



# Motor Vehicle Accidents in Victoria

**Ibrahim Al-Hindi**  
Master of Business Analytics

**Arek Chouzadjian**  
Master of Business Analytics

**Kaihao Chen**  
Master of Business Analytics

Report for  
VicRoads

**Department of Econometrics  
and Business Statistics**

📞 (03) 9905 2478  
✉️ [questions@company.com](mailto:questions@company.com)

ABN: 12 377 614 630

**29 May 2021**

## Introduction

Road safety is a major public policy issue in the state of Victoria. Every year, there are thousands of accidents on the state's road network, in which hundreds of people are killed or injured. The aim of this report is to explore the trends and patterns in Victoria's road safety data, identifying which factors are associated with a higher risk of accident or death, and seeking to explain those relationships.

## Data

The data used in the report was obtained from VicRoads, and contains comprehensive information on approximately 200,000 accidents that occurred in the state from 2006 to 2020. The dataset is licensed with Creative Commons Attribution 4.0 International.

## Research questions

The research questions are as follows:

1. The impact of temporal factors, such as year, weekday and hour, on the number of accidents.
2. The relationship of speed and the age of vehicles with the death rate from accidents.
3. The effect of age and gender on accident numbers, as well as which roads in Victoria are most accident-prone and most deadly.

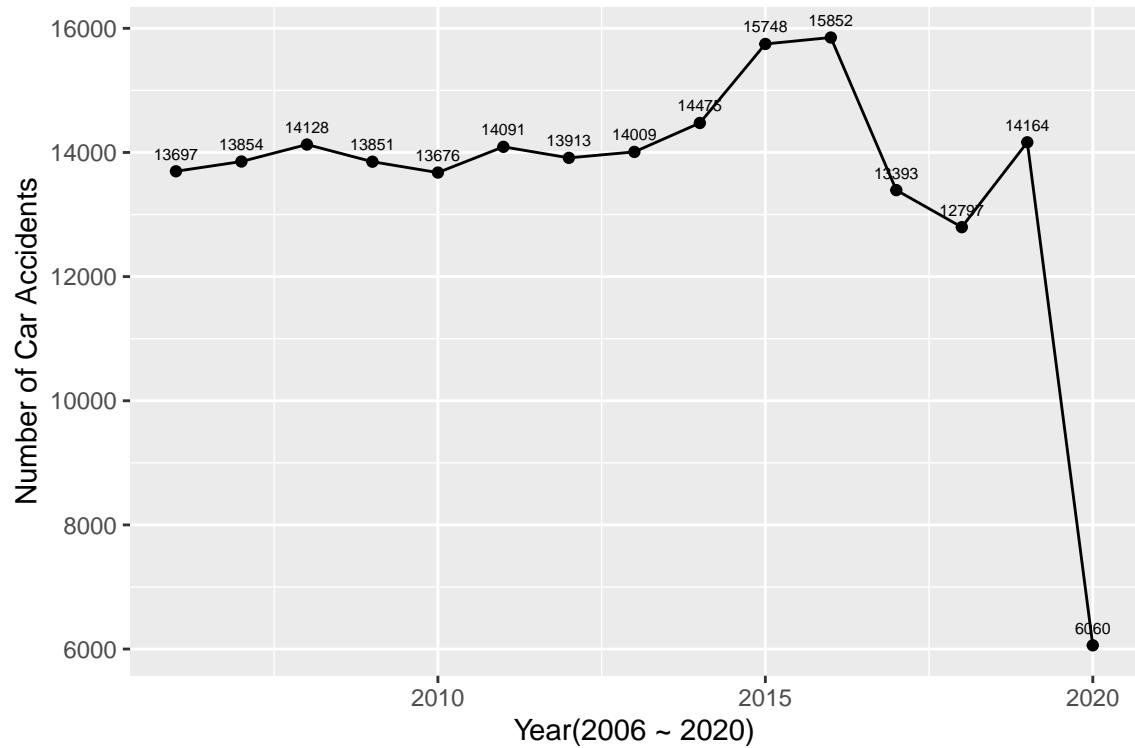
## Temporal factors analysis

### Accidents per year

According to plot 1, the increasing number of car accidents remained relatively stable from 2006 to 2014, around 14000. After 2014, the increasing speed became faster and then reached the first peak in 2015, second peak in 2016: 15852. After 2016, it started to drop. There is one outstanding change between 2019 and 2020, which it plummeted down from 14164 to 6060. The possible reason is the coming of covid 19 pandemic and the lockdown of Victoria, which made less car on roads, less accidents happened.

