1. Introduction

In the year 2011 & 2012, Victoria had 25,485 road accidents, resulting in 563 deaths, 9,799 serious injuries, and 22,687 other injuries requiring medical attention. The Victorian Association of Local Government (VALG) wants to reduce these numbers, because they have some responsibility both to the public and because they oversee road maintenance and parking. This responsibility is limited because of its shared nature with the federal and state governments, and therefore their actions are limited in what they can do, and this means targeted acts driven by data are even more important.

This report aims to produce those targeted actions driven by data by firstly producing insights, each with a link to a further piece of analysis, followed by recommended actions, which are related to one of those specific insights, followed by an analysis of 2-years of data from 2011 and 2012.

2. Insights

The key insight is strategic, VALG believed that focusing on reducing total accidents will lead to a reduction in deaths on the road and injuries. While this makes logical sense, in that a road accident must happen for someone to die or be injured in a road accident (Appendix 5.K). This turned out not to be true**, fatalities and serious injuries are only weakly and moderately correlated to accident count** (section 4.1)**.** This is because accident counts, and medical injuries are mainly associated with schools (section 4.2 & section 4.3). While fatal and serious injuries are mainly associated with high speeds (section 4.4) and drink driving (section 4.5).

Some other potentially insights were captured in this project. One is that 99% of accidents are related to 2 roads but most 2nd roads have very little information gathered (section 4.6). While another is that dusk (and to a lesser extent dawn) also relate to accidents seemingly from light shining in people's eyes (section 4.7).

3. Recommended Actions

The following are recommended actions from this project:

1. **Decide which is the true goal and target KPIs for that**. Is it reducing fatal and serious injury car crashes or number of accidents and medical injuries?
2. **If the former, focus on reducing speeds and stopping drink driving**. Some suggested actions for reducing speeds are decreasing spend on road maintenance on highways to while increasing spend at parking near public transportation, this will discourage driving and encourage public transportation use. Or another methods may include reducing speed limits at targeted spots. While for reducing drink driving (especially around the early mornings on weekends) perhaps more public transport close to night spots, closing bars earlier, reduce availability of parking around these areas, and offering blood-alcohol tests close to these zones.
3. **If the latter, focus on reducing vehicular collusions around schools.** Some suggested actions include decreasing number of cars in the area by increasing school bus routes, offering free public transportation to students. Other methods may include having classes with students about road safety and increasing signage around schools.
4. **Gather better data on 2nd roads and the weather at time of crash**. If 2nd roads are very impactful and weather conditions, like the dusk/dawn lighting impact discussed earlier, having better quality data may lead to important new actionable insights. There is some evidence that potentially the data quality is not the best in the current sources (section 4.8)

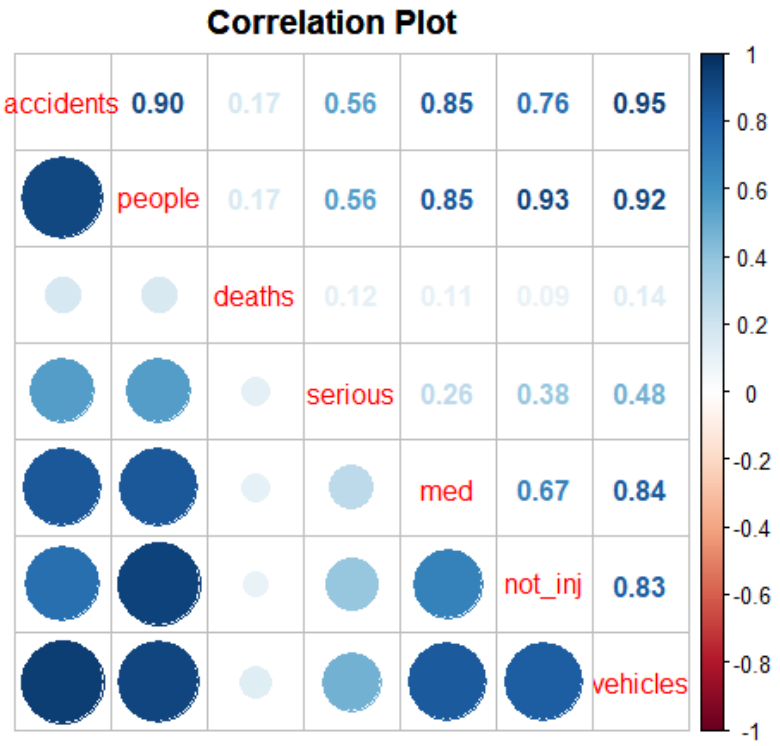
5. **Encourage the community to change time they start work, so less driving during dusk**. Maybe by changing LAGV start time to 1 hour earlier or later, so they are not driving at dusk and this will then have hopefully flow on effects.

4. Analysis

An analysis was done on the 2011 and 2012 data which was seemingly provided from an accident register (section 4.8). This analysis will go through the following sections:

- Correlation between accident count and deaths, serious injuries, and medical injuries
- A time series analysis of accidents
- A seasonality analysis of accidents
- A vehicle speed analysis
- A Weekend Early Morning analysis
- 2nd Road's analysis
- Dusk/Dawn analysis
- Other Data Quality

4.1 Correlation between accident count and deaths, serious injuries, and medical injuries



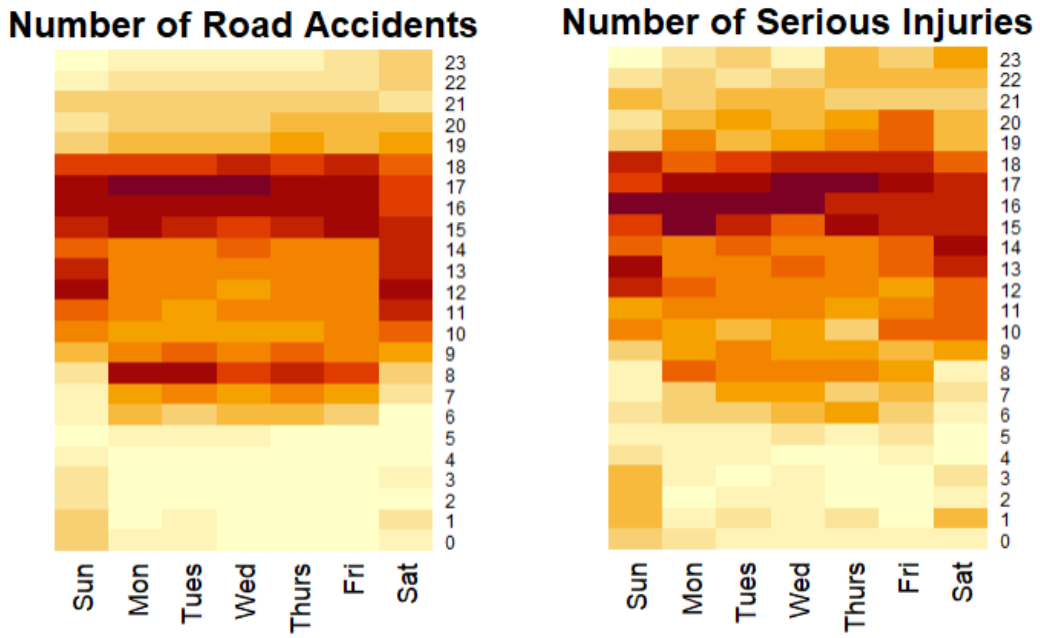
The above is a Correlation Plot which measures the level of association between a change in 1 variable and a change in another. A 1 would mean when one variable increases the other must and vice versa if one decreases the other must decrease. While a -1 implies the opposite, one increases the other must decrease, and a 0 implies no association. The graph has the daily total of all of the different variables

Interestingly accidents/day is weakly correlated with deaths, moderately with serious injuries, and highly with medical injuries.

Serious injuries seem to be moderately correlated with number of vehicles and amount of people too, this may be related to numbers of accidents though (e.g., accidents are moderately correlated to serious injuries, but accidents are highly correlated to vehicles and people, therefore vehicles is moderately correlated but slightly weaker to serious injuries, which is what we see).

4.2 A weekly time series analysis of accidents

The below 2 graphs show the hour of the day on the vertical axis, and the day of the week on the horizontal axis, with each cell being one hour day combination for the two years of data provided.



We can observe that road accidents have a strong weekday cycle with lots of accidents during the morning and afternoon rush hours and school runs, while the weekends show an increase in the afternoon and the early mornings.

The serious injuries show a similar association but much less injuries in the weekday mornings and more serious injuries in the late nights and early mornings on the weekends, presumably caused by people being inebriated.

