16	***
Vehicle.Type.Desc - Taxi	-1.536613	0.270455	-5.682	1.33E-08	***
Vehicle.Type.Desc - Utility	-1.621621	0.118416	-13.694	< 2e-16	***
ROAD_SURFACE_TYPE - Unpaved	-0.227594	0.292266	-0.779	0.436143	
ROAD_SURFACE_TYPE - Gravel	0.387702	0.098827	3.923	0.0000874	***

Model 1 Results - Discussion

Perhaps somewhat surprisingly, we can see that gender and age, with the exception of over 70 year olds, are all incredibly insignificant to the chance of a casualty. Given how biologically ill-equipped humans are at dealing with road crashes, as effectively highlighted in Towards Zero's 'Meet Graham' campaign(http://www.meetgraham.com.au/), we wouldn't necessarily expect men or women, nor young or middle-age to deviate significantly from a physical point of handling a crash. However, we might have expected a leaning towards male and young demographics to engage in more aggressive driving and incur more serious consequences. As this doesn't appear to be the case, we should look to remove these and construct a new variable to count the number of people over 70 years old.

Also notable is that the precise level of rainfall doesn't have a significant impact. This could be due to a number of reasons, perhaps that as rainfall goes up there are fewer people on the roads and that, in a general sense, people tend to drive more carefully during torrential conditions. Therefore, it might be a better idea to model using a binary variable set at a threshold of rainfall capturing whether the road is wet or not. As the United States Geological Survey states that slight rain is under 0.5mm an hour, we'll adopt a 1.5mm daily rain cut-off. (USGS, 2019)