### Accidents by weekday

Regarding to plot 2, it indicates that there is a stable increasing number of car accidents from Sunday to Friday then reach the highest number on Friday. We could understand it from the following reasons:



**Figure 1:** *Car Accidents per year*

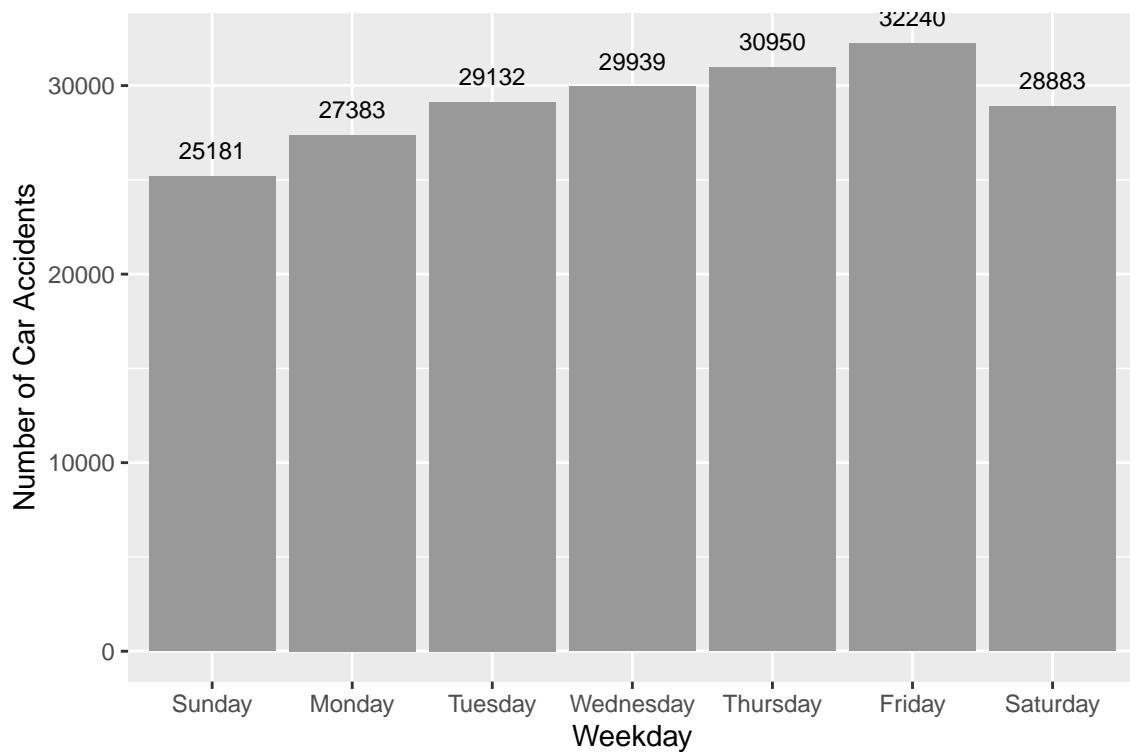
**Table 1:** *Number of Car Accidents happened by weekday*

Weekday	Accidents
Sunday	25181
Monday	27383
Tuesday	29132
Wednesday	29939
Thursday	30950
Friday	32240
Saturday	28883

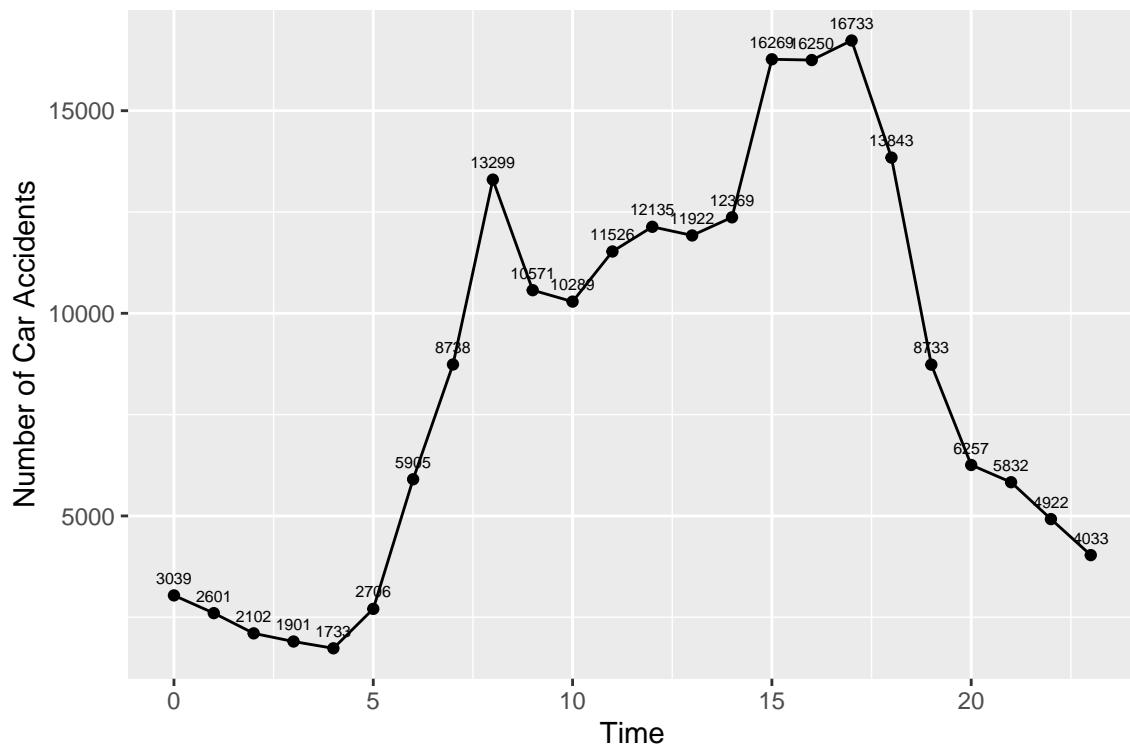
people are getting more and more exhausted during the whole working week, and many people will choose to hang out on Friday night which increase the percentage of driving drunk or reckless.

