11/3/2019

Analysis of Road Accidents' Severity using Multilevel Modeling

Abhisek Gautam [13679042]

Introduction

Road accidents may cause deaths and it may leave a lot of people seriously injured. 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 is increasing with more than 1.5% per annum, inline with a similar increase in the population (Australian Bureau of Statistics 2019, Australian Bureau of Statistics 2018). There is thus an increasing call for road safety by the government, which can be noticed with the Towards Zero campaign setup by the Victorian government. The dataset that was analysed was collected also collected from a government website; from the Department of Planning, Transport and Infrastructure. Previously, our team had done a binomial logistic regression on the dataset to find out what were the causes of high severity and fatalities in road accidents of Victorian Roads. The team had concluded that type of accident, light condition, type of road and type of vehicle during the time of the accident had the most effect on the severity of road accidents.

Low p (hence, statistically significant) values were observed on the vehicle's part, such as the age of the vehicle and the type of vehicle (like buses, vans, taxis and cars). This study dives deeper into the vehicle aspect of road accidents to answer the question of whether vehicle's make and model have an impact on the severity of road accidents or not.

Details on the Previous Study

The previous study included a detailed exploratory data analysis (EDA) for the road accident data, which was used to choose the independent variables. The final result of the analysis is shown in Appendix 1. It shows that the type of vehicle is significant in determining the severity of the accident. However, there may be an underlying effect within, where the make and the model of the vehicles would be causing an impact on the accidents' severity. Such a structure is not considered by the previous model and without looking into more details, it cannot be assumed that Vehicle Type, in general, is significant. A simple multilevel model would be able to factor the underlying variables.

```
Vehicle.Type.DescBus/Coach                                        -2.246087   0.406339  -5.528 3.25e-08 ***
Vehicle.Type.DescCar                                              -1.598060   0.096165 -16.618  < 2e-16 ***
Vehicle.Type.DescHeavy Vehicle (Rigid) > 4.5 Tonnes               -1.945070   0.247829  -7.848 4.21e-15 ***
Vehicle.Type.DescLight Commercial Vehicle (Rigid) <= 4.5 Tonnes GVM -1.989273   0.193751 -10.267  < 2e-16 ***
Vehicle.Type.DescMini Bus(9-13 seats)                             -3.474495   0.598475  -5.806 6.41e-09 ***