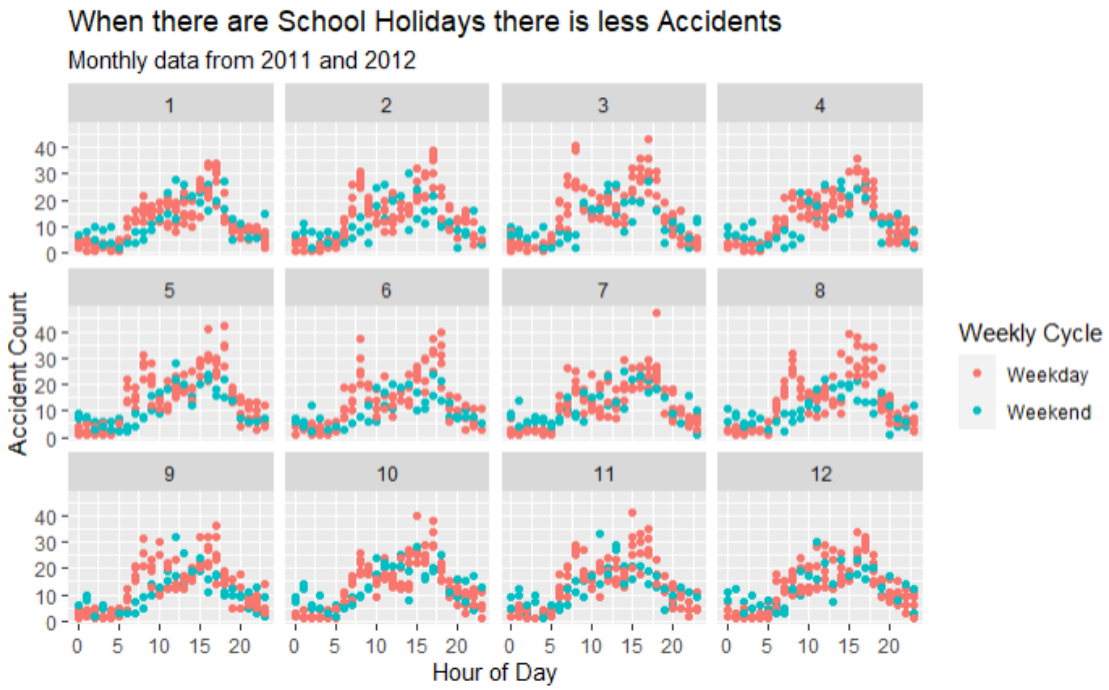
Referring to appendix 5.A, it can be shown that deaths do not seem to have a weekly pattern apart from maybe the early hours in the weekend. While medical injuries are similar to road accidents in distribution.

4.3 A seasonality analysis of accidents

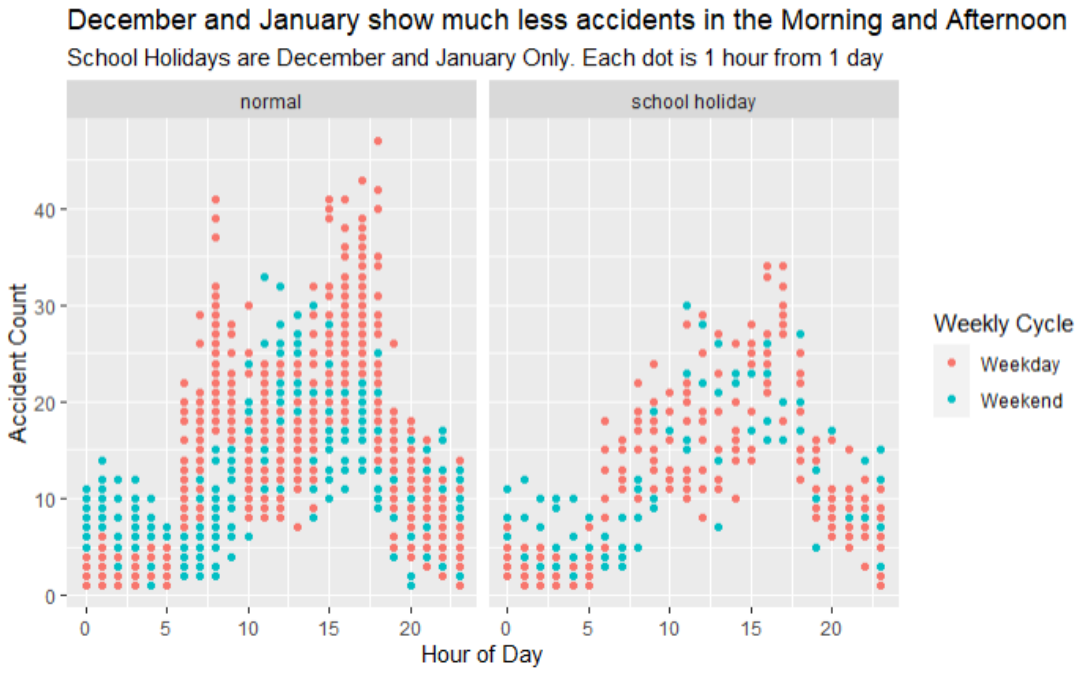
The below graph shows accidents by time of day by month, with each graph headlined by the number of the month.

While the early morning peak on the weekend is clearly visible as shown in the graphs in section 4.2. There is an interesting change during the January, April, July, and October weekdays with all having

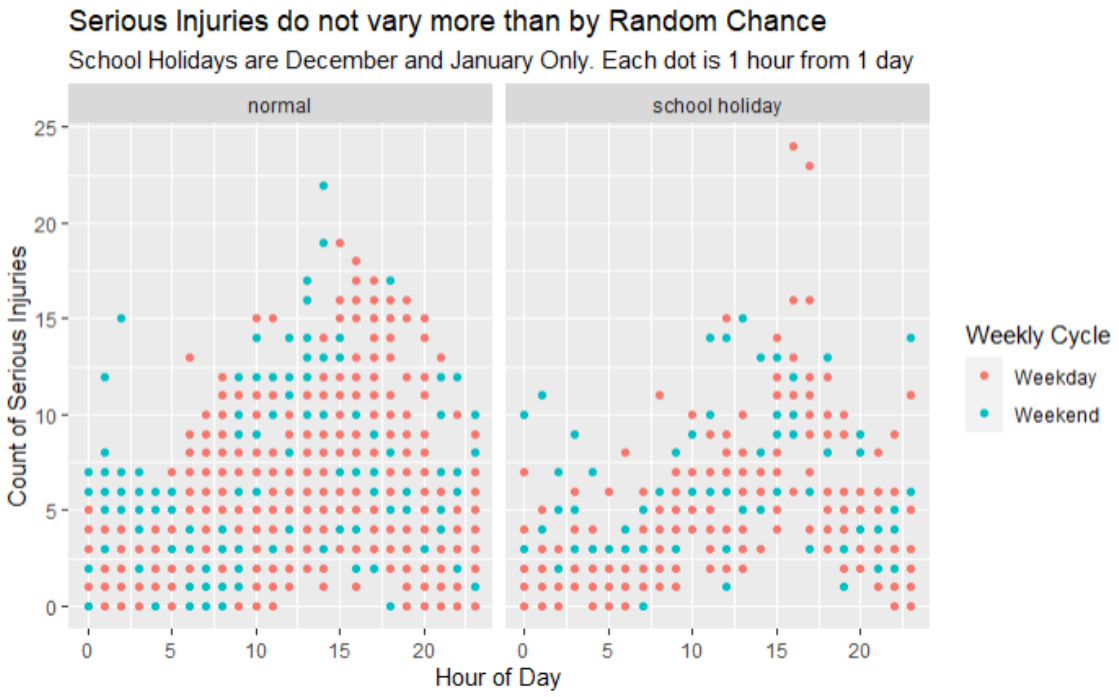
smaller early morning peaks and potentially afternoon peaks, these months generally have school holidays.



If we just look at December and January, where the entire month is school holidays, we can see the effect quite strongly. This trend is also seen in medical injuries (Appendix 5.B)

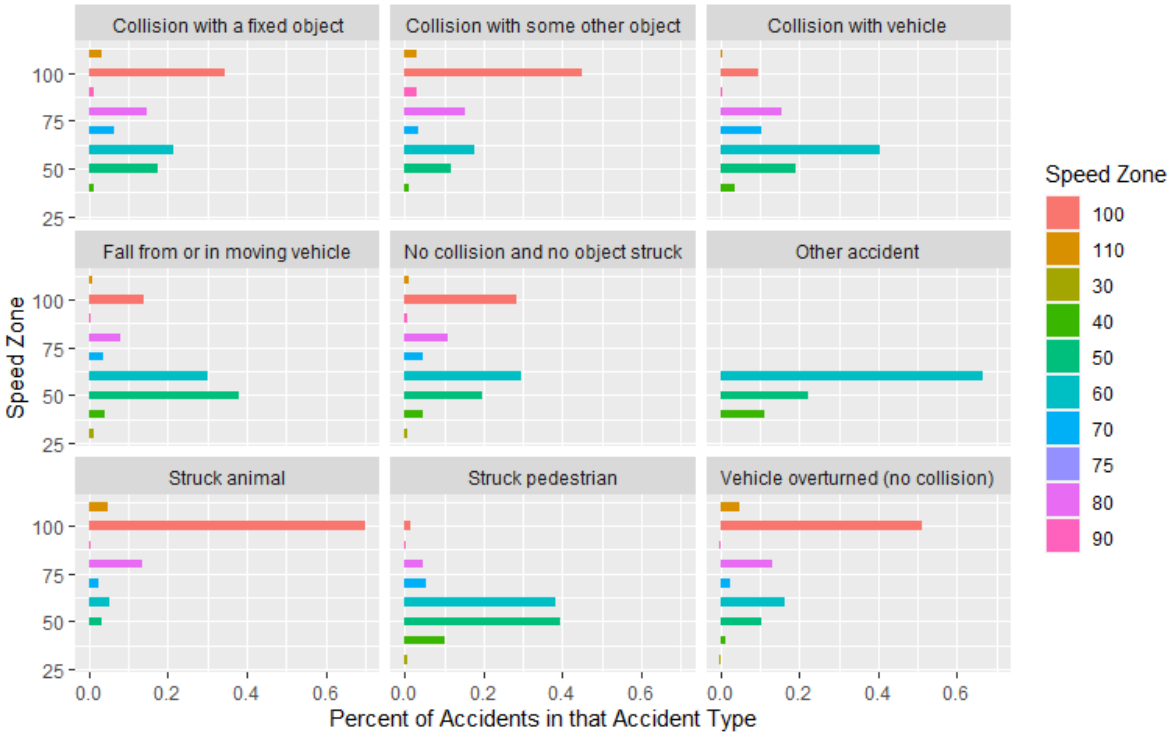


But the accidents that are associated with school holidays do not seem to effect serious injuries and deaths, as shown in the chart below and in Appendix 5.B.