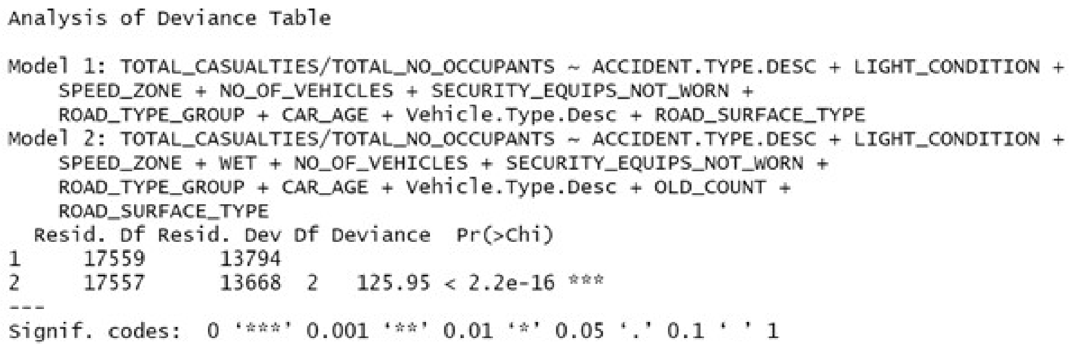
Model 2 Results



Model 2					
Variable	**Estimate**	**Std. Error**	**Z Value**	**Pr (>|z|)**	**Sig.**
(Intercept)	-1.481108	0.261928	-5.655	1.56E-08	***
ACCIDENT.TYPE.DESC - Collision with some other object	-0.682085	0.231043	-2.952	0.003155	**
ACCIDENT.TYPE.DESC - Collision with vehicle	-1.091767	0.071256	-15.322	< 2e-16	***
ACCIDENT.TYPE.DESC - Fall from or in moving vehicle	-0.240552	0.27716	-0.868	0.385438	
ACCIDENT.TYPE.DESC - No collision and no object struck	-0.858902	0.129686	-6.623	3.52E-11	***
ACCIDENT.TYPE.DESC - Other accident	-1.542601	1.105522	-1.395	0.162907	
ACCIDENT.TYPE.DESC - Struck animal	-1.355794	0.223325	-6.071	1.27E-09	***
ACCIDENT.TYPE.DESC - Struck Pedestrian	-3.223902	0.308494	-10.45	< 2e-16	***
ACCIDENT.TYPE.DESC - Vehicle overturned (no collision)	-0.805123	0.109362	-7.362	1.81E-13	***
LIGHT_CONDITION - Dusk/Dawn	0.148143	0.094907	1.561	0.118541	
LIGHT_CONDITION - Dark street lights on	0.367797	0.057834	6.359	2.02E-10	***
LIGHT_CONDITION - Dark street lights off	0.521815	0.074925	6.964	3.3E-12	***
SPEED_ZONE	0.019869	0.001443	13.771	< 2e-16	***
WET	-0.098552	0.055208	-1.785	0.074245	.
NO_OF_VEHICLES	-0.196332	0.037758	-5.2	0.0000002	***
SECURITY_EQUIPS_NOT_WORN	0.646444	0.074816	8.64	< 2e-16	***
ROAD_TYPE_GROUP - AVENUE	0.683313	0.253434	2.696	0.007013	**
ROAD_TYPE_GROUP - DRIVE	0.669899	0.259062	2.586	0.009713	**
ROAD_TYPE_GROUP - FREEWAY	0.050708	0.236471	0.214	0.830206	
ROAD_TYPE_GROUP - HIGHWAY	0.66444	0.222732	2.983	0.002853	**
ROAD_TYPE_GROUP - OTHER	0.51071	0.229739	2.223	0.026216	*
ROAD_TYPE_GROUP - ROAD	0.487456	0.216819	2.248	0.024562	*
ROAD_TYPE_GROUP -STREET	0.190352	0.223909	0.85	0.39525	
CAR_AGE	0.030021	0.003057	9.82	< 2e-16	***
Vehicle.Type.Desc - Bus/Coach	-2.246087	0.406339	-5.528	3.25E-08	***
Vehicle.Type.Desc - Car	-1.59806	0.096165	-16.618	< 2e-16	***
Vehicle.Type.Desc - Heavy Vehicle (Rigid) > 4.5 Tonnes	-1.94507	0.247829	-7.848	4.21E-15	***
Vehicle.Type.Desc - Light Commercial Vehicle (Rigid) <= 4.5 Tonnes GVM -1.971612	0.193751	-10.267 <		2E-16	***
Vehicle.Type.Desc - Mini Bus(9-13 seats)	-3.474495	0.598475	-5.806	6.41E-09	***
Vehicle.Type.Desc - Moped	1.793562	1.233744	1.454	0.146014	
Vehicle.Type.Desc - Motor Cycle	0.128102	0.10425	1.229	0.219148	
Vehicle.Type.Desc - Motor Scooter	0.397302	0.288404	1.378	0.168331	
Vehicle.Type.Desc - Other Vehicle	-0.822224	0.473474	-1.737	0.082462	.
Vehicle.Type.Desc - Panel Van	-1.803996	0.181585	-9.935	< 2e-16	***
Vehicle.Type.Desc - Plant machinery and Agricultural equipment	-2.010826	0.811952	-2.477	0.013267	*
Vehicle.Type.Desc - Prime Mover - Single Trailer	-1.863824	0.255439	-7.297	2.95E-13	***
Vehicle.Type.Desc - Prime Mover B-Double	-2.157834	0.556935	-3.874	0.000107	***
Vehicle.Type.Desc - Prime Mover B-Triple	-2.550906	1.053847	-2.421	0.015496	*
Vehicle.Type.Desc - Prime Mover Only	-1.832172	0.505969	-3.621	0.000293	***
Vehicle.Type.Desc - Station Wagon	-1.933816	0.10351	-18.682	< 2e-16	***
Vehicle.Type.Desc - Taxi	-1.627183	0.271357	-5.996	2.02E-09	***
Vehicle.Type.Desc - Utility	-1.639282	0.116304	-14.095	< 2e-16	***
OLD_COUNT	0.496068	0.043751	11.338	< 2e-16	***
ROAD_SURFACE_TYPE - Unpaved	-0.250241	0.290449	-0.862	0.388927	
ROAD_SURFACE_TYPE - Gravel	0.353946	0.098549	3.592	0.000329	***

Model 2 Results - Interpretation

This result is an improvement; the number of over 70 year olds is incredibly significant and the wet variable, while not quite at the desired level of 5% significance, isn’t too far off and holds more predictive power than our rainfall variable did.