### Accidents by hour and Death Rate by hour

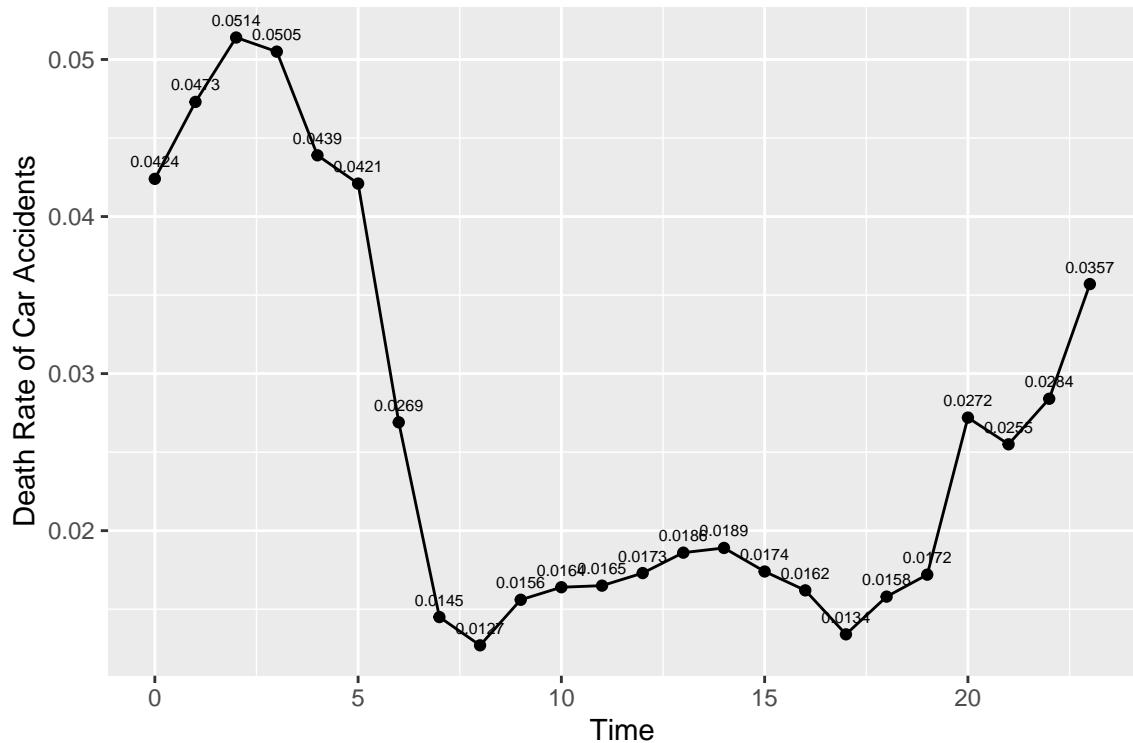
When comparing plot 3 and plot 4, we could notice that these two trend are indicating opposite story, which is the higher number of accidents actually with lower death rate during the same specific time. For example, 8 o'clock in the morning reached the first peak of car accidents. However, the death rate of it was the lowest. And the point of 5 o'clock tells a similar story. The 4 actually shows although 2 o'clock in the midnight has almost the lowest volume of car accidents, it has the highest death rate.