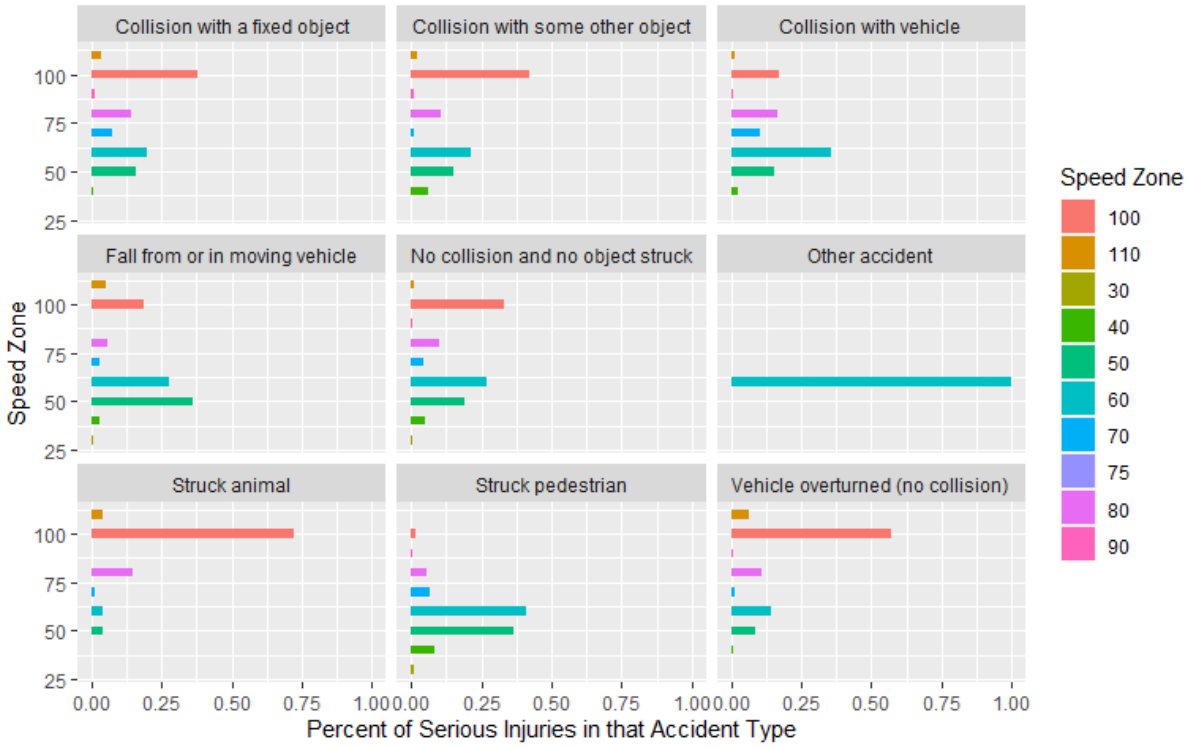


This seems to imply that schools are associated with accidents but not school children. If it was school children related, we would see an elevated level during the summer school holidays, but it would be uniformly distributed throughout the day, as the school children have lots of free time. The fact that it is elevated in these peaks imply it is associated with the time school children are getting in and out of school rather than the children itself.

4.4 A vehicle speed analysis



A potential cause of accidents may be the speed that the vehicle is travelling, which maybe signified by the speed limit at the time of driving. The above and below charts shows an interesting difference between the two. With collusions with fixed objects, other objects, and with vehicles more likely to have serious injuries at high speeds than their proportions of total accidents. This makes intuitive sense as high-speed impacts would seem to be more dangerous. This trend is also seen in medical injuries and deaths with medical injuries being closer to accident count as medical injuries are much more likely to occur than fatalities or serious injuries. While deaths from accidents are more likely to come from high speeds too (see Appendix 5.D)

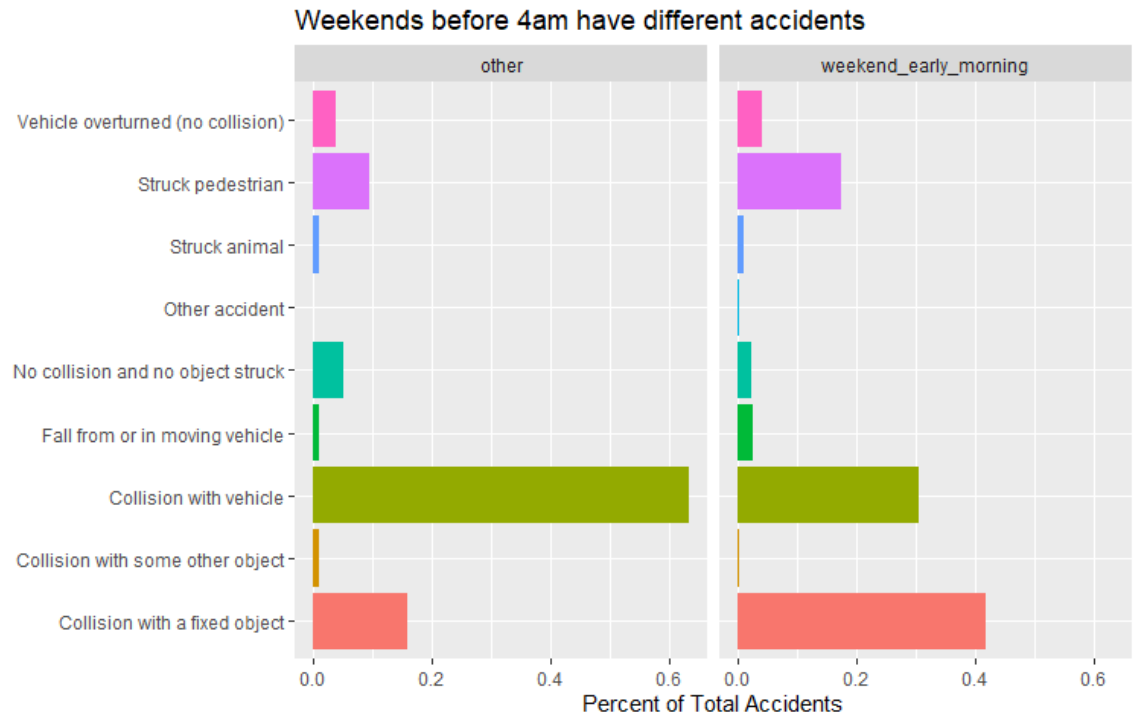


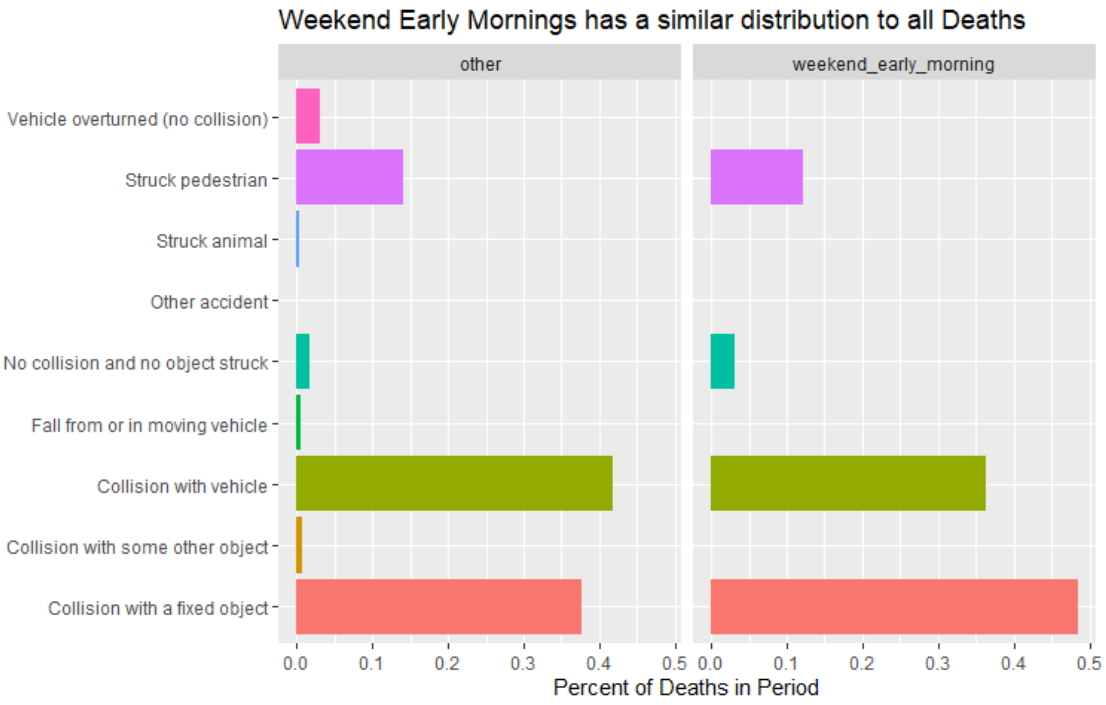
Considering most high-speed zones are on highways or in rural areas, while low speed zones are in urban and industrial areas, this does imply that targeting reduced deaths and serious injuries requires a different approach than targeting total accidents and less serious injuries.

4.5 A weekend early morning analysis

As mentioned in section 4.2 and 4.3, the early morning on the weekend has not only a step-up in accidents but in serious injuries and deaths. The below 2 charts and appendix 5.E show why that may be the case.

The first graph shows the percent of total accidents by accident type, with the left graph showing all time periods, and the right weekends before 4am. The distribution of accidents is different with less vehicle collusions and more pedestrians and object collusions. This is important, as the graph after shows that deaths from vehicle accidents are much more likely to occur from accidents featuring pedestrians and objects collusions than in other kinds of accidents. This demonstrates that while early morning accidents may be rare compared to other times of the week (see section 4.2), they are more likely to cause serious injuries and potentially deaths.





