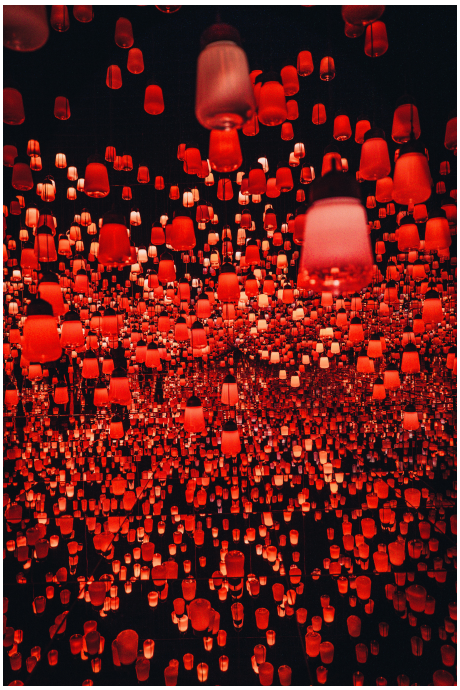

Assignment 4 ggplot2

Siyi Li
29018102

Yusen Wang
27496538

Alexi Tsakis
29682525

Report for
Australian Government

4 June 2021

Our consultancy
Alexi &
Siyi &
Yusen

(03) 9905 2478
questions@company.com

ABN: 12 377 614 630

1 Introduction

The seatbelts dataset (Harvey and Durbin 1986) is a collection of monthly data from 1968 to 1984 collected by the Department of Transportation. The dataset includes data on number of deaths - for car drivers, front seat passengers, rear seat passengers and van drivers - as well as distance driven, petrol price and a dummy variable that denotes whether the law mandating that front seat passengers and drivers wear seatbelts was in effect. In this report, we aim to use this data to provide insight into the following questions

- 1) Did seatbelts have a notable effect on number of deaths?
- 2) Is there any relationship between driver deaths and the other variables in this dataset?
- 3) What is the trend and trough of driver deaths?
- 4) Is there any seasonality to number of deaths?

2 Compulsory Seatbelts 1983 - Alexi Tsakis

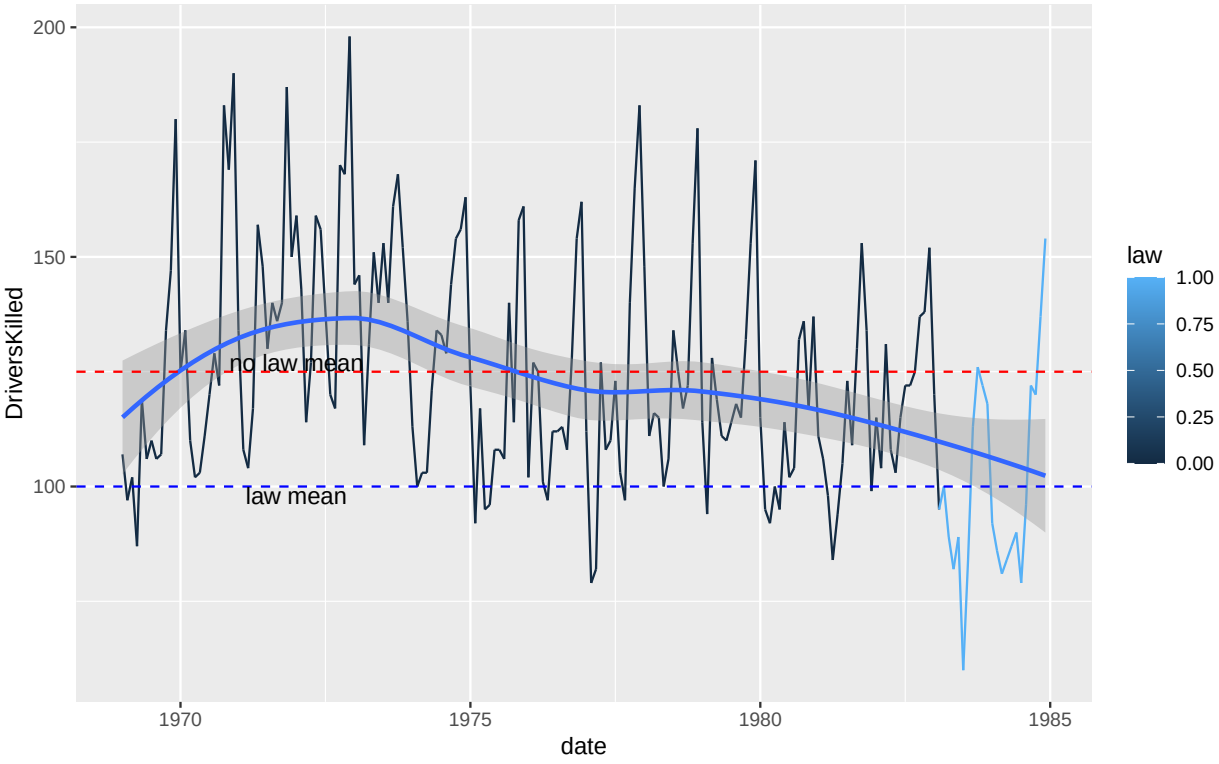

Figure 1: *Drivers killed in accidents over time*

In this section we look at the effect of seatbelts on driver deaths in the UK. Front seatbelts were compulsory on all new cars registered in the UK from 1968, however it was not required for them to