**Figure 2:** Car Accidents by weekday



**Figure 3:** Car Accidents by hour



**Figure 4: Death Rate by hour**

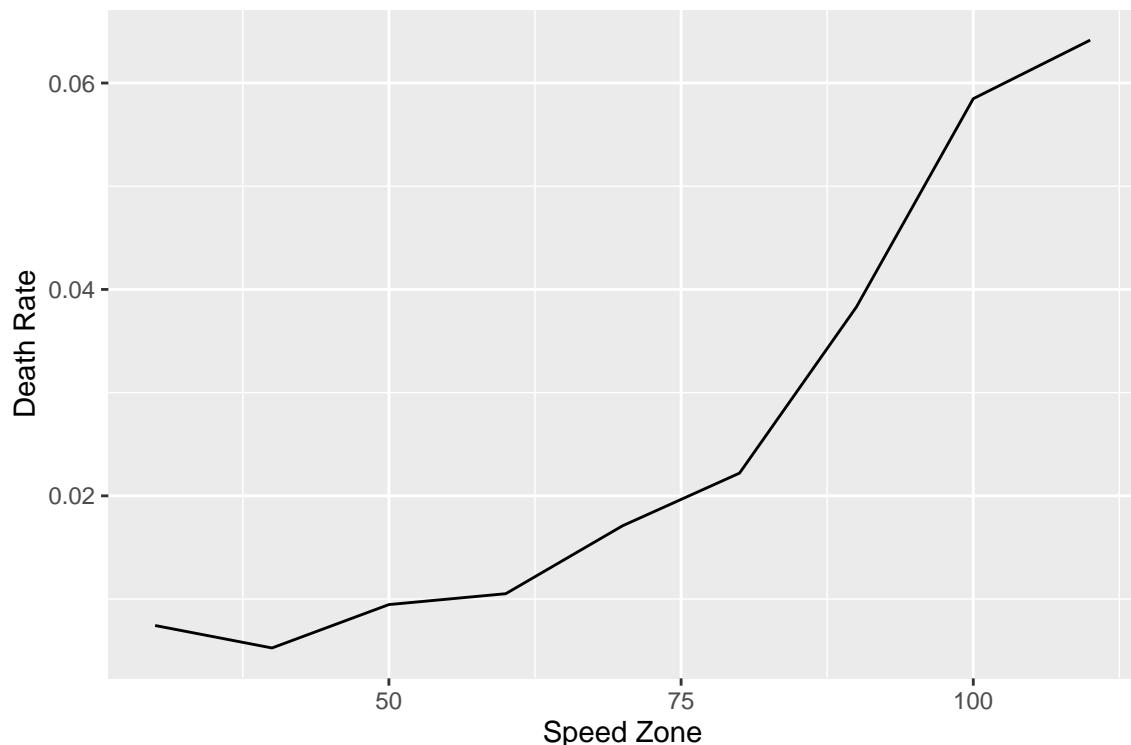
The possible reasons for above results are: both 8 o'clock in the morning and 5 o'clock in the afternoon are the commuter time, which make the traffic busier and more cars on roads, so more accidents. However, most of the accidents won't be too server due to the packed traffic. The 2 o'clock in the midnight is different. First, drivers would be more sleepy and more drivers would take the risk of drunk driving after attending parties in the night which make the accidents have a higher death rate.

### Effect of speed and vehicle age on death rate

**Table 2: Death Rate per Speed Zone**

Speed Zone	Accidents	Deaths	Death Rate
030	269	2	0.0074349
040	8937	47	0.0052590
050	36149	342	0.0094608
060	69133	727	0.0105160
070	15145	259	0.0171014
075	62	2	0.0322581
080	27794	617	0.0221990
090	940	36	0.0382979
100	31240	1827	0.0584827
110	2151	138	0.0641562
777	249	1	0.0040161
888	930	6	0.0064516
999	10709	30	0.0028014