It is likely that this is caused by people driving while having alcohol in their symptoms through areas with night clubs, bars, and pubs. This may explain why pedestrians are more likely to be struck as they may be running across the road, while fixed objects may be street lights and other such items as drivers swerve off the road.

4.6 2nd Road's analysis

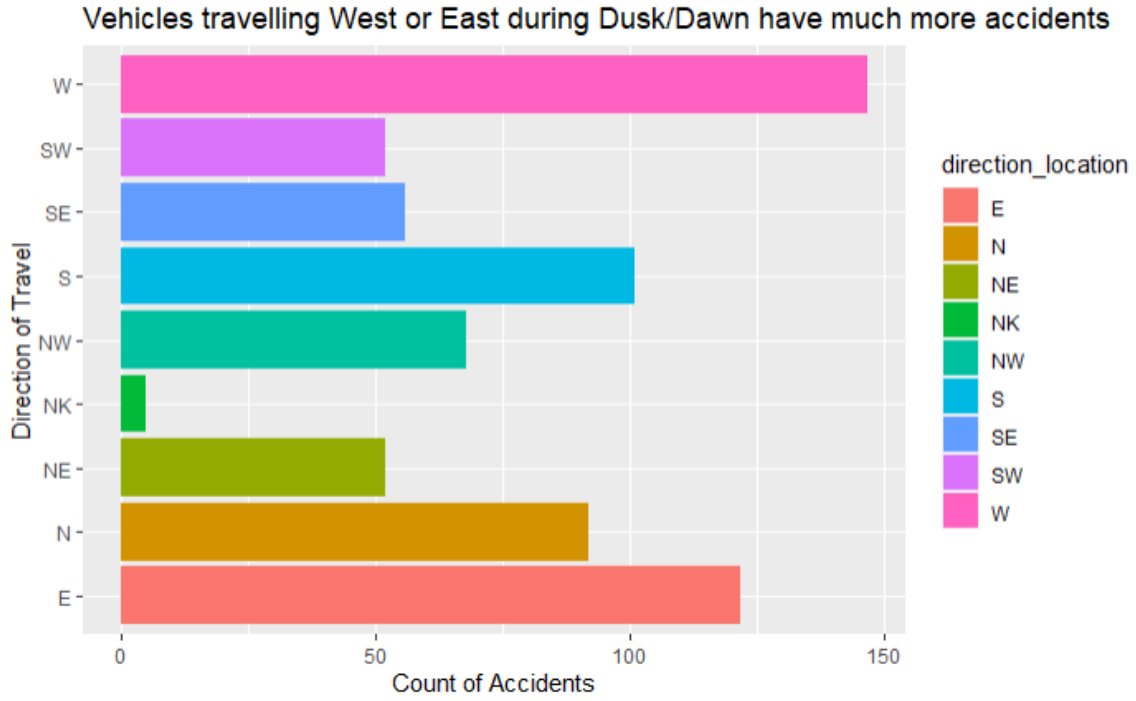
Of the 25,484 observations, there was missing road type for the 2nd road involved in an accident in 1416 observations, but names of the roads were missing in 33 observations (Appendix 5.F). This

means potentially important data on the 2nd road was missing in the analysis. 66% of the road 2 had the title of “Unnamed” which means that a lot of the data is missing on these 2nd roads (Appendix 5.H).

It is possible that "Unnamed" means something different to what it sounds like, for instance it may be used for no road, or a driveway and that the data entry person needs trained or the system improved. Perhaps call it a driveway instead or no second road if those are true. There is some evidence of that as shown in appendix 5.G.

4.7 Dusk/Dawn analysis

The below chart shows a count of all accidents during dusk and dawn times and the direction the vehicle is driving in.



If there was no association between the direction the vehicle was travelling in and the time period, the number of accidents should be uniformly distributed between the different direction of travel. East and West directions are much more likely than other directions, which may be caused by the dusk and dawn sunlight being in the drivers’ eyes. This effect is seen across medical injuries, serious injuries and deaths as shown in Appendix 5.C.

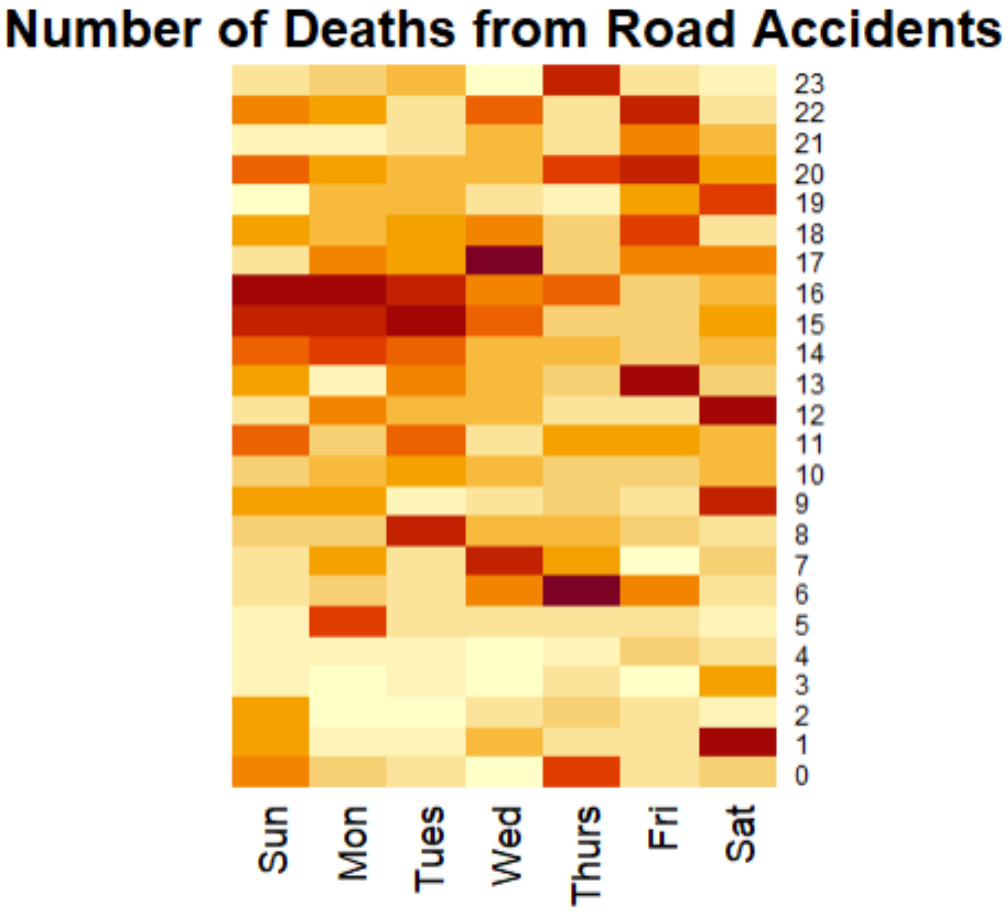
4.8 Other Data Quality issues

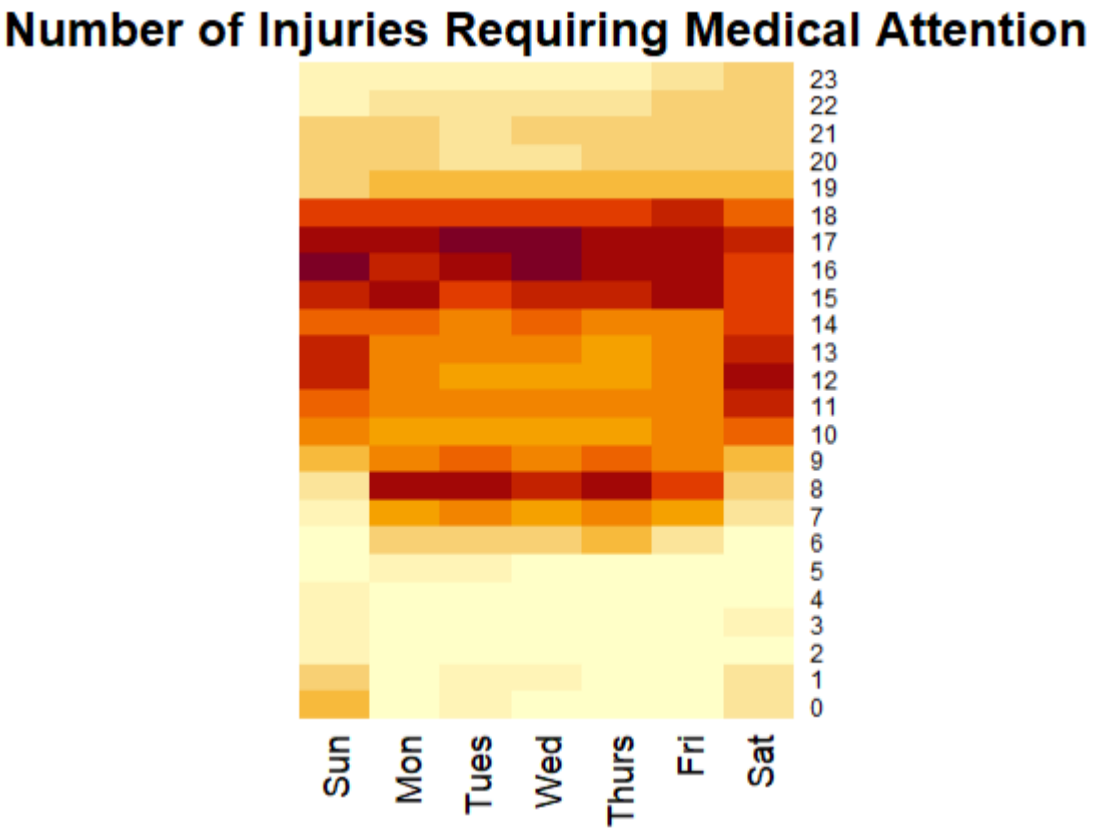
There were also numerous data irregularities in the accident data set provided. Apart from the “Unnamed Roads”, there was the below to add:

- Direction of accidents are missing more than half the time (see Appendix 5.F).
- Accidents are filtered from a bigger report rather than their own database and may not be entered in a timely period (see 5.I).
- Serious injuries do not reconcile to the Victorian Government with 9,799 in this report and 10,500 in the Victoria Road Safety Trauma report (see Appendix 5.J).