Vehicle.Type.DescMoped                                             1.793562   1.233744   1.454 0.146014
Vehicle.Type.DescMotor Cycle                                       0.128102   0.104250   1.229 0.219148
Vehicle.Type.DescMotor Scooter                                     0.397302   0.288404   1.378 0.168331
Vehicle.Type.DescOther Vehicle                                    -0.822224   0.473474  -1.737 0.082462 .
Vehicle.Type.DescPanel Van                                        -1.803996   0.181585  -9.935  < 2e-16 ***
Vehicle.Type.DescPlant machinery and Agricultural equipment        -2.010826   0.811952  -2.477 0.013267 *
Vehicle.Type.DescPrime Mover - Single Trailer                     -1.863824   0.255439  -7.297 2.95e-13 ***
Vehicle.Type.DescPrime Mover B-Double                             -2.157834   0.556935  -3.874 0.000107 ***
Vehicle.Type.DescPrime Mover B-Triple                             -2.550906   1.053847  -2.421 0.015496 *
Vehicle.Type.DescPrime Mover Only                                 -1.832172   0.505969  -3.621 0.000293 ***
Vehicle.Type.DescStation Wagon                                    -1.933816   0.103510 -18.682  < 2e-16 ***
Vehicle.Type.DescTaxi                                             -1.627183   0.271357  -5.996 2.02e-09 ***
Vehicle.Type.DescUtility                                          -1.639282   0.116304 -14.095  < 2e-16 ***
OLD_COUNT                                                          0.496068   0.043751  11.338  < 2e-16 ***
ROAD_SURFACE_TYPE2                                                -0.250241   0.290449  -0.862 0.388927
ROAD_SURFACE_TYPE3                                                 0.353946   0.098549   3.592 0.000329 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 1: A snapshot of previous analysis showing Vehicle Type variable and its significance

It is also worthwhile to find out how predictors on the different levels account in increasing or decreasing the severity of accidents. The group predictors (such as the age of the vehicle, damage done to a vehicle, its speed) can be conjugate with different hierarchies, and their effects can be analysed separately using a multilevel model. Using multilevel fixed effects model, inferences can be derived beyond a group. Using multilevel random effects model, the effect of both types of group level predictor variables and group level dummy variables can be estimated (Magnusson, 2015).

Structure of the dataset

The total dataset consists of 28 Vehicle types, from light such as bicycles to heavy ones. With the aim of making this study more specific, it will factor out two wheelers and heavy wheelers and will only take everyday vehicles from the small category of vehicles only, viz. car, station wagon, utility and taxi only.