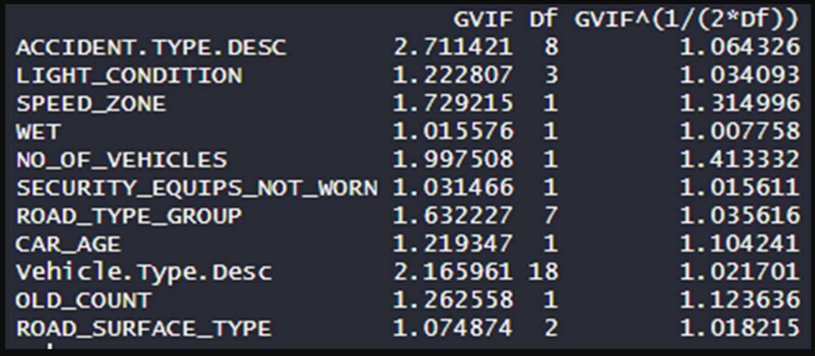


```
Analysis of Deviance Table

Model 1: TOTAL_CASUALTIES/TOTAL_NO_OCCUPANTS ~ ACCIDENT.TYPE.DESC + LIGHT_CONDITION +
    SPEED_ZONE + NO_OF_VEHICLES + SECURITY_EQUIPS_NOT_WORN +
    ROAD_TYPE_GROUP + CAR_AGE + Vehicle.Type.Desc + ROAD_SURFACE_TYPE
Model 2: TOTAL_CASUALTIES/TOTAL_NO_OCCUPANTS ~ ACCIDENT.TYPE.DESC + LIGHT_CONDITION +
    SPEED_ZONE + WET + NO_OF_VEHICLES + SECURITY_EQUIPS_NOT_WORN +
    ROAD_TYPE_GROUP + CAR_AGE + Vehicle.Type.Desc + OLD_COUNT +
    ROAD_SURFACE_TYPE
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     17559      13794
2     17557      13668  2   125.95 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Comparing this model to the initial model with insignificant variables removed, we can see a statistically significant improvement. Additionally, we receive confirmation that we’re attaining the highest possible AIC given our set of variables with a stepwise regression returning the same model. A number of other variables were tested in the model, however, by the nature of variable selection being a lengthy and difficult process, further analysis here is curtailed in the interest of brevity.

Testing for Multicollinearity

In order to ratify one of the key assumptions of our model, we delve into measuring multicollinearity, Multicollinearity is the situation where a significant correlation exists between explanatory variables and its minimal occurrence is one of our assumptions. Significant multicollinearity is generally considered to produce variance inflation factors (VIF) of between 5 to 10 (STHDA, 2019). As can be seen, our explanatory variables are easily clear of this threshold.