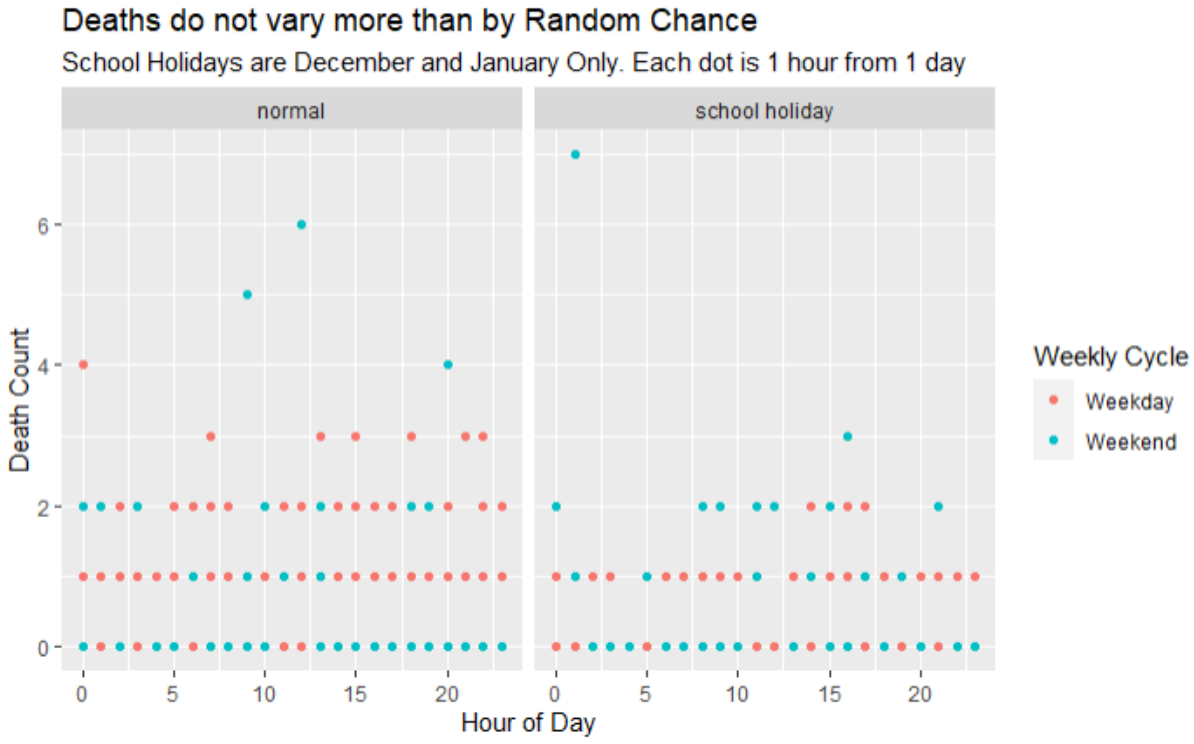
5. Appendix

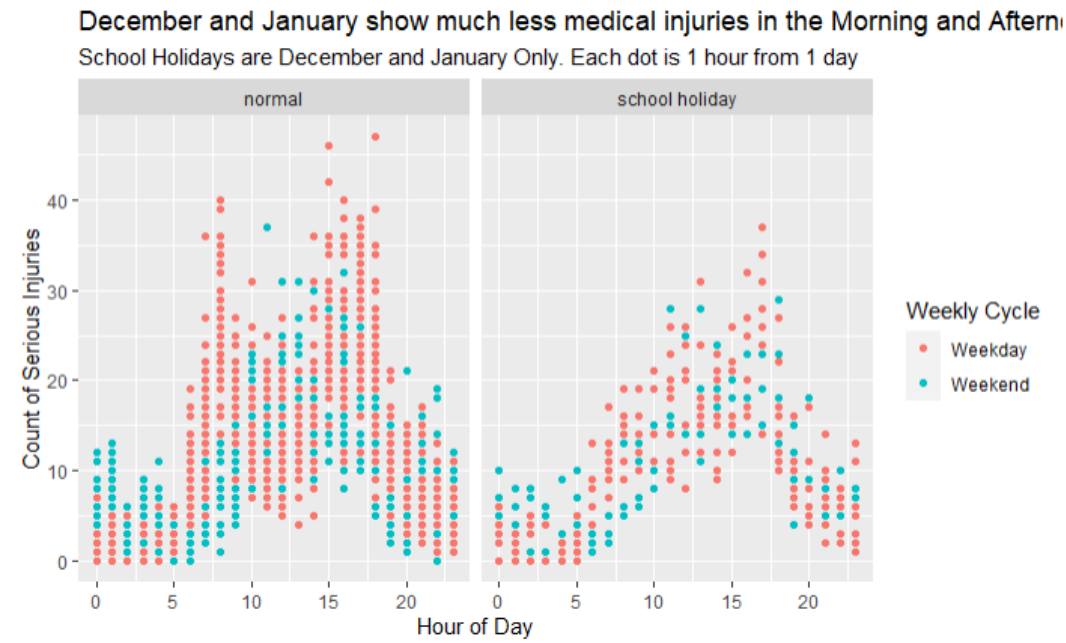
5.A



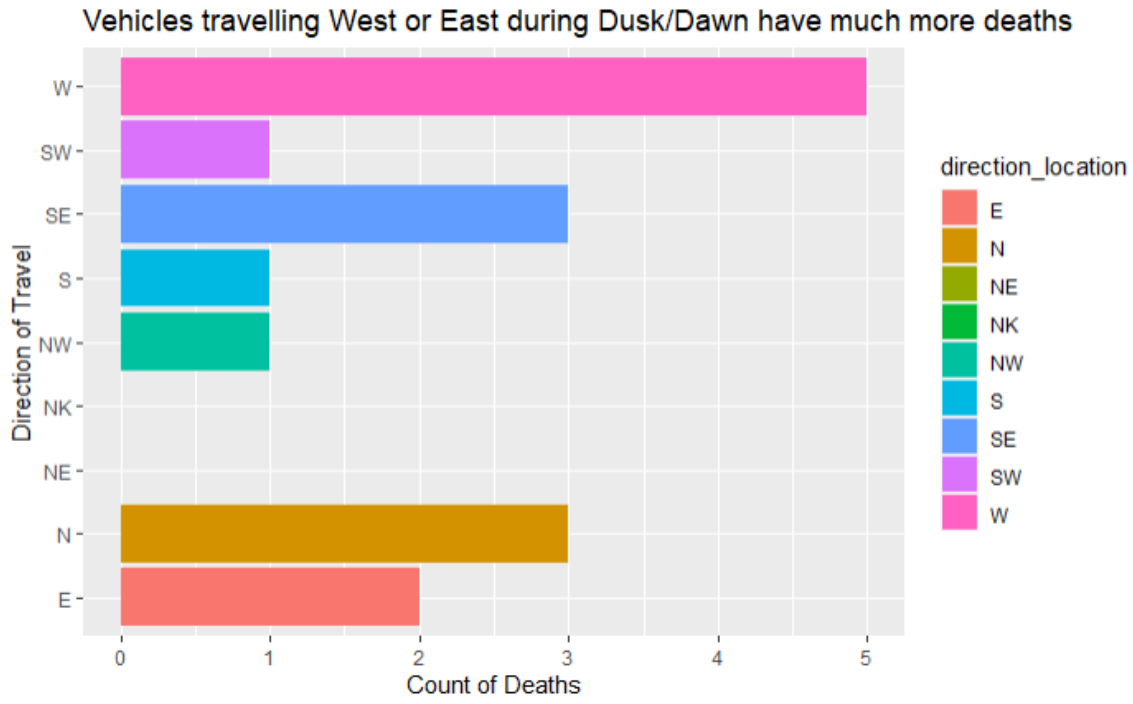

Number of Injuries Requiring Medical Attention
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Sun
Mon
Tues
Wed
Thurs
Fri
Sat