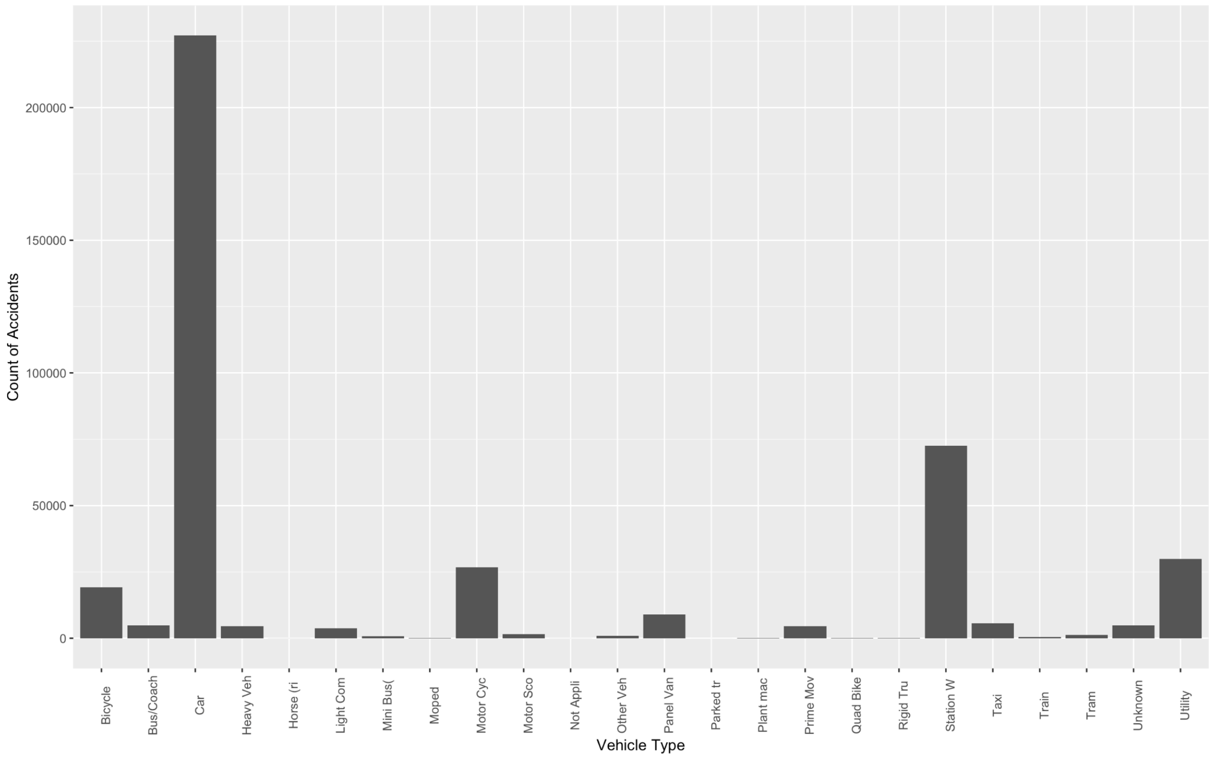

Figure 2: A bar graph showing the count of individual vehicle types in the dataset

The data-frame used in the analysis is as follows:

Column Name	**Description**
ACCIDENT_NO	The Unique ID of the accident
ACCIDENTDATE	The date of the accident
VEHICLE_MAKE	The Make of the vehicle
VEHICLE_MODEL	The Model of the vehicle
Vehicle_Type_Desc	Description of the Type of the Vehicle (Car, Station Wagon)
VEHICLE_YEAR_MANUF	The year of Manufacture of the Vehicle
LEVEL_OF_DAMAGE	Damage Level: 1-6, 1 includes severely damaged
SPEED_ZONE	The speed zone of the road which the accident occurred (This is assumed to be the speed of vehicle during crash and is taken as continuous- as speed of the vehicle is not always equal to speed zone's speed)
ACCIDENT_YEAR	Year of the accident, derived from ACCIDENTDATE (continuous)
NUM_OF_YEARS	Age of Car, obtained from ACCIDENT_YEAR – VEHICLE_YEAR_MANUF
SEVERE	Is the accident severe? Binary: Yes and No

Now, checking for collinearity between the continuous variables NUM_OF_YEARS and SPEED_ZONE, we see that there it is minimal. Hence, both of them can be used in our model.

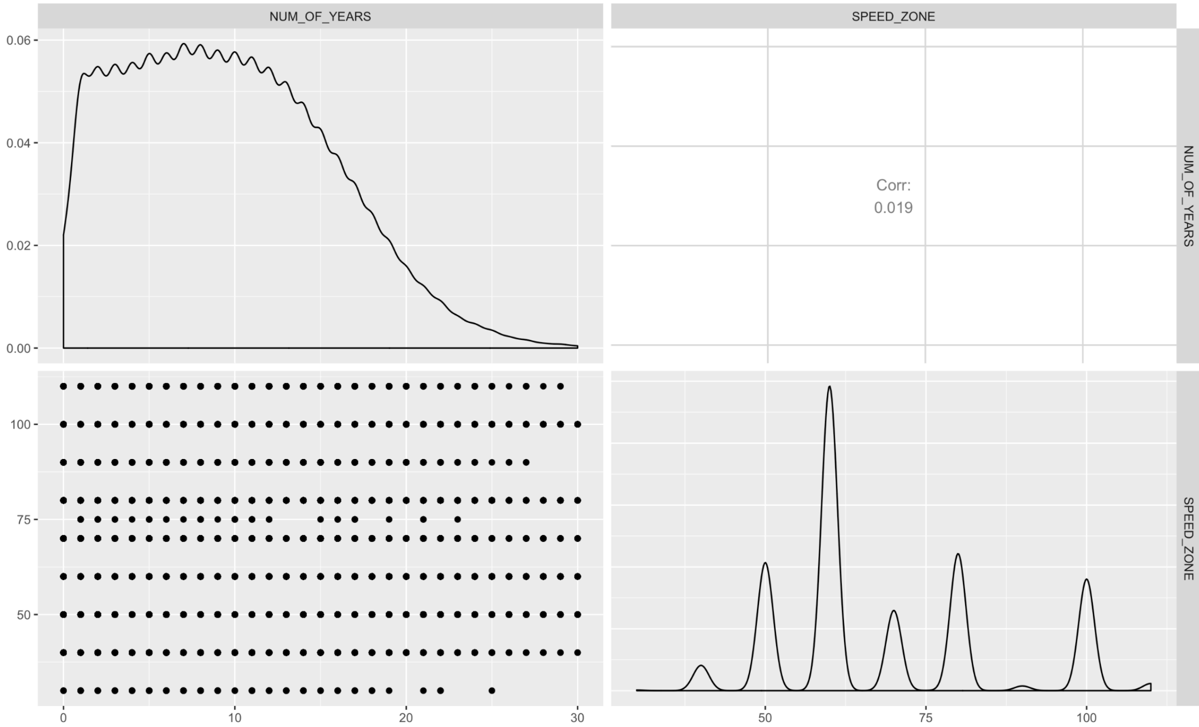

Figure 3: Checking collinearity between continuous residual variables

Looking at the distribution of severity on continuous variables (Figure 4), Number of Years (i.e. the age of car) and Speed Zone, we find that there is not much impact of the age of car on the severity of accidents, whereas with regards to the speed zone, accidents are more severe at higher speed zone areas.

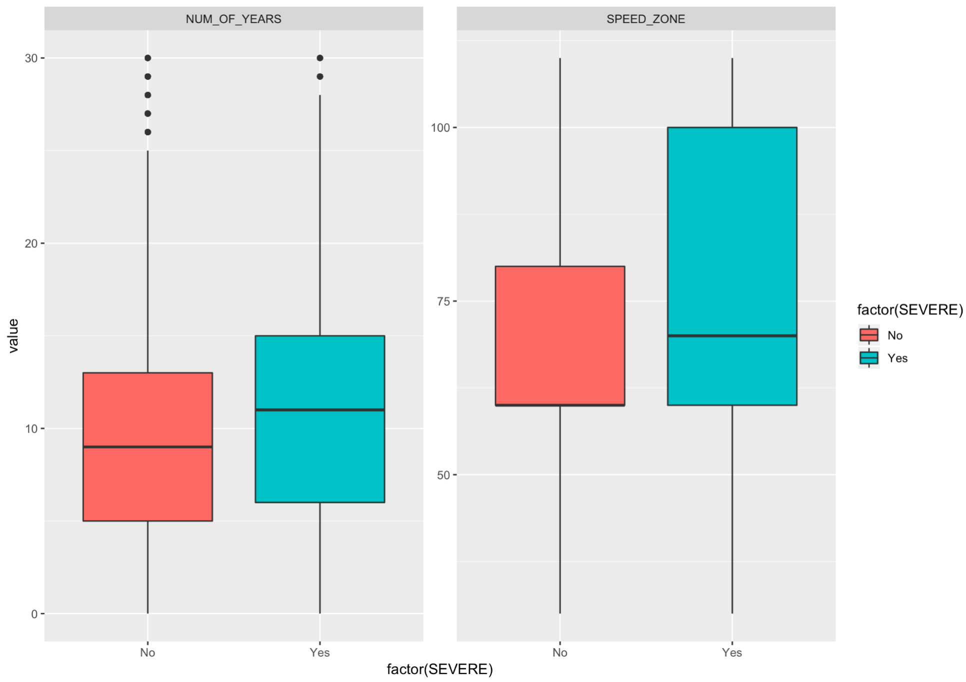

Figure 4: A box plot showing the distribution of continuous variables

Second test is to check for the ratios of the response variable, SEVERE. From Figure 5, it can be seen that there are no zeros, and also some of the ratios are large which may be due to the dataset being from the year 2006 to 2018. We will need to be careful with this during modeling our data.

	VEHICLE_MAKE	No	Yes	No_Percentage	Yes_Percentage
1	AUDI	885	84	91.33	8.669
2	B M W	1523	183	89.27	10.727
3	CHRYS	190	31	85.97	14.027
4	DAEWOO	719	187	79.36	20.640
5	DAIHAT	431	121	78.08	21.920
6	FORD	26546	5439	83.00	17.005
7	HOLDEN	33840	7299	82.26	17.742
8	HONDA	8592	1370	86.25	13.752
9	HYUNDAI	7562	1728	81.40	18.601
10	ISUZU	118	10	92.19	7.812
11	JEEP	819	91	90.00	10.000
12	KIA	2429	413	85.47	14.532
13	LAND R	371	71	83.94	16.063
14	LEXUS	291	21	93.27	6.731
15	M MOVE	202	36	84.87	15.126
16	MAZDA	9812	1692	85.29	14.708
17	MERC B	824	110	88.22	11.777
18	MITSUB	11503	2502	82.13	17.865
19	NISSAN	10451	1907	84.57	15.431
20	PEUG	331	51	86.65	13.351
21	REN	334	49	87.21	12.794
22	SUBARU	4020	631	86.43	13.567
23	SUZUKI	2041	440	82.27	17.735
24	TOYOTA	32456	5840	84.75	15.250