```
                             GVIF Df GVIF^(1/(2*Df))
ACCIDENT.TYPE.DESC       2.711421  8        1.064326
LIGHT_CONDITION          1.222807  3        1.034093
SPEED_ZONE               1.729215  1        1.314996
WET                      1.015576  1        1.007758
NO_OF_VEHICLES           1.997508  1        1.413332
SECURITY_EQUIPS_NOT_WORN 1.031466  1        1.015611
ROAD_TYPE_GROUP          1.632227  7        1.035616
CAR_AGE                  1.219347  1        1.104241
Vehicle.Type.Desc        2.165961 18        1.021701
OLD_COUNT                1.262558  1        1.123636
ROAD_SURFACE_TYPE        1.074874  2        1.018215
```

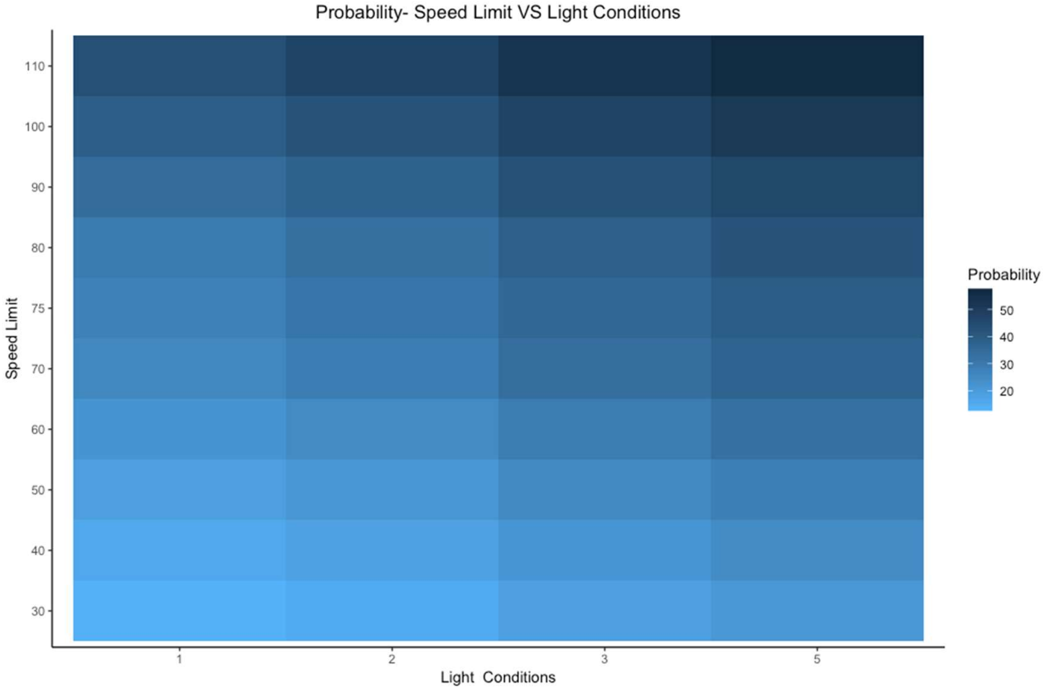
Interpretation and Findings

The coefficients of a logistic regression model must be interpreted in the vyse of logit nature of the model, requiring it to be exponentiated to uncover the chance it has against the reference. Similarly, categorical variables should be interpreted in relation to the reference variable from which their coefficient relates (included in the variable table).

We can clearly observe heavily significant relationships existing in each variable, with the exception of wet weather. Wet weather, too, carries a negative relationship with the chance of a casualty, suggesting the central hypothesis of this investigation is incorrect.

However, a number of relationships we'd expect to hold true are reinforced, with speed zone, missing seat belts and old age heavy indicators of adverse outcomes. Assuming a set of standard conditions, we can visually demonstrate the likelihood.

This is consistent with Towards Zero's current focus on safer speeds, which this investigation thus echoes.



6. Limitations

Only crash-data available

The rich road crash data set employed in this investigation unfortunately doesn’t extend to journeys where a crash didn’t occur. This limits the number of research questions which can effectively be answered and in some cases curbs the strength of predictions made for those that can.

It should be addressed, though, that traffic data sets incorporated in the future will face significant challenges, due to their multi-temporal and multi-locational properties.

Limited to crashes with injury