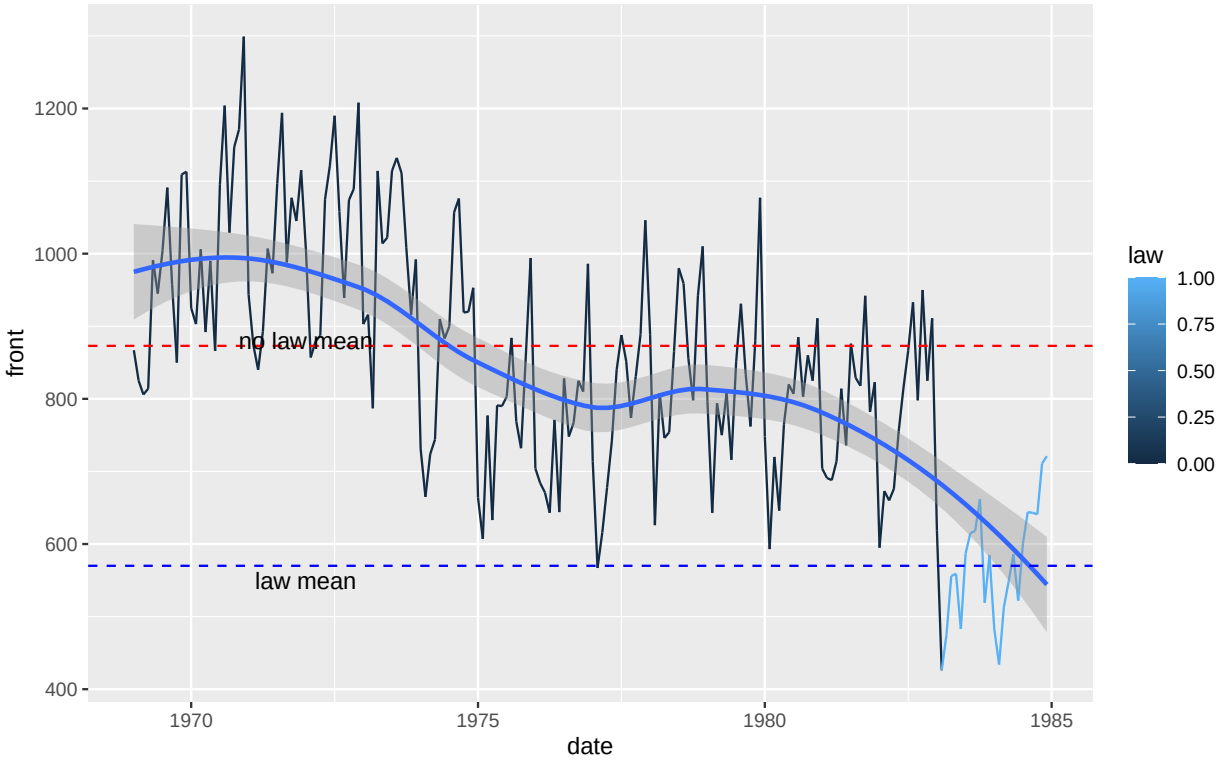

Figure 2: *Frontseat passengers killed in accidents over time*

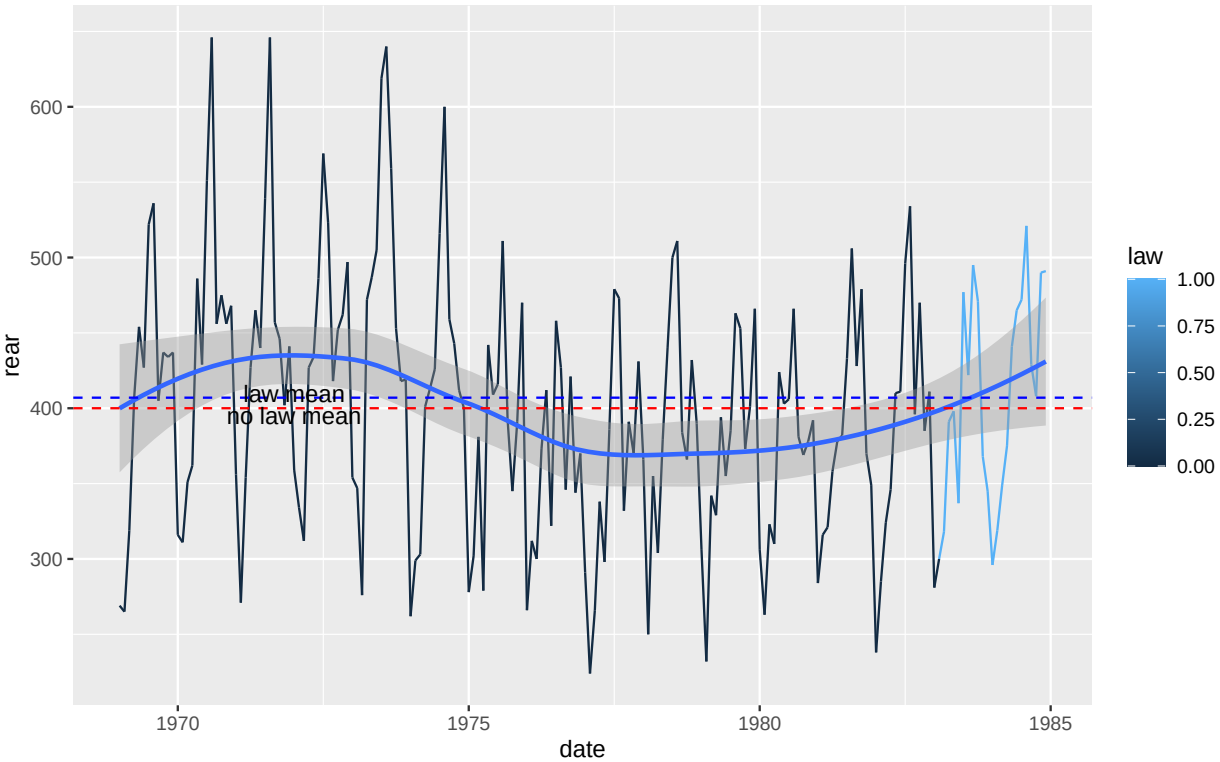

Figure 3: *Backseat passengers killed in accidents over time*

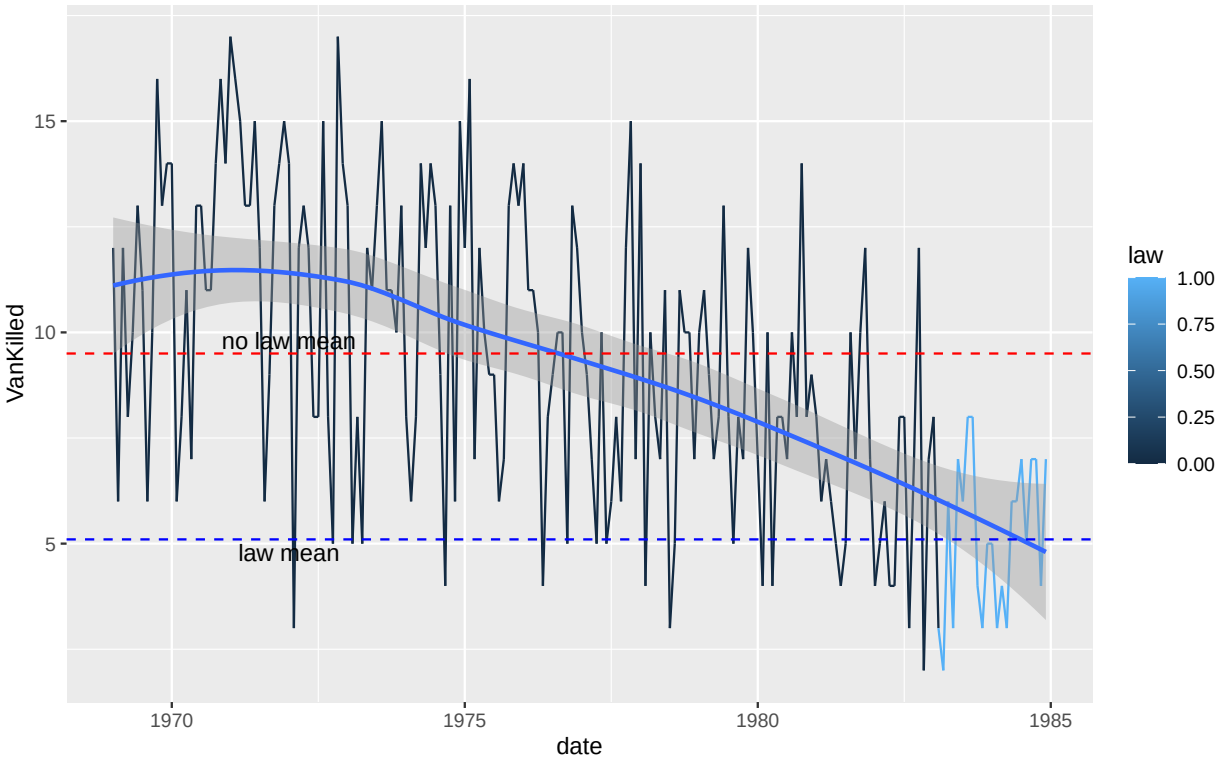

Figure 4: *Light goods vehicle drivers killed in accidents over time*

Table 1: *Drivers killed before and after law*

law	DriversKilled_Average
0	125.8698
1	100.2609

Table 2: *Back seat passengers killed before and after law*

law	Rear_Average
0	400.3195
1	407.7391

Table 3: *Front seat passengers killed before and after law*

law	Front_Average
0	873.4556
1	570.9565

Table 4: *Light good vehicle drivers killed before and after law*

law	Van_Average
0	9.585799
1	5.173913

be worn until 1983. In this section, we will answer the question - did mandatory wearing of frontrow seatbelts prevent deaths for drivers and passengers?