25	VOLKSW	3818	408	90.35	9.655
26	VOLVO	230	15	93.88	6.122

Figure 5: A table showing the number and percentage of Response variable

Now, regarding the levels in the data, a vehicle's make and Vehicle's model is a sample of wider population of model and make respectively. Hence, they can be used as levels for multilevel analysis. It is useful to compare between higher level units; hence we will be comparing the vehicle's makes.

The data has a 2-level nested structure hierarchy. The lowest level (Level 2) is vehicle's model, and highest level (Level 1) is the vehicle's make, as shown in Figure 6.

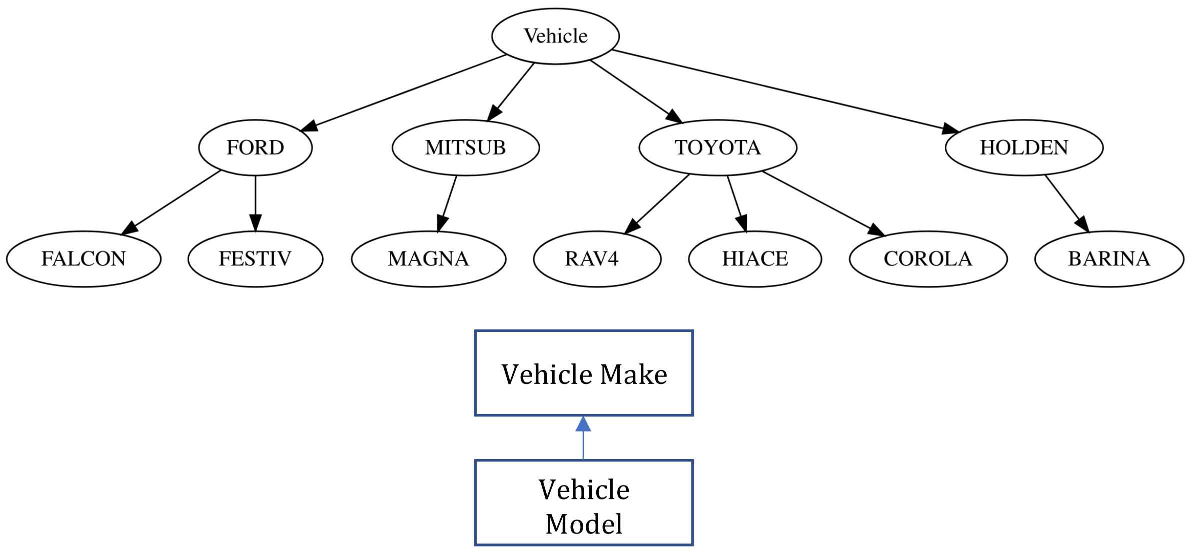

Figure 6: A plot and a relationship diagram showing 2-level hierarchy in the dataset

Thus, we will be taking the VEHICLE_MAKE and VEHICLE_MODEL as our two levels for multilevel modeling. There are no predictor variables for VEHICLE_MAKE in this dataset. For VEHICLE_MODEL level, will take LEVEL_OF_DAMAGE, SPEED_ZONE and NUM_OF_YEARS as predictor variables.

The response variable, SEVERE is binary, having values 'No' and 'Yes'. Hence, from the nature of the data and the response variable, the type of regression that we will be using is a logistic multilevel regression to access whether the vehicle's attributes play a significant impact on the severity of accidents or not.

Multilevel Model Design and Interpretation

Multilevel model treats units of analysis as independent observations and helps in drawing conclusions about the effects in those observations individually. The motivation to take up multilevel modeling can be due to the wish to make amendments to previously made inferences, or an interest in group effects of a research, or even to estimate the group effects using the group level predictors together. Whatever the motivation might be, there are a series of steps that are performed while doing multilevel regression. We will be performing each of those steps, taking inferences and moving onto the next until the best combination of variables is found.

We will be using `glmer` function of `lme4` package in R to compute the multi-level models with the optimizer `bobyqa` and value of nAGQ as 0, for optimizing the speed of the modeling. To compare the models, we will use the function AIC which stands for 'An Information Criterion' with log likelihood of the model. Lower value of AIC means a better model.

Now, the first step of doing multilevel modeling is to create a null model, with no effects, just the intercepts. A null model can be taken as the base model for analysis, and it will help to determine whether controlling the factor is useful or not.

From the data, we have more 'No' and less 'Yes'. The lower coded group is 'No' in this analysis.