### Death rate by speed zone



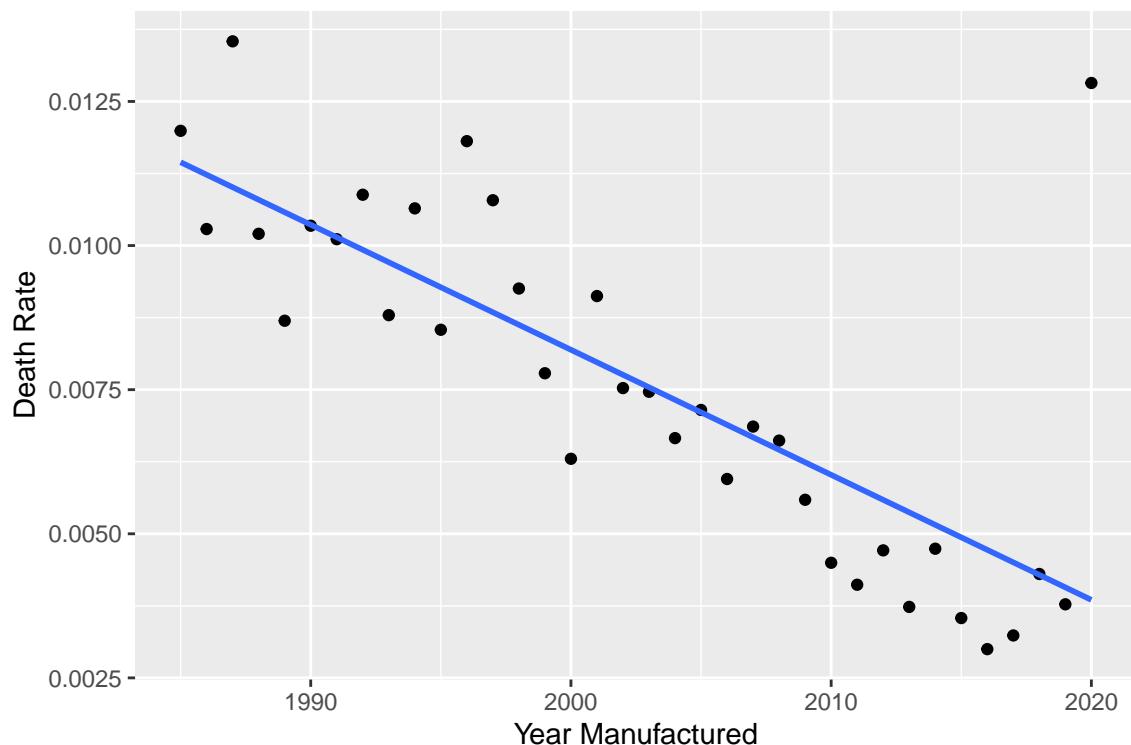
For this question, the variable Death Rate is defined as the number of deaths in a given speed zone over the time period analysed, divided by the number of accidents. All of these figures can be seen in the table.

The analysis for this question works on the reasonable assumption that accidents that occur in higher speed zones occur at higher speeds. The analysis shows that there is a very strong association between

the level of speed permitted in a certain zone, and the likelihood of dying in an accident in that zone. As the level of speed permitted increases, the probability of dying in an accident rises sharply. The death rate from accidents in 40km/h zones is around 0.005, whereas in 110km/h zones, the death rate is around 0.064; that is, you are nearly 13 times more likely to be killed in an accident in a 110km/h zone, as opposed to a 40km/h zone.

The reason for this difference in the death rate is fairly obvious; higher speeds contribute greatly to the severity of accidents. A person's car may be lightly to moderately damaged in a low-speed collision, but is much more likely to suffer massive damage in a high speed collision. This in turn drastically increases the risk of serious injury or death for the occupants.

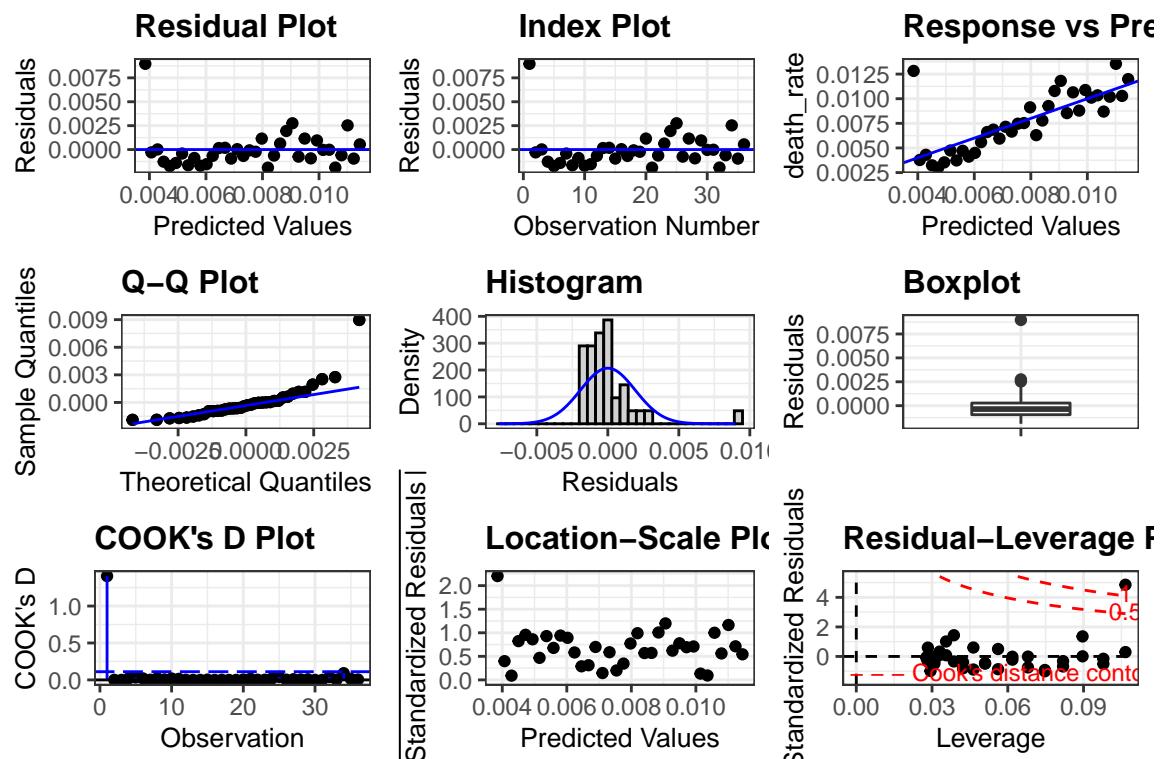
### Death rate by year of vehicle manufacture