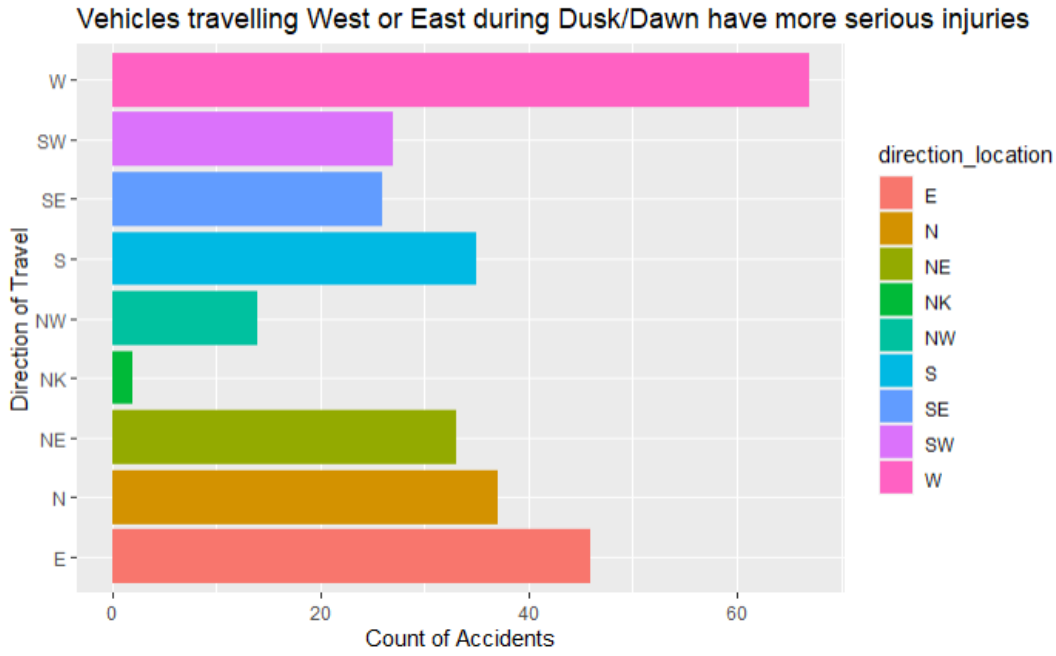
5.B

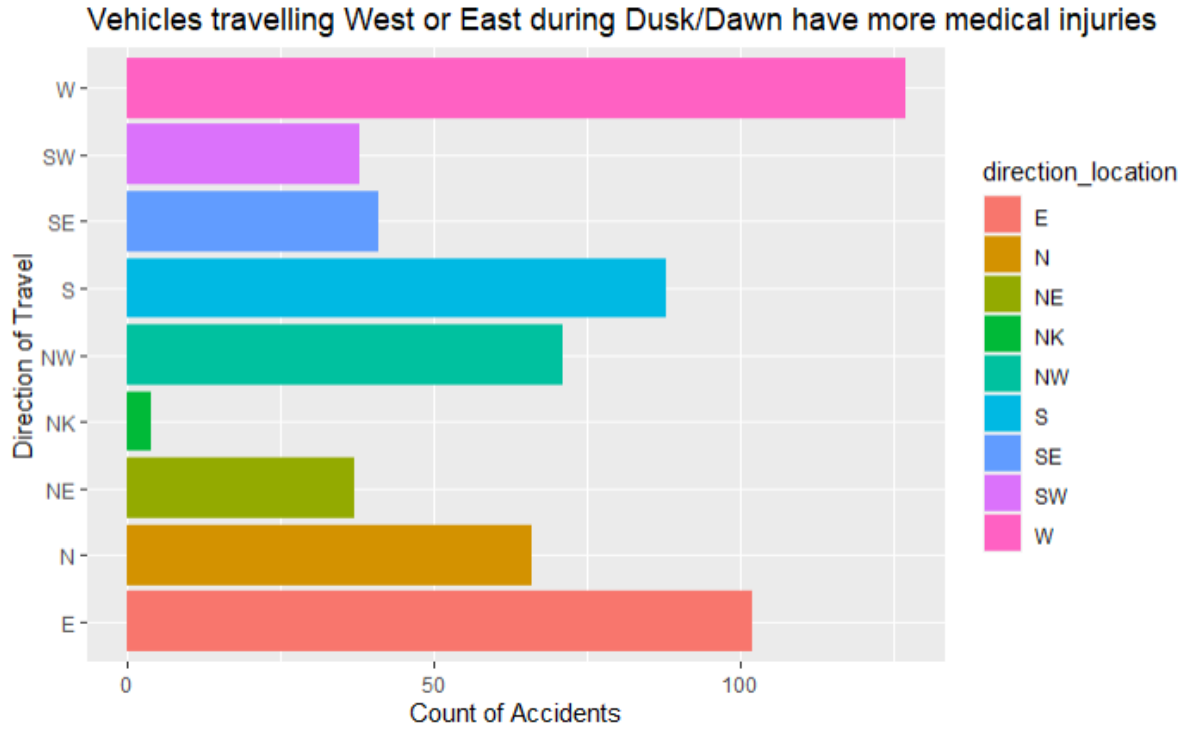

Deaths do not vary more than by Random Chance
School Holidays are December and January Only. Each dot is 1 hour from 1 day
normal
school holiday
Death Count
0
2
4
6
0
5
10
15
20
Hour of Day
Weekly Cycle
Weekday
Weekend


December and January show much less medical injuries in the Morning and Aftern
School Holidays are December and January Only. Each dot is 1 hour from 1 day
normal
school holiday
Count of Serious Injuries
0
10
20
30
40
0
5
10
15
20
Hour of Day
Weekly Cycle
Weekday
Weekend

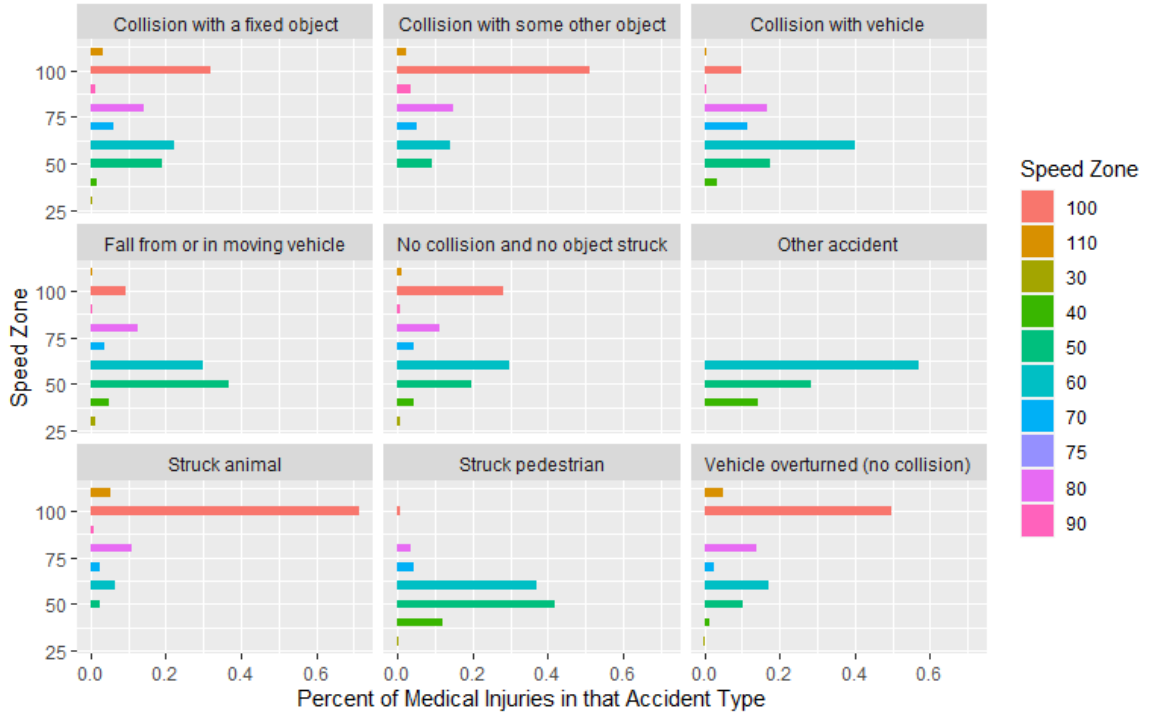
5.C

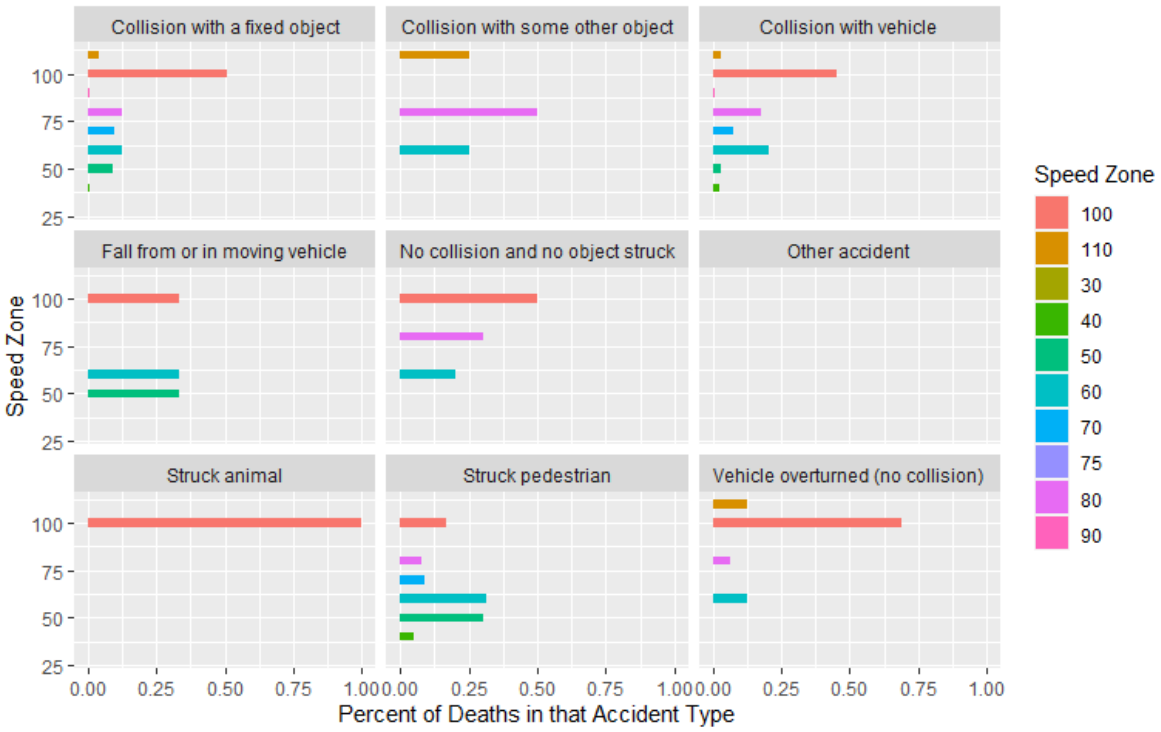

Vehicles travelling West or East during Dusk/Dawn have much more deaths
Direction of Travel
W
SW
SE
S
NW
NK
NE
N
E
0
1
2
3
4
5
Count of Deaths
direction_location
E
N
NE
NK
NW
S
SE
SW
W