Only those crashes which resulted in at least one participant receiving a minor injury or worse were recorded, which is logically the case given that authorities of the state were the collectors of this data. This means that our data set is skewed towards more dangerous accidents.

Rainfall Estimations

Rainfall figures, as previously described, are estimated through the use of weather station readings across the state. Inevitably, these estimates will often deviate with the true rainfall incurred at that location during that period (Beesley, Frost & Zajaczkowski 2009).

Furthermore, the rainfall data is provided on a daily basis, without information about time of day.

Confounding variables

With the inclusion of multiple explanatory variables comes the potential that any number of variables could be obfuscating relationships that other independent variables hold with the response variable.

7. Privacy and Ethics

All data sets used in this research are publically available online and are issued by Government authorities, hence are addressed by the data privacy policies. Additionally, individuals recorded in the accident dataset are not identified, and for those who the victims may be determinable (given additional knowledge surrounding the crash), no personally sensitive information is revealed.

The research activities which are sought to be performed are in the interest of saving lives and the general public benefit.

8. Conclusions and further questions

This investigation sought to uncover relationships between a plethora of variables, namely rainfall, to the occurrence of death and severe injuries in car crashes. Ultimately, it was established that the nominal level of rainfall does not bear a statistically significant relationship with this outcome, while the binary categorisation holds weak significance. The latter's relationship is negative, leading us to conclude that Towards Zero should not drastically alter their current campaigns.

The deeply rich data sets afforded to this topic of investigation beg the opportunity to dig into an enormous number of further lines of inquiry, including but not limited to the role of other weather phenomena and the prediction of crash frequency. Predictive applications of this model, too, is an area ripe for further investigation.