term	estimate	std.error	statistic	p.value
(Intercept)	0.4420226	0.0627854	7.040211	0e+00
vehicle_year_manuf	-0.0002169	0.0000314	-6.918470	1e-07

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	dev
0.5846833	0.5724681	0.0019542	47.86523	1e-07	1	174.5064	-343.0127	-338.2622	0.000

### Regression model residual panel



### Goodness of fit tables

For this question, the variable Death Rate is defined as the number of deaths in accidents involving cars from a particular year of manufacture, divided by the number of accidents.

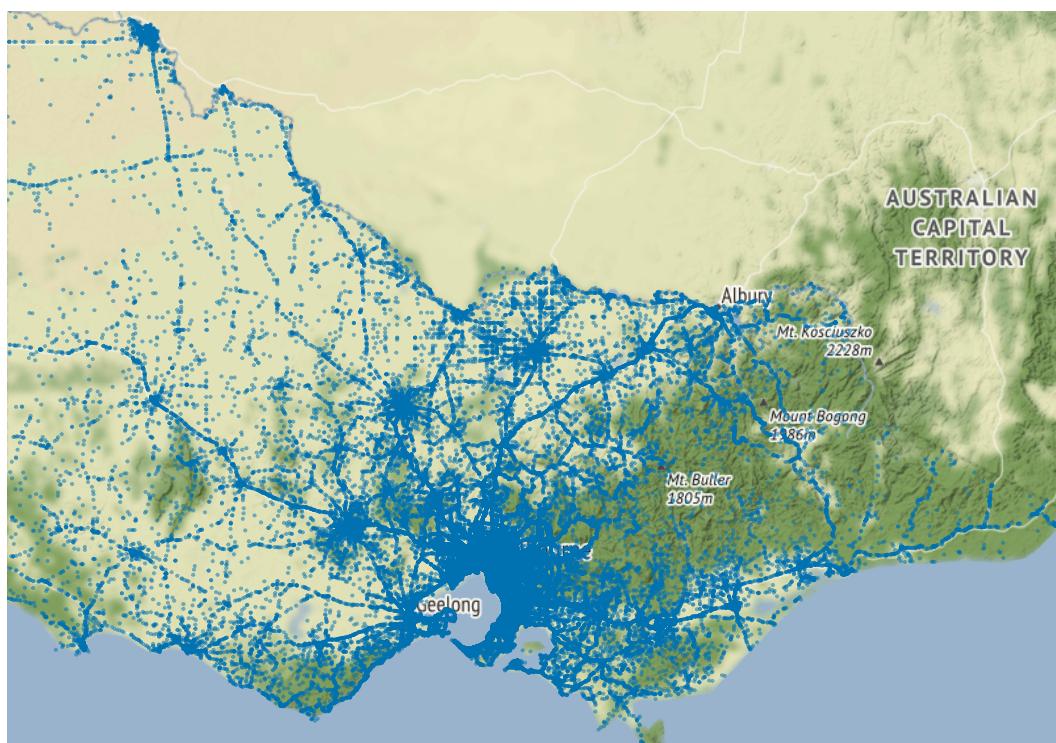
The analysis of the death rate by year of vehicle manufacture shows that there is a relatively strong negative correlation between the recentness of the year of manufacture of a vehicle, and the likelihood that a vehicle manufactured in that year will be involved in a fatal accident. The graph shows that a person is more than twice as likely to be killed in an accident if they are in a car manufactured in the late 1980s, as opposed to a car manufactured in the last 5 years.