```
Call:
glm(formula = SEVERE ~ 1, family = binomial(link = "logit"),
    data = train)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-0.592  -0.592  -0.592  -0.592   1.912

Coefficients:
            Estimate Std. Error z value            Pr(>|z|)
(Intercept) -1.65208    0.00623    -265 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 168537  on 191066  degrees of freedom
Residual deviance: 168537  on 191066  degrees of freedom
AIC: 168539

Number of Fisher Scoring iterations: 3
```

Figure 7: A null model with only the response variable SEVERE

After the null model, a random intercept model is made using `glmer` using random intercept of VEHICLE_MAKE. From this model, we can see that the intercept's estimate has changed towards the 'No' category. The variance due to random effect VEHICLE_MAKE is very small, so it is not significant.

```
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite
  Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: SEVERE ~ 1 + (1 | VEHICLE_MAKE)
   Data: train
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  168000   168020   -83998   167996   191065

Scaled residuals:
   Min     1Q Median     3Q    Max
-0.515 -0.464 -0.424 -0.399  3.305

Random effects:
 Groups       Name        Variance Std.Dev.
 VEHICLE_MAKE (Intercept) 0.0999   0.316
Number of obs: 191067, groups:  VEHICLE_MAKE, 26

Fixed effects:
            Estimate Std. Error z value            Pr(>|z|)
(Intercept)  -1.8361     0.0656     -28 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> AIC(logLik(m2))
[1] 168000
>
```

Figure 8: A random intercept model with VEHICLE_MAKE independent variable

Now, again adding a second random intercept VEHICLE_MODEL to the above model to note the changes in the variance of the intercept (Figure 9), we can see that the estimate again changes more towards 'No'. The variance of VEHICLE_MODEL is a little more than that of VEHICLE_MAKE, hence it can be said that VEHICLE_MODEL plays a bigger role.