9. References

- Beesley, C.A., Frost, A.J. & Zajaczkowski, J. 2009 ‘A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets’, paper presented to the 18th World IMACS / MODSIM Congress, Cairns, 13-17 July.
- Jeffrey, S.j., Carter, J.O, Moodie, K.B & Beswick, A.R 2001 ‘Using spatial interpolation to construct a comprehensive archive of Australian climate data’, *Environmental Modelling & Software*, vol. 16, pp.309-330.
- Bureau of Infrastructure, Transport and Regional Economics 2018, *International road safety comparisons 2016*, BITRE, Canberra.
- Australian Bureau of Statistics 2019, *9309.0 - Motor vehicle census, Australia, 31 Jan 2019*, ABS, viewed 3 October 2019 <https://www.abs.gov.au/ausstats/abs@.nsf/mf/9309.0>.
- Australian Bureau of Statistics 2018, *3101.0 - Australian Demographic Statistics, Dec 2018*, ABS, viewed 3 October 2019 <https://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/1988DE98D5424933CA258479001A75A5?opendocument>.
- STHDA 2019, <http://www.sthda.com/english/articles/36-classification-methods-essentials/148-logistic-regression-assumptions-and-diagnostics-in-r/>
- United States Geological Survey 2019, <https://water.usgs.gov/edu/activity-howmuchrain-metric.html>
- Towards Zero 2019 , *Lives Lost - Annual*, Towards Zero, viewed 3 October 2019 <http://www.tac.vic.gov.au/road-safety/statistics/lives-lost-annual>.
- Towards Zero 2019, *What is Towards Zero*, Towards Zero, viewed 3 October 2019 <https://www.towardszero.vic.gov.au/what-is-towards-zero/road-safety-action-plan>.
- VicRoads 2019, *Licences,* VicRoads, viewed 4 October 2019, <https://www.vicroads.vic.gov.au/licences>.
- Baguley, C. (2019). The importance of a road accident data system and its utilisation.. [online] Pdfs.semanticscholar.org. Available at: https://pdfs.semanticscholar.org/2d8e/c35ac2bd46cc41934d84e1ee2024da6dbec1.pdf [Accessed 1 Sep. 2019].
- Goodwin, L. (2019). Weather Impacts on Arterial Traffic Flow. [online] Citeseerx.ist.psu.edu. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.587.9563&rep=rep1&type=pdf [Accessed 1 Sep. 2019].

Appendix

A Data Sets & Merging

VEHICLE.CSV
ACCIDENT_NO
VEHICLE_ID
VEHICLE_YEAR_MANUF
VEHICLE_DCA_CODE