First, to answer if any relationships shown are causation and not simply correlation, we ask - why do seatbelts prevent deaths? According to the National Highway Traffic Safety Administration, the primary reasons why seatbelts prevent death is that “Buckling up helps keep you safe and secure inside your vehicle, whereas not buckling up can result in being totally ejected from the vehicle in a crash, which is almost always deadly.” (NHTSA 2021)

These scenarios are referred to as “secondary collisions” - i.e. a crash event that results from the impact of an initial collison, such as ejection from a vehicle, or even a high speed impact with objects inside the vehicle.

Now that we understand the theory behind the impact of seatbelts on preventing death, let us look at the data:

There is a clear negative correlation between frontrow passengers and light good vehicle drivers killed since the introduction of the law (figure 2, figure 4), data from the introduction of the law shown in blue).

There is also a clear reduction in number of drivers killed (figure 1) since the introduction of compulsory seatbelts - the lack of change since the introduction of the law, but the large change since approximately 1973 can be described as " drivers that are least likely to use seat belts might be those that are more likely to be involved in an accident" (Cohen and Einav 2003) - i.e. that careful drivers adopted seatbelts already without the law, and those likely to get involved in accidents would be unlikely to wear seatbelts even with the law in place.

There is little change in number of rear seat passengers killed (figure 3) - this is because the regulations and law were only for front seat seatbelts, and not backseat seatbelts, suggesting the total number of accidents didn’t change.

We can also see this in tabular format, so it is easier to see the effect of the law numerically (table 1, table 2, table 3, table 4).

Table 5: *Summary of the three variabls*

	DriversKilled	drivers	front
	Min. : 60.0	Min. :1057	Min. : 426.0
	1st Qu.:104.8	1st Qu.:1462	1st Qu.: 715.5
	Median :118.5	Median :1631	Median : 828.5
	Mean :122.8	Mean :1670	Mean : 837.2
	3rd Qu.:138.0	3rd Qu.:1851	3rd Qu.: 950.8
	Max. :198.0	Max. :2654	Max. :1299.0

3 Reseach Questions - Siyi Li

- Is there any relationship between the DriversKilled variable and other variables?
- Can we build an model among these variables?

4 Analysis

4.1 Is there any relationship between the DriversKilled variable and other variables?

From the table 5 shows:

- the minimum number of drivers killed is 60 and the maximum is 198.
- the minimum number of drivers is 1057 and the maximum is 2654.
- the minimum number of front-seat passengers killed or seriously injured is 426 and the maximum is 1299.
- when the number of the drivers increase, the number of the drivers killed also increase. Therefore, I guess there is a positive relationship between the variable drivers and the variable drivers killed.

In order to develop the relationship between the number of drivers and car drivers killed, we can use GGally package (Schloerke et al. 2020) to produce the correlation graph.