Vehicles travelling West or East during Dusk/Dawn have more serious injuries
Direction of Travel
W
SW
SE
S
NW
NK
NE
N
E
0
20
40
60
Count of Accidents
direction_location
E
N
NE
NK
NW
S
SE
SW
W

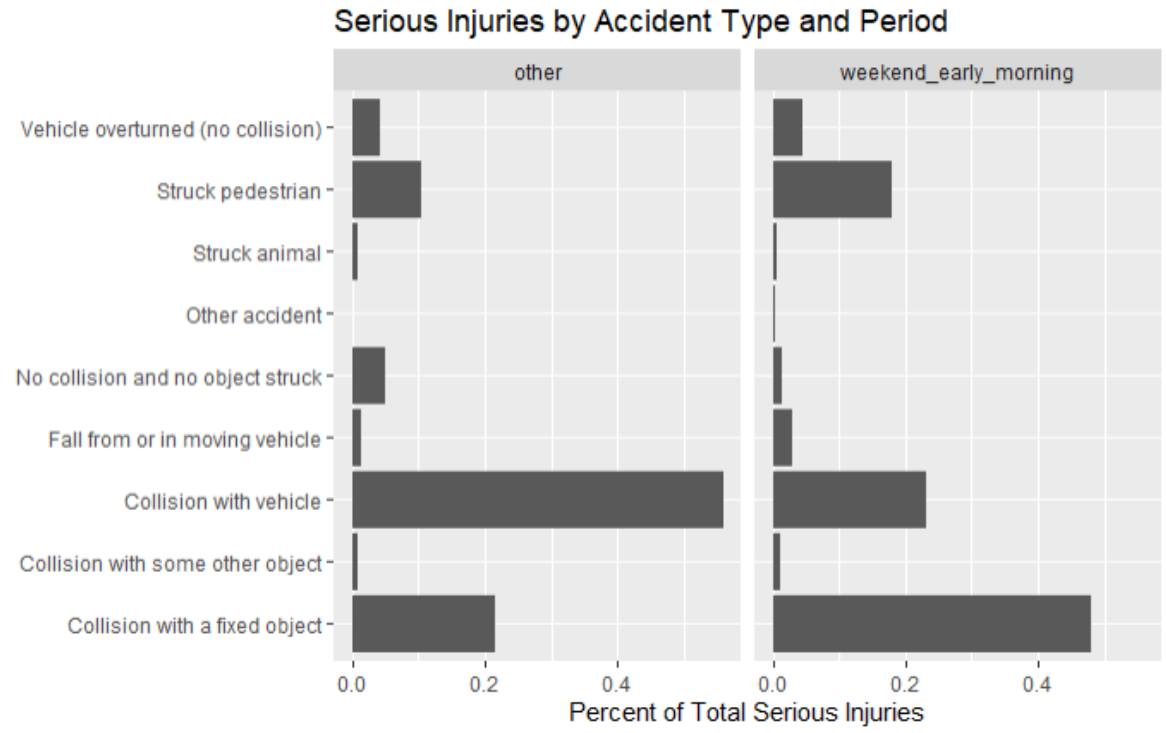

Vehicles travelling West or East during Dusk/Dawn have more medical injuries
Direction of Travel
W
SW
SE
S
NW
NK
NE
N
E
0
50
100
Count of Accidents
direction_location
E
N
NE
NK
NW
S
SE
SW
W

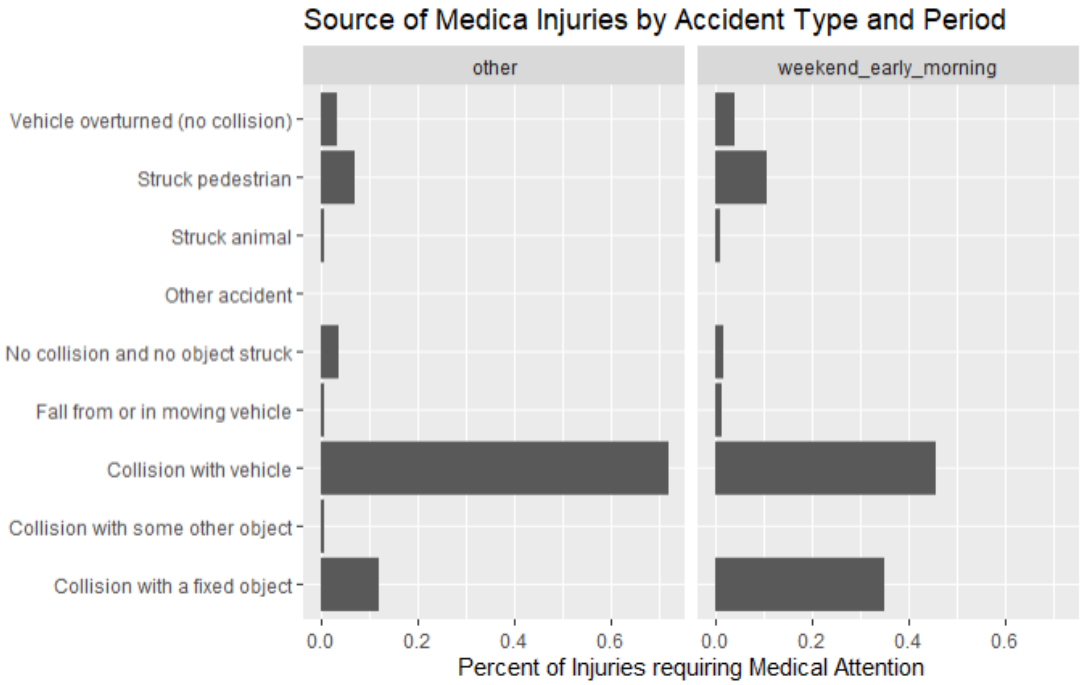
5.D


Collision with a fixed object
Collision with some other object
Collision with vehicle
Fall from or in moving vehicle
No collision and no object struck
Other accident
Struck animal
Struck pedestrian
Vehicle overturned (no collision)
Speed Zone
100
75
50
25
0.0
0.2
0.4
0.6
Percent of Medical Injuries in that Accident Type
Speed Zone
100
110
30
40
50
60
70
75
80
90


Collision with a fixed object
Collision with some other object
Collision with vehicle
Fall from or in moving vehicle
No collision and no object struck
Other accident
Struck animal
Struck pedestrian
Vehicle overturned (no collision)
Speed Zone
100
75
50
25
0.00
0.25
0.50
0.75
1.00
Percent of Deaths in that Accident Type
Speed Zone
100
110
30
40
50
60
70
75
80
90

5.E


Serious Injuries by Accident Type and Period
other
weekend_early_morning
Vehicle overturned (no collision)
Struck pedestrian
Struck animal
Other accident
No collision and no object struck
Fall from or in moving vehicle
Collision with vehicle
Collision with some other object
Collision with a fixed object
0.0
0.2
0.4
Percent of Total Serious Injuries


Source of Medica Injuries by Accident Type and Period
other
weekend_early_morning
Vehicle overturned (no collision)
Struck pedestrian
Struck animal
Other accident
No collision and no object struck
Fall from or in moving vehicle
Collision with vehicle
Collision with some other object
Collision with a fixed object
0.0
0.2
0.4
0.6
Percent of Injuries requiring Medical Attention