INITIAL_DIRECTION
ROAD_SURFACE_TYPE
Road Surface Type Desc
REG_STATE
VEHICLE_BODY_STYLE
VEHICLE_MAKE
VEHICLE_MODEL
VEHICLE_POWER
VEHICLE_TYPE
Vehicle Type Desc
VEHICLE_WEIGHT
CONSTRUCTION_TYPE
FUEL_TYPE
NO_OF_WHEELS
NO_OF_CYLINDERS
SEATING_CAPACITY
TARE_WEIGHT
TOTAL_NO_OCCUPANTS
CARRY_CAPACITY
CUBIC_CAPACITY
FINAL_DIRECTION
DRIVER_INTENT
VEHICLE_MOVEMENT
TRAILER_TYPE
VEHICLE_COLOUR_1
VEHICLE_COLOUR_2
CAUGHT_FIRE
INITIAL_IMPACT
LAMPS
LEVEL_OF_DAMAGE
OWNER_POSTCODE
TOWED_AWAY_FLAG
TRAFFIC_CONTROL
Traffic Control Desc

ATMOSPHERIC_COND.CSV
ACCIDENT_NO
ATMOSPH_COND
ATMOSPH_COND_SEQ
Atmosph Cond Desc

NODE.CSV
ACCIDENT_NO
NODE_ID
NODE_TYPE
AMG_X
AMG_Y
LGA_NAME
Lga Name All
Region Name
Deg Urban Name
Lat
Long
Postcode No

PERSON.CSV
ACCIDENT_NO
PERSON_ID
VEHICLE_ID
SEX
AGE
Age Group
INJ_LEVEL
Inj Level Desc
SEATING_POSITION
HELMET_BELT_WORN
ROAD_USER_TYPE
Road User Type Desc
LICENCE_STATE
PEDEST_MOVEMENT
POSTCODE
TAKEN_HOSPITAL
EJECTED_CODE

ACCIDENT_EVENT.CSV
ACCIDENT_NO
EVENT_SEQ_NO
EVENT_TYPE
Event Type Desc
VEHICLE_1_ID
VEHICLE_1_COLL_PT
Vehicle 1 Coll Pt Desc
VEHICLE_2_ID
VEHICLE_2_COLL_PT
Vehicle 2 Coll Pt Desc
PERSON_ID
OBJECT_TYPE
Object Type Desc

ACCIDENT_LOCATION.CSV
ACCIDENT_NO
NODE_ID
ROAD_ROUTE_1
ROAD_NAME
ROAD_TYPE
ROAD_NAME_INT
ROAD_TYPE_INT
DISTANCE_LOCATION
DIRECTION_LOCATION
NEAREST_KM_POST
OFF_ROAD_LOCATION

ROAD_SURFACE_COND.CSV
ACCIDENT_NO
SURFACE_COND
Surface Cond Desc
SURFACE_COND_SEQ

RAINFALL
Date
Lat
Long
Rainfall (mm)

ACCIDENT.CSV
ACCIDENT_NO
ACCIDENTDATE
ACCIDENTTIME
ACCIDENT_TYPE
Accident Type Desc
DAY_OF_WEEK
Day Week Description
DCA_CODE
DCA Description
DIRECTORY
EDITION
PAGE
GRID_REFERENCE_X
GRID_REFERENCE_Y
LIGHT_CONDITION
Light Condition Desc
NODE_ID
NO_OF_VEHICLES
NO_PERSONS
NO_PERSONS_INJ_2
NO_PERSONS_INJ_3
NO_PERSONS_KILLED
NO_PERSONS_NOT_INJ
POLICE_ATTEND
ROAD_GEOMETRY
Road Geometry Desc
SEVERITY
SPEED_ZONE

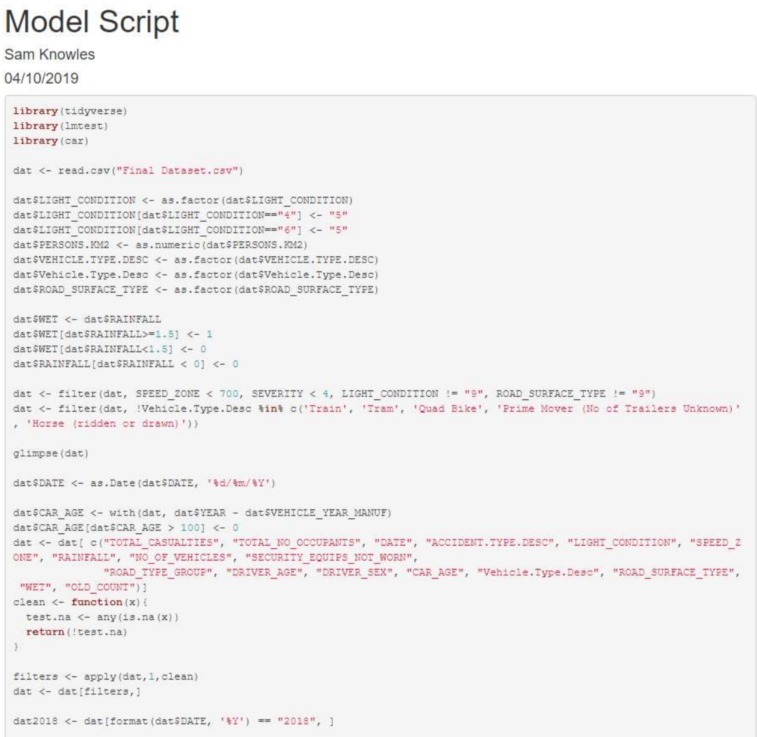
ACCIDENT_CHAINAGE.CSV
Node Id
Route No
Chainage Seq
Route Link No
Chainage

SUB_DCA.CSV
ACCIDENT_NO
SUB_DCA_CODE
SUB_DCA_SEQ
Sub Dca Code Desc

VICTORIA_POPULATION
Code
Label
Year
Estimated Resident Population - Total
Population Density

Appendix

B Scripts

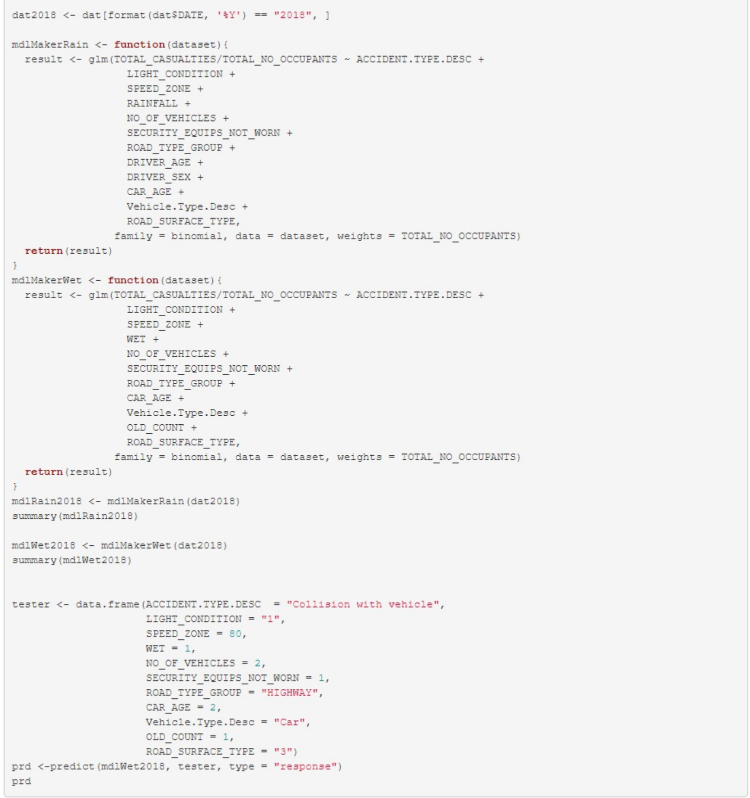


Model Script

Sam Knowles

04/10/2019