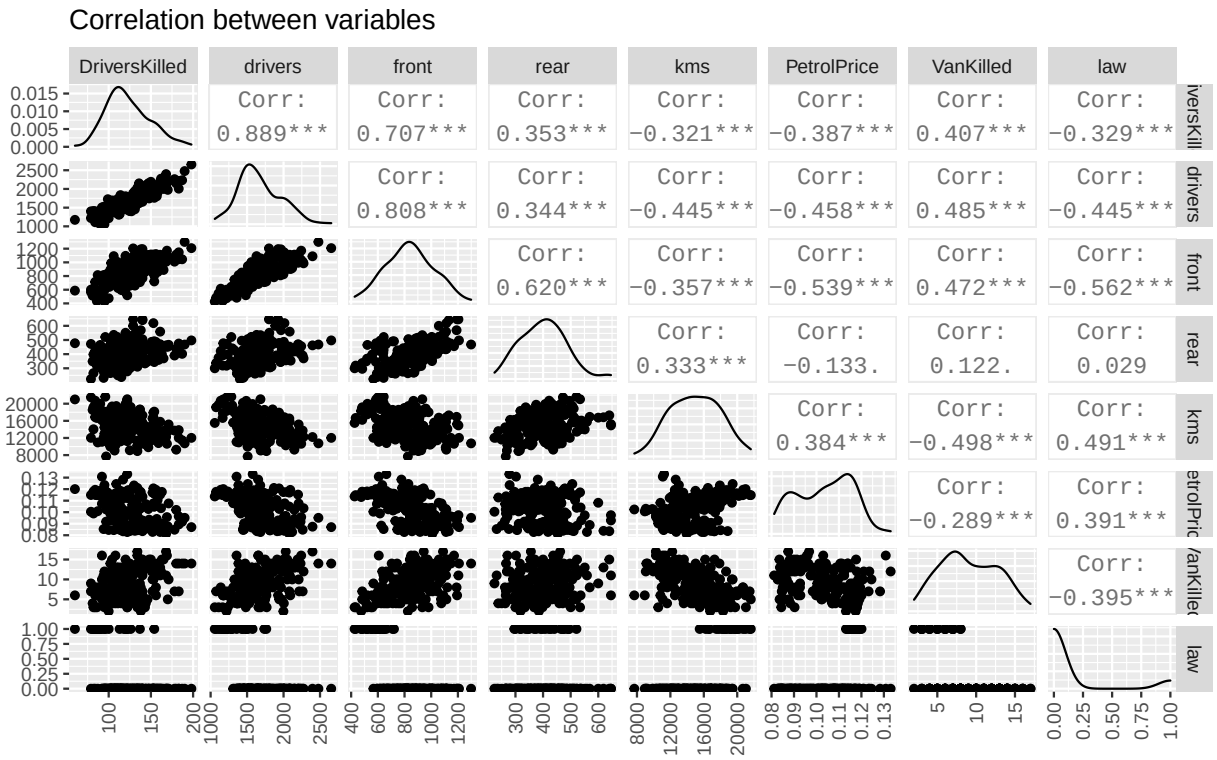

Figure 5: *Correlation between variables*

- The figure 5 shows that there are positive relationships between drivers variable and Driverskilled variable, front variable and Driverskilled variable, VanKilled and Driverskilled, rear variable and Driverskilled variable. Especially, we can see that the variables of front and drivers are highly positive related to the Driverskilled variable since the correlations between that are 0.707 and 0.889 respectively.

- It also indicates that the relationships between kms and Driverskilled variable, PetrolPrice and Driverskilled, and law and Driverskilled are negative.

- In fact, there are some correlations between each pair of variables but i just focus on the relationship between the response variable Driverskiied and other predictor.

4.2 Can we build an model among these variables?

```
##
## Call:
## lm(formula = DriversKilled ~ ., data = Seatbelt)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -29.489  -7.620  -0.531   7.676  34.700
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.330e+01  1.472e+01  -1.583    0.115
## drivers      8.327e-02  5.544e-03  15.018   <2e-16 ***
## front       -3.838e-03  1.723e-02  -0.223    0.824
## rear         5.189e-03  2.474e-02   0.210    0.834
## kms          5.875e-04  5.025e-04   1.169    0.244
## PetrolPrice -1.629e+01  8.643e+01  -0.188    0.851
## VanKilled    5.928e-02  2.911e-01   0.204    0.839
## law1         4.070e+00  4.076e+00   0.998    0.319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.58 on 184 degrees of freedom
## Multiple R-squared:  0.7993, Adjusted R-squared:  0.7917
## F-statistic: 104.7 on 7 and 184 DF,  p-value: < 2.2e-16
```

The linear model shows that just the drivers variable are significant which means that the drivers are fit the model. Other variables include the Intercept are not significant since their p-value are larger than 0.05, which means that the model can not fit all the variables very well.

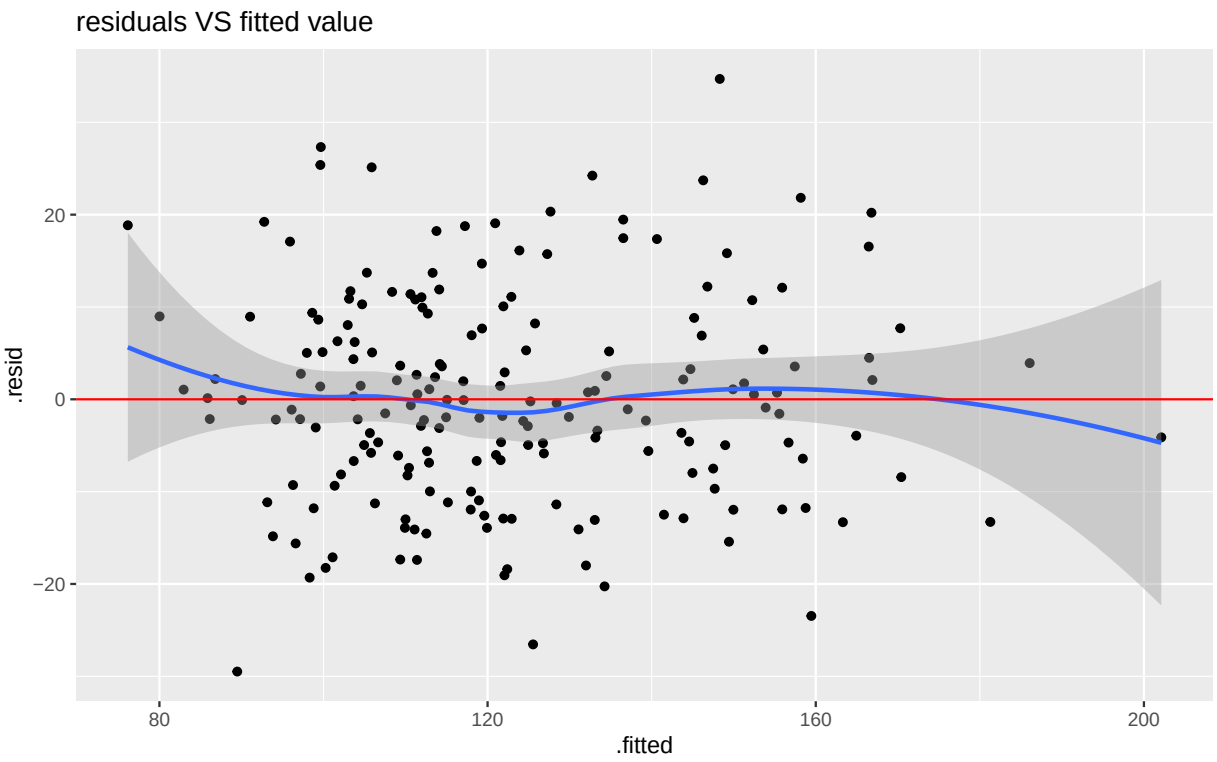

Figure 6: *residuals VS fitted value*

In order to analyze the residuals, I use the **broom** package(Robinson, Hayes, and Couch 2020).