```
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite
  Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: SEVERE ~ 1 + (1 | VEHICLE_MAKE) + (1 | VEHICLE_MODEL)
   Data: train
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  166726   166757   -83360   166720   191064

Scaled residuals:
   Min     1Q Median     3Q    Max
-0.703 -0.472 -0.427 -0.347  3.856

Random effects:
 Groups        Name        Variance Std.Dev.
 VEHICLE_MODEL (Intercept) 0.1366   0.370
 VEHICLE_MAKE  (Intercept) 0.0617   0.248
Number of obs: 191067, groups:  VEHICLE_MODEL, 204; VEHICLE_MAKE, 26

Fixed effects:
            Estimate Std. Error z value            Pr(>|z|)
(Intercept)  -1.8620     0.0624   -29.8 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> AIC(logLik(m2.1))
[1] 166726
>
```

Figure 9: A random intercept model with VEHICLE_MAKE and VEHICLE_MODEL independent variables

Adding fixed variables NUM_OF_YEARS and SPEED_ZONE to the model, we have slight change in the intercepts. The positive value of Estimate indicates that adding these variables has increases the chance of getting a 'Yes', though they are very small. This means that there is a slight increase in the chances of having a higher severity when the vehicle gets older, or when the vehicle is in the higher speed zone (essentially meaning that the vehicle is in higher speed). This is shown in Figure 10.

```
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite
  Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: SEVERE ~ NUM_OF_YEARS + SPEED_ZONE + (1 | VEHICLE_MAKE) + (1 |
    VEHICLE_MODEL)
   Data: train
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  154658   154709   -77324   154648   181755

Scaled residuals:
   Min      1Q Median      3Q     Max
-1.167 -0.464 -0.378 -0.295  5.632

Random effects:
 Groups        Name        Variance Std.Dev.
 VEHICLE_MODEL (Intercept) 0.0818   0.286
 VEHICLE_MAKE  (Intercept) 0.0443   0.211
Number of obs: 181760, groups:  VEHICLE_MODEL, 204; VEHICLE_MAKE, 26

Fixed effects:
              Estimate Std. Error z value            Pr(>|z|)
(Intercept)  -3.907021   0.059940   -65.2 <0.0000000000000002 ***
NUM_OF_YEARS  0.033773   0.001223    27.6 <0.0000000000000002 ***
SPEED_ZONE    0.025728   0.000366    70.3 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) NUM_OF
NUM_OF_YEAR -0.180
SPEED_ZONE  -0.441  0.008
> AIC(logLik(m3))
[1] 154658
```

Figure 10: A random intercept model with added Level 2 fixed variables

Now, the last thing to do is to add a random slope in the model. A subset of the dataset where the accident year is 2018 was sued to see the effect as the computation did not complete for the entire dataset. It can be seen that the NUM_OF_YEARS has had a little impact on the Variance of the Random Effects model, so it is also not that significant. Adding other intercept values also does not have considerable effect on the data. Hence, we will be excluding it from our analysis.

```
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite
  Quadrature, nAGQ = 0) [glmerMod]
 Family: binomial  ( logit )
Formula: SEVERE ~ NUM_OF_YEARS + SPEED_ZONE + (NUM_OF_YEARS | VEHICLE_MAKE) +
    (1 | VEHICLE_MODEL)
   Data: train.sub
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
    8252     8303    -4119     8238    11453

Scaled residuals:
   Min     1Q Median     3Q    Max
-1.229 -0.402 -0.318 -0.255  5.387

Random effects:
 Groups        Name         Variance Std.Dev. Corr
 VEHICLE_MODEL (Intercept)  0.000000 0.0000
 VEHICLE_MAKE  (Intercept)  0.018564 0.1363
               NUM_OF_YEARS 0.000265 0.0163   -1.00
Number of obs: 11460, groups:  VEHICLE_MODEL, 198; VEHICLE_MAKE, 26