```
library(tidyverse)
library(lmtest)
library(car)

dat <- read.csv("Final Dataset.csv")

dat$LIGHT_CONDITION <- as.factor(dat$LIGHT_CONDITION)
dat$LIGHT_CONDITION[dat$LIGHT_CONDITION=="4"] <- "5"
dat$LIGHT_CONDITION[dat$LIGHT_CONDITION=="6"] <- "5"
dat$PERSONS.KM2 <- as.numeric(dat$PERSONS.KM2)
dat$VEHICLE.TYPE.DESC <- as.factor(dat$VEHICLE.TYPE.DESC)
dat$Vehicle.Type.Desc <- as.factor(dat$Vehicle.Type.Desc)
dat$ROAD_SURFACE_TYPE <- as.factor(dat$ROAD_SURFACE_TYPE)

dat$WET <- dat$RAINFALL
dat$WET[dat$RAINFALL>=1.5] <- 1
dat$WET[dat$RAINFALL<1.5] <- 0
dat$RAINFALL[dat$RAINFALL < 0] <- 0

dat <- filter(dat, SPEED_ZONE < 700, SEVERITY < 4, LIGHT_CONDITION != "9", ROAD_SURFACE_TYPE != "9")
dat <- filter(dat, !Vehicle.Type.Desc %in% c('Train', 'Tram', 'Quad Bike', 'Prime Mover (No of Trailers Unknown)'
, 'Horse (ridden or drawn)'))

glimpse(dat)

dat$DATE <- as.Date(dat$DATE, '%d/%m/%Y')

dat$CAR_AGE <- with(dat, dat$YEAR - dat$VEHICLE_YEAR_MANUF)
dat$CAR_AGE[dat$CAR_AGE > 100] <- 0
dat <- dat[ c("TOTAL_CASUALTIES", "TOTAL_NO_OCCUPANTS", "DATE", "ACCIDENT.TYPE.DESC", "LIGHT_CONDITION", "SPEED_Z
ONE", "RAINFALL", "NO_OF_VEHICLES", "SECURITY_EQUIPS_NOT_WORN",
              "ROAD_TYPE_GROUP", "DRIVER_AGE", "DRIVER_SEX", "CAR_AGE", "Vehicle.Type.Desc", "ROAD_SURFACE_TYPE",
 "WET", "OLD_COUNT")]
clean <- function(x){
  test.na <- any(is.na(x))
  return(!test.na)
}

filters <- apply(dat,1,clean)
dat <- dat[filters,]

dat2018 <- dat[format(dat$DATE, '%Y') == "2018", ]
```



```
dat2018 <- dat[format(dat$DATE, '%Y') == "2018", ]

mdlMakerRain <- function(dataset){
  result <- glm(TOTAL_CASUALTIES/TOTAL_NO_OCCUPANTS ~ ACCIDENT.TYPE.DESC +
                  LIGHT_CONDITION +
                  SPEED_ZONE +
                  RAINFALL +
                  NO_OF_VEHICLES +
                  SECURITY_EQUIPS_NOT_WORN +
                  ROAD_TYPE_GROUP +
                  DRIVER_AGE +
                  DRIVER_SEX +
                  CAR_AGE +
                  Vehicle.Type.Desc +
                  ROAD_SURFACE_TYPE,
                family = binomial, data = dataset, weights = TOTAL_NO_OCCUPANTS)
  return(result)
}
mdlMakerWet <- function(dataset){
  result <- glm(TOTAL_CASUALTIES/TOTAL_NO_OCCUPANTS ~ ACCIDENT.TYPE.DESC +
                  LIGHT_CONDITION +
                  SPEED_ZONE +
                  WET +
                  NO_OF_VEHICLES +
                  SECURITY_EQUIPS_NOT_WORN +
                  ROAD_TYPE_GROUP +
                  CAR_AGE +
                  Vehicle.Type.Desc +
                  OLD_COUNT +
                  ROAD_SURFACE_TYPE,
                family = binomial, data = dataset, weights = TOTAL_NO_OCCUPANTS)
  return(result)
}
mdlRain2018 <- mdlMakerRain(dat2018)
summary(mdlRain2018)

mdlWet2018 <- mdlMakerWet(dat2018)
summary(mdlWet2018)


tester <- data.frame(ACCIDENT.TYPE.DESC  = "Collision with vehicle",
                     LIGHT_CONDITION = "1",
                     SPEED_ZONE = 80,
                     WET = 1,
                     NO_OF_VEHICLES = 2,
                     SECURITY_EQUIPS_NOT_WORN = 1,
                     ROAD_TYPE_GROUP = "HIGHWAY",
                     CAR_AGE = 2,
                     Vehicle.Type.Desc = "Car",
                     OLD_COUNT = 1,
                     ROAD_SURFACE_TYPE = "3")
prd <-predict(mdlWet2018, tester, type = "response")
prd
```

Further scripts available upon request

Appendix

C Additional Resources

Interactive Road Crash Map:

http://rpubs.com/ganesharun237/534975