From the figure 6, it indicates that the residuals does not include obvious pattern of the variables since the residuals are randomly around the red line. It means that the model can capture most information about the response variable DriversKilled. But we can see that the tail at the beginning and end is not like a straight since the observations are not enough at the end and the beginning.

5 Research Questions - Yusen Wang

- From the data on the number of driver deaths, which time period was the peak, and what historical facts accompany the change after the peak ? How about the trend before and after law introduced ?
- Is there any relationship between season and driverskilled?

6 Data Analysis

6.1 Analysis of trends in driverkilled before and after 1983

According to Figure 7, as we can see:

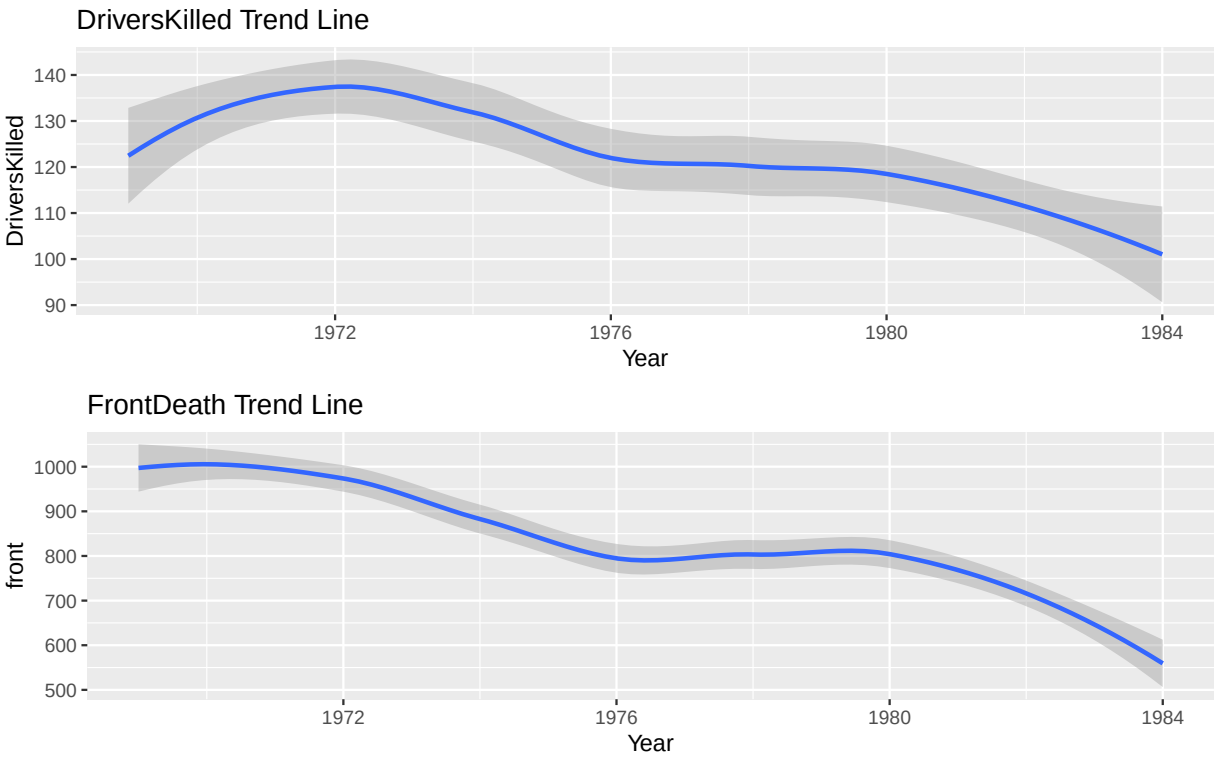

Figure 7: *Trend Lines*

- the trend line of DriversKilled and FrontDeath after 1970 began to decline.

The reason is Successive UK governments proposed, but failed to deliver, seat belt legislation throughout the 1970s. Front seat belts were compulsory equipment on all new cars registered in the UK from 1968, although it did not become compulsory for them to be worn until 1983. (Richens, Imrie, and Copas 2000)

Table 6: *Before Law summary stats*

	Year	Month	DriversKilled	drivers	front	rear	kms
	Min. :1969	Jan :15	Min. : 79.0	Min. :1309	Min. : 567.0	Min. :224.0	Min. : 768
	1st Qu.:1972	Feb :14	1st Qu.:108.0	1st Qu.:1511	1st Qu.: 767.0	1st Qu.:344.0	1st Qu.:12
	Median :1976	Mar :14	Median :121.0	Median :1653	Median : 860.0	Median :401.0	Median :1
	Mean :1976	Apr :14	Mean :125.9	Mean :1718	Mean : 873.5	Mean :400.3	Mean :144
	3rd Qu.:1979	May :14	3rd Qu.:140.0	3rd Qu.:1926	3rd Qu.: 986.0	3rd Qu.:454.0	3rd Qu.:16
	Max. :1983	Jun :14	Max. :198.0	Max. :2654	Max. :1299.0	Max. :646.0	Max. :210
	NA	(Other):84	NA	NA	NA	NA	NA