Fixed effects:
             Estimate Std. Error z value             Pr(>|z|)
(Intercept)  -4.58950    0.13793  -33.27 < 0.0000000000000002 ***
NUM_OF_YEARS  0.05235    0.00693    7.55    0.000000000000042 ***
SPEED_ZONE    0.02941    0.00157   18.78 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) NUM_OF
NUM_OF_YEAR -0.489
SPEED_ZONE  -0.849  0.013
```

Figure 11: A random slope random intercept model applied on a subset of dataset to analyse the effects

We will now be taking the AIC to compare between our models and finding out the best model among them. Figure 12 shows the AIC values for all the models discussed above. It shows that the model `m3`, which takes NUM_OF_YEARS and SPEED_ZONE as the fixed variables and the random intercepts of our multilevel variables VEHICLE_MAKE and VEHICLE_MODEL gives the most efficient model.

```
> m1$aic
[1] 168539
> AIC(logLik(m2))
[1] 168000
> AIC(logLik(m2.1))
[1] 166726
> AIC(logLik(m3))
[1] 154658
>
```

Figure 12: Performance of models, compared using AIC

Though we have compared our models and found that our model m3 is the best fitting model, it has variance of intercepts only 0.081 for VEHICLE_MODEL and 0.0617 for VEHICLE_MAKE. Since the model is a same slope and random intercept model, the parallel regression lines which are drawn are separated by a very small distance. This implies that the regression lines drawn by using our independent variables will be very close to each other, which means that there is no effect of either VEHICLE_TYPE or VEHICLE_MODEL on our response variable individually. So, the answer to whether the vehicle's make or model impacts severity of accidents is negative.

Conclusion

The original study conducted by the group found that type of vehicle played a significant role in the severity of accidents. It had given an impression that except two wheelers, all the vehicle types were statistically significant. However, this assumption is contradicted by analysing deeper and performing multilevel modeling by taking a subset and using the vehicle's make and model. The reason that the previous study had given those to be statistically significant may be due to overfitting, since the data points were in the order of 180,000.

The main thing to take from this study is that the simple regression models, though would give a good approximation, may be incorrect. It is necessary to go deeper into details if one were to set a firm assumption on the model. Lastly, this study has come up with the conclusion that the vehicle's attributes are not statistically significant when it comes to the severity of the accidents.

References

Taimur, U., Liping, F., & F. M. M., Luis. (2016). *Injury severity analysis: comparison of multilevel logistic regression models and effects of collision data aggregation*. Retrieved from https://link.springer.com/article/10.1007/s40534-016-0096-4

Gudmundur, F., & Fred, L. M. (2004). *Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0001457502001355

Bureau of Infrastructure, Transport and Regional Economics 2018, *International road safety comparisons 2016*, BITRE, Canberra.

Magnusson, K. (2015). *Training for multilevel modeling*. Retrieved from http://www.bristol.ac.uk/cmm/learning/multilevel-models/samples.html

Taylor, J. J. (2011). *Confusing Stats Terms Explained: Residual*. Retrieved from http://www.statsmakemecry.com/smmctheblog/confusing-stats-terms-explained-residual.html

Blissett, R. (2017). *Multilevel Modeling in R*. Retrieved from https://rpubs.com/rslbliss/r_mlm_ws

Dr. Buchanan, E. M. (2019). *R - Binary Logistic Multilevel Models*. Viewed from https://www.youtube.com/watch?v=6MexZiX-2W8

Jovanovic, M. (2015). *R Playbook: Introduction to Multilevel/Hierarchical Models*. Retrieved from https://complementarytraining.net/r-playbook-introduction-to-multilevelhierarchical-models/

Borowski, T. G. *Chapter 16: Multilevel Modeling*. Retrieved from https://ademos.people.uic.edu/Chapter16.html

Appendices

Appendix I