In the linear model which was fitted to the data, for every year older a car is, the death rate increases by around 0.0002 deaths per accident. This linear model fits the data quite well; the R-squared is around 0.6, and the residuals are fairly evenly spaced around 0.

The reason for the decline in the death rate associated with more recently manufactured vehicles is improved safety standards. Cars built today contain far more structural features designed to protect occupants in the event of an accident. They also possess better braking capability, as well as extensive electronic systems that warn drivers of impending hazards.

## Accidents by Locations, Gender and Road User Type

### Accidents Map



**Figure 5:** Map of accident locations in Victoria

As shown in figure 5, accidents are most highly concentrated around metropolitan Melbourne, and gradually reduce in volume the further we move from Melbourne, with pockets of concentrations in the regional cities such as Bendigo, Ballarat and Geelong. This is due to the population being most present in metropolitan Melbourne, resulting in more accidents, and the population declining as we drift away, resulting in less accidents.

### Roads with Most Accidents and Highest Death Rates

According to table 3, Princes Highway, High Street and Nepean Highway are the three most accident prone roads in Victoria with 3581, 3096, 2376 accidents, respectively. It is important to note that

**Table 3:** Accidents by road

Road	Accidents
PRINCES HIGHWAY	3581
HIGH STREET	3096
NEPEAN HIGHWAY	2376
SPRINGVALE ROAD	1663
SOUTH GIPPSLAND HIGHWAY	1538
SYDNEY ROAD	1538
MONASH FREEWAY	1533
MAROONDAH HIGHWAY	1335

**Table 4:** Deadliest roads

Road	Accidents	Deaths	Deaths_per_accident
GLENELG HIGHWAY	231	28	0.1212121
GOULBURN VALLEY HIGHWAY	344	39	0.1133721
WIMMERA HIGHWAY	156	17	0.1089744
MURRAY VALLEY HIGHWAY	727	76	0.1045392
STRZELECKI HIGHWAY	106	11	0.1037736
HAMILTON HIGHWAY	223	21	0.0941704
MELBA HIGHWAY	211	18	0.0853081
NORTHERN HIGHWAY	300	25	0.0833333

roads with the name “High Street” are quite common and are present in various suburbs in Victoria, therefore the “High Street” displayed in the table is likely a combination of all the accidents that occurred in all the High Streets. When examining the deadliest roads from table 4, it is immediately apparent that **Highways** are the deadliest type of road in Victoria, this is likely a result of highways being locations of higher speed zones, which as we have seen from the previous section, lead to higher death rates.

### Accidents by Gender

Table 5 shows that there are more accidents committed by males (171,043) than by females (118,307). Figure 6 extends that differential by showing that males commit more accidents than females at all age groups. This could be a result of larger male presence on the roads than females, for example the majority of truck drivers and taxi/uber drivers are male, therefore representing higher numbers