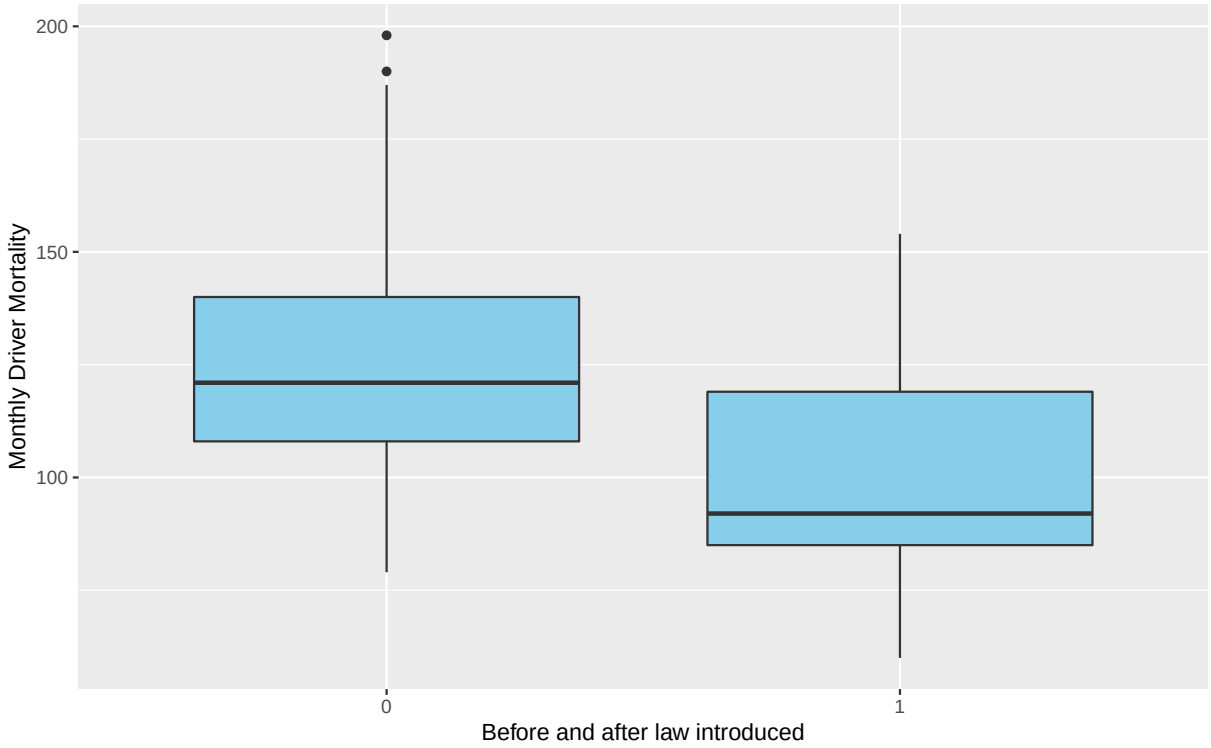

Figure 8: *Law Introduction*

From the figure 8, it is clear that the range of monthly deaths among drivers before the law was higher than after it was enacted.

From the tables 6 and 7 :

- Before law introduced, the median of DriversKilled is 121, the minimum number is 79 and the maximum number is 198.
- After introduced, the median of DriversKilled is 92, the minimum number is 60 and the maximum number is 154.

Table 7: *After Law summary stats*

	Year	Month	DriversKilled	drivers	front	rear	kms
	Min. :1983	Feb : 2	Min. : 60.0	Min. :1057	Min. :426.0	Min. :296.0	Min. :1551
	1st Qu.:1983	Mar : 2	1st Qu.: 85.0	1st Qu.:1171	1st Qu.:516.0	1st Qu.:347.0	1st Qu.:179
	Median :1984	Apr : 2	Median : 92.0	Median :1282	Median :585.0	Median :408.0	Median :191
	Mean :1984	May : 2	Mean :100.3	Mean :1322	Mean :571.0	Mean :407.7	Mean :1889
	3rd Qu.:1984	Jun : 2	3rd Qu.:119.0	3rd Qu.:1464	3rd Qu.:629.5	3rd Qu.:471.5	3rd Qu.:199
	Max. :1984	Jul : 2	Max. :154.0	Max. :1763	Max. :721.0	Max. :521.0	Max. :2162
	NA	(Other):11	NA	NA	NA	NA	NA

We can see that the law is effective in helping drivers stay away from death, since it is introduced on 31st January 1983.