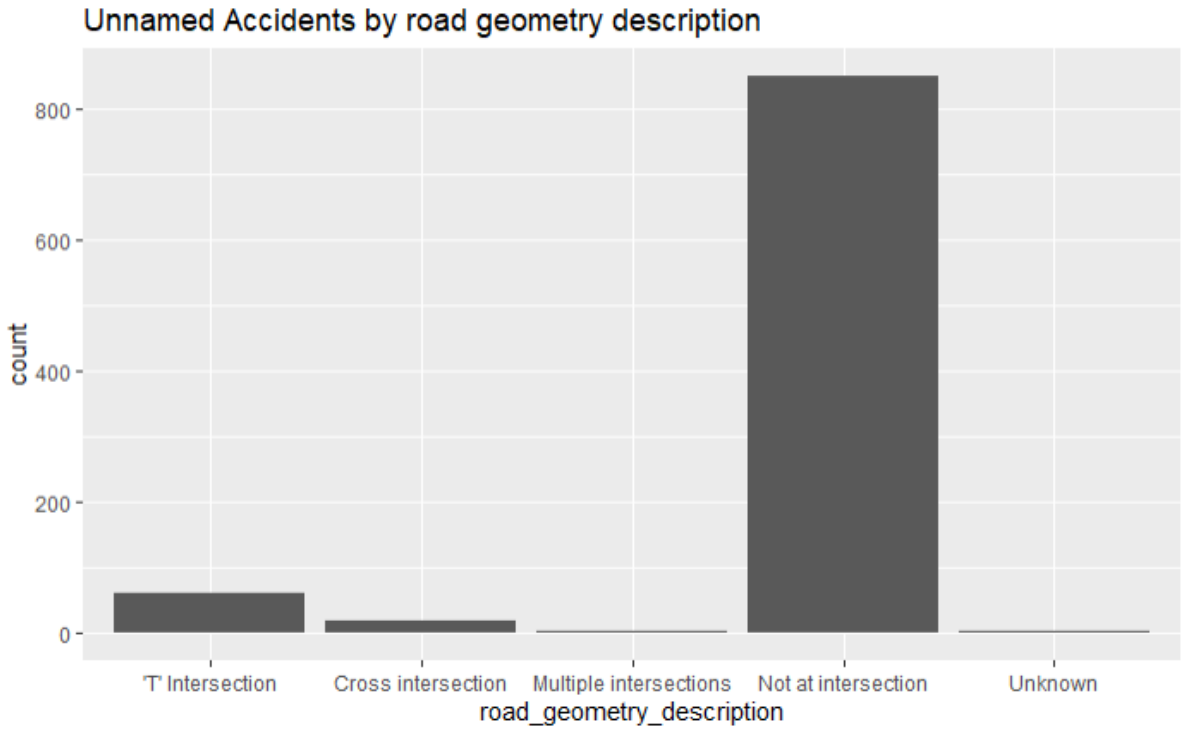
5.F

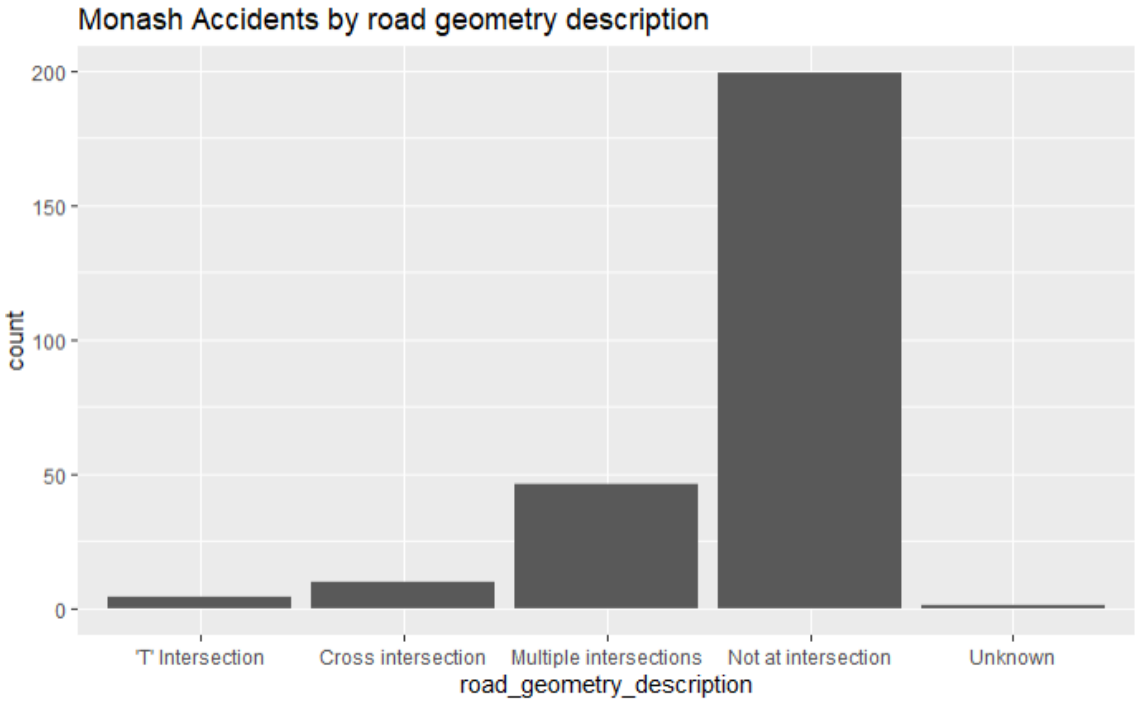
```
#Nulls and NAs
```{r}
colSums(is.na(df_crash_vic_raw))
dim(df_crash_vic_raw)
```
```

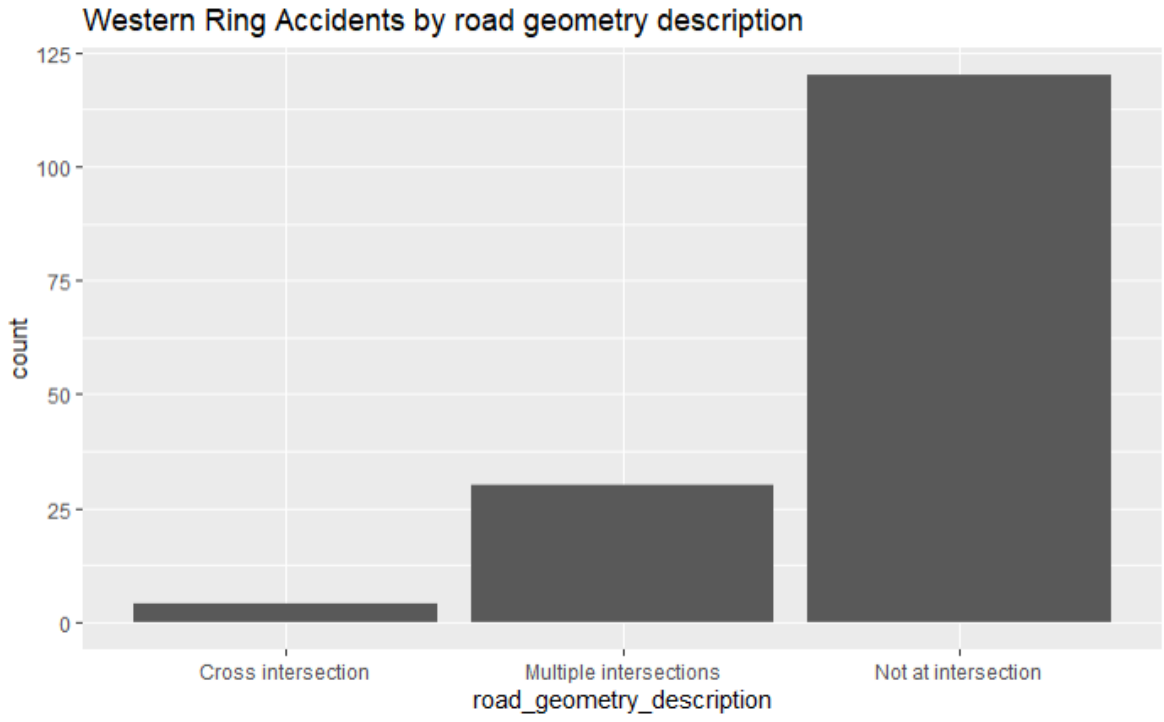
```
                    accident_no                   accident_date                   accident_time 
                              0                               0                               0 
            accident_description            dca_code_description                   dca_code_type 
                              0                               0                               0 
      light_condition_description              no_persons_involved                no_persons_killed 
                              0                               0                               0 
         no_persons_inj_serious no_persons_inj_medical_attention                no_persons_not_inj 
                              0                               0                               0 
                 no_of_vehicles          police_attend_description          road_geometry_description 
                              0                               0                               0 
                       severity              severity_description                      speed_zone 
                              0                               0                               0 
                  location_type                  locals_involved                             lon 
                              0                               0                               0 
                            lat                     road_name_1                     road_type_1 
                              0                               1                             441 
                    road_name_2                     road_type_2               direction_location 
                             33                            1416                           13478 
                      directory                         edition                            page 
                              0                               0                               0 
                grid_reference_x                grid_reference_y                        lga_name 
                              0                               0                               0 
                    region_name                    hour_number                  speed_zone_char 
                              0                               0                               0 
                   accident_date2 
                              0 
[1] 25484    37
```

5.G

The unnamed roads have a slightly different description if looking intersections versus 2 other regions.




Monash Accidents by road geometry description
200
150
100
50
0
count
'T' Intersection
Cross intersection
Multiple intersections
Not at intersection
Unknown
road_geometry_description

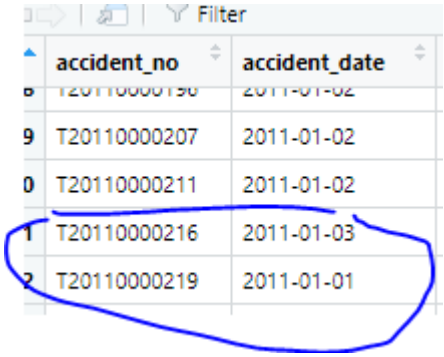

Western Ring Accidents by road geometry description
125
100
75
50
25
0
count
Cross intersection
Multiple intersections
Not at intersection
road_geometry_description

5.H

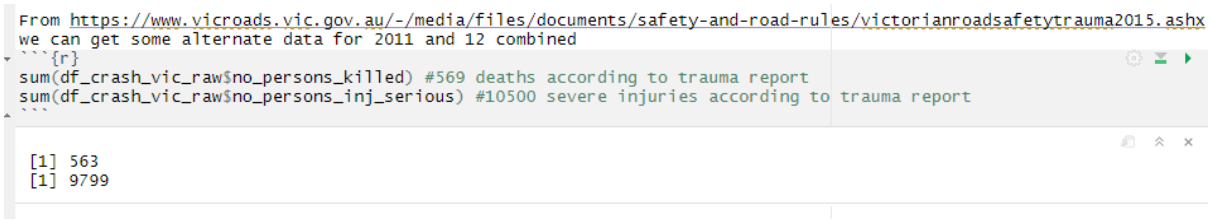
road_name_2 <chr>	count <int>	percent <dbl>
UNNAMED	935	0.6603107345
WESTERN LINK TOLLWAY	66	0.0466101695
EASTLINK	39	0.0275423729
NA	33	0.0233050847
SOUTHERN LINK TOLLWAY	27	0.0190677966
WESTERN LINK & WEST GATE	24	0.0169491525
THE AVENUE	20	0.0141242938
HUME & METROPOLITAN	14	0.0098870056
PRINCES & WESTERN RING	14	0.0098870056
SHEPPARTON ALTERNATIVE ROUT	14	0.0098870056

1-10 of 97 rows

5.I



5.J



5.K

“Your (fictional) client, The Victorian Association of Local Government (that contains representatives from the 79 councils in VIC) have given two years worth of road accident data to you. Using your analytical skill, they want you analyse this data and come up with some real policies/initiatives they can implement that, according to the data, will improve the road accident rate (and resultant rate of injuries and deaths).” From “Dataset A_BNE – Road Accident Data.xlsx”