```
                                                            Estimate Std. Error z value Pr(>|z|)    
(Intercept)                                                -1.481108   0.261928  -5.655 1.56e-08 ***
ACCIDENT.TYPE.DESCCollision with some other object         -0.682085   0.231043  -2.952 0.003155 ** 
ACCIDENT.TYPE.DESCCollision with vehicle                   -1.091767   0.071256 -15.322  < 2e-16 ***
ACCIDENT.TYPE.DESCFall from or in moving vehicle           -0.240552   0.277160  -0.868 0.385438    
ACCIDENT.TYPE.DESCNo collision and no object struck        -0.858902   0.129686  -6.623 3.52e-11 ***
ACCIDENT.TYPE.DESCOther accident                           -1.542601   1.105522  -1.395 0.162907    
ACCIDENT.TYPE.DESCStruck animal                            -1.355794   0.223325  -6.071 1.27e-09 ***
ACCIDENT.TYPE.DESCStruck Pedestrian                        -3.223902   0.308494 -10.450  < 2e-16 ***
ACCIDENT.TYPE.DESCVehicle overturned (no collision)        -0.805123   0.109362  -7.362 1.81e-13 ***
LIGHT_CONDITION2                                            0.148143   0.094907   1.561 0.118541    
LIGHT_CONDITION3                                            0.367797   0.057834   6.359 2.02e-10 ***
LIGHT_CONDITION5                                            0.521815   0.074925   6.964 3.30e-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  -5.200 2.00e-07 ***
SECURITY_EQUIPS_NOT_WORN                                    0.646444   0.074816   8.640  < 2e-16 ***
ROAD_TYPE_GROUPAVENUE                                       0.683313   0.253434   2.696 0.007013 ** 
ROAD_TYPE_GROUPDRIVE                                        0.669899   0.259062   2.586 0.009713 ** 
ROAD_TYPE_GROUPFREEWAY                                      0.050708   0.236471   0.214 0.830206    
ROAD_TYPE_GROUPHIGHWAY                                      0.664440   0.222732   2.983 0.002853 ** 
ROAD_TYPE_GROUPOTHER                                        0.510710   0.229739   2.223 0.026216 *  
ROAD_TYPE_GROUPROAD                                         0.487456   0.216819   2.248 0.024562 *  
ROAD_TYPE_GROUPSTREET                                       0.190352   0.223909   0.850 0.395250    
CAR_AGE                                                     0.030021   0.003057   9.820  < 2e-16 ***
Vehicle.Type.DescBus/Coach                                 -2.246087   0.406339  -5.528 3.25e-08 ***
Vehicle.Type.DescCar                                       -1.598060   0.096165 -16.618  < 2e-16 ***
Vehicle.Type.DescHeavy Vehicle (Rigid) > 4.5 Tonnes        -1.945070   0.247829  -7.848 4.21e-15 ***
Vehicle.Type.DescLight Commercial Vehicle (Rigid) <= 4.5 Tonnes GVM -1.989273   0.193751 -10.267  < 2e-16 ***
Vehicle.Type.DescMini Bus(9-13 seats)                      -3.474495   0.598475  -5.806 6.41e-09 ***
Vehicle.Type.DescMoped                                      1.793562   1.233744   1.454 0.146014    
Vehicle.Type.DescMotor Cycle                                0.128102   0.104250   1.229 0.219148    
Vehicle.Type.DescMotor Scooter                              0.397302   0.288404   1.378 0.168331    
Vehicle.Type.DescOther Vehicle                             -0.822224   0.473474  -1.737 0.082462 .  
Vehicle.Type.DescPanel Van                                 -1.803996   0.181585  -9.935  < 2e-16 ***
Vehicle.Type.DescPlant machinery and Agricultural equipment -2.010826   0.811952  -2.477 0.013267 *  
Vehicle.Type.DescPrime Mover - Single Trailer              -1.863824   0.255439  -7.297 2.95e-13 ***
Vehicle.Type.DescPrime Mover B-Double                      -2.157834   0.556935  -3.874 0.000107 ***
Vehicle.Type.DescPrime Mover B-Triple                      -2.550906   1.053847  -2.421 0.015496 *  
Vehicle.Type.DescPrime Mover Only                          -1.832172   0.505969  -3.621 0.000293 ***
Vehicle.Type.DescStation Wagon                             -1.933816   0.103510 -18.682  < 2e-16 ***
Vehicle.Type.DescTaxi                                      -1.627183   0.271357  -5.996 2.02e-09 ***
Vehicle.Type.DescUtility                                   -1.639282   0.116304 -14.095  < 2e-16 ***
OLD_COUNT                                                   0.496068   0.043751  11.338  < 2e-16 ***
ROAD_SURFACE_TYPE2                                         -0.250241   0.290449  -0.862 0.388927    
ROAD_SURFACE_TYPE3                                          0.353946   0.098549   3.592 0.000329 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16570  on 17601  degrees of freedom
Residual deviance: 13668  on 17557  degrees of freedom
  (14 observations deleted due to missingness)
AIC: 14477
```

Appendix I: The final model of our group's study on Road Accidents