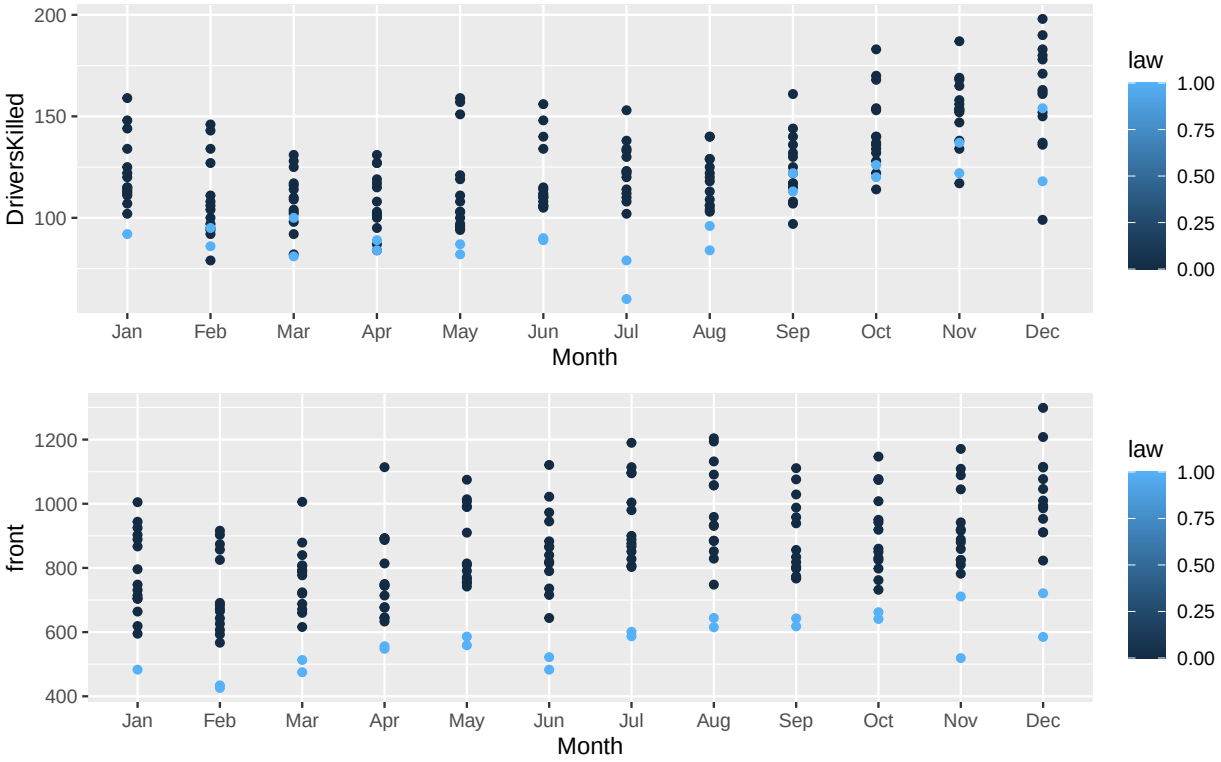

Figure 9: *Seasonality*

From the figure 9, there seems to be some seasonal correlation between driver deaths:

- it is quite obvious that, there is more driver deaths and front deaths in autumn/winter than in spring/summer, somehow it is with seasonal effect. Snow, rain, fog, and other bad weather is the main reason, not only affect the driver's vision, but also increase the braking distance.
- based on the color plots, it can be seem that the number of driver deaths is the lowest after the front seatbelt legislation been introduced.

7 Conclusion

Based on the analysis done in this report, our findings were as follows

- The mandatory seatbelt production law of 1968, as well as the mandatory seatbelt wearing law of 1983 did have a statistically significant impact in reducing number of drivers, van drivers and front seat passengers killed. The laws slightly increased number of rear seat passengers killed, but not significantly so.
- There is no statistically significant relationship between DriversKilled and any other variable in the dataset except number of drivers, as such we can't model DriversKilled very well as Drivers does not explain enough of the variance.
- The trough of drivers killed is after the seatbelt legislation was introduced.
- There is more deaths in autumn and winter due to weather conditions.

---

8 Packages used

Wickham et al. 2019 Wickham and Hester 2020 Zhu 2021 Xie 2021 Auguie 2017

References

Auguie, B (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3. https://CRAN.R-project.org/package=gridExtra.

Cohen, A and L Einav (2003). The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities. *The Review of Economics and Statistics* **85**(4), 828–843.

Harvey, AC and J Durbin (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society* **149**(Series A), 187–227.

NHTSA (2021). Seat Belts.

Richens, J, J Imrie, and A Copas (2000). Condoms and seat belts: the parallels and the lessons. *The Lancet* **355**(9201), 400–403.

Robinson, D, A Hayes, and S Couch (2020). *broom: Convert Statistical Objects into Tidy Tibbles*. R package version 0.7.0. https://CRAN.R-project.org/package=broom.

Schloerke, B, D Cook, J Larmarange, F Briatte, M Marbach, E Thoen, A Elberg, and J Crowley (2020). *GGally: Extension to 'ggplot2'*. R package version 2.0.0. https://CRAN.R-project.org/package=GGally.

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.

Wickham, H and J Hester (2020). *readr: Read Rectangular Text Data*. R package version 1.4.0. https://CRAN.R-project.org/package=readr.

Xie, Y (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.22. https://github.com/rstudio/bookdown.

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.