**Table 5:** Number of accidents by gender

Sex	Accidents
Male	171043
Female	118307
Unknown	8831



**Figure 6:** Number of accidents by gender and age

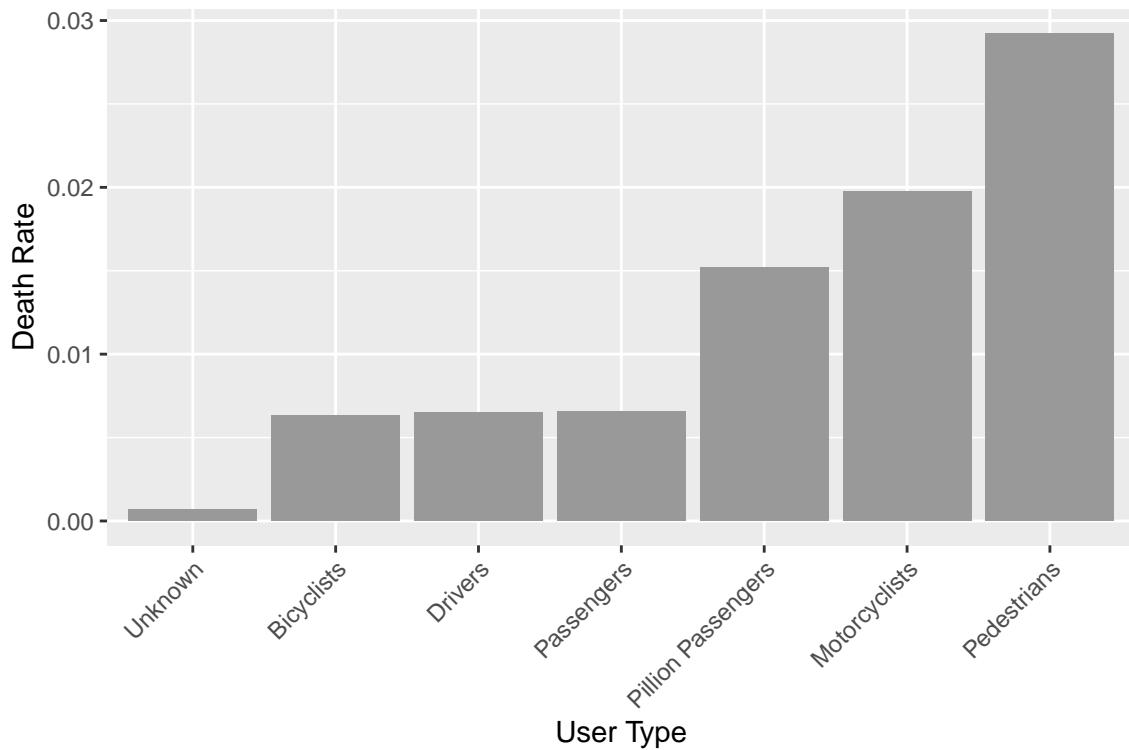
and longer times spent on the road. What is common between both genders, however, is that the accident numbers are highest for young and inexperienced drivers before steadily declining as age and experience increase. This is consistent with the findings from Gislason et al. (1997).

### User Type Death Rates

As per figure 7, **pedestrians** are at the most risk of death per accident, this is expected as pedestrians have no protection at all. **Motorcyclists** and **pillion passengers** (motorcycle passengers) occupy the second and third highest death rate per accident. It is surprising, however, that **bicyclists**'s death rate is similar to that of car drivers and passengers, as one would expect that bicyclists would have a death rate similar to that of pedestrians or motorcyclists, due to the lack of protection besides a helmet.

### Conclusions

The analysis produced several pieces of informative results. First, it was found that accidents increase gradually throughout the working week, and that whilst accidents are most common in evening peak hour, that period is the least deadly time of day in which to have an accident. Higher speeds dramatically increase the risk of death in an accident, and more recent car models are far less prone



**Figure 7:** Death rate by road user type

to fatal accidents than older varieties. Finally, being young and male is most strongly associated with having an accident, and regional highways tend to be the deadliest roads in the state.

## Software and Packages

This report was created using R Core Team (2021). We also used packages created by Goode and Rey (2019), Robinson, Hayes, and Couch (2021), Wickham et al. (2019), Grolemund and Wickham (2011), Firke (2021), Xie (2021), Zhu (2021), Xie (2020), Kahle and Wickham (2013), Arnold (2021).

## References

- Arnold, JB (2021). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 4.2.4.  
<https://CRAN.R-project.org/package=ggthemes>.
- Firke, S (2021). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. R package version 2.1.0.  
<https://CRAN.R-project.org/package=janitor>.
- Gislason, T, K Tomasson, H Reynisdottir, J Björnsson, and H Kristbjarnarson (1997). Medical risk factors amongst drivers in single-car accidents. *Journal of internal medicine* **241**(3), 217–223.

- Goode, K and K Rey (2019). *ggResidpanel: Panels and Interactive Versions of Diagnostic Plots using 'ggplot2'*. R package version 0.3.0. <https://CRAN.R-project.org/package=ggResidpanel>.
- Grolemund, G and H Wickham (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software* **40**(3), 1–25.
- Kahle, D and H Wickham (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal* **5**(1), 144–161.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Robinson, D, A Hayes, and S Couch (2021). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.5. <https://CRAN.R-project.org/package=broom>.
- Wickham, H, M Averick, J Bryan, W Chang, LD McGowan, R François, G Grolemund, A Hayes, L Henry, J Hester, M Kuhn, TL Pedersen, E Miller, SM Bache, K Müller, J Ooms, D Robinson, DP Seidel, V Spinu, K Takahashi, D Vaughan, C Wilke, K Woo, and H Yutani (2019). Welcome to the tidyverse. *Journal of Open Source Software* **4**(43), 1686.
- Xie, Y (2020). bookdown: Authoring Books and Technical Documents with R Markdown. R package version 0.21. <https://github.com/rstudio/bookdown>.
- Xie, Y (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.32. <https://yihui.org/knitr/>.
- Zhu, H (2021). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4. <https://CRAN.R-project.org/package=kableExtra>.