Does Covid encourage assaults?*

A statistical analysis of how the rise of Covid in 2020 affected assault rates in various neighborhoods in Toronto

Varun Vijay

27 April 2022

Abstract

This paper presents an analysis on the effect of Covid on assault rates in Toronto neighbourhoods in 2020. The paper uses datasets obtained from opendatatoronto to show the Covid's positive effect on the number assaults in a neighbourhood. The analysis consists of using the statistical programming language R to model the data using regression and constructing various plots and graphs to visually represent the data. THe results obtained will allow us to better predict people's behavior. With the better prediction of people's behaviour, the police can better manage the public order and government will be able to make reasonable and effective policy to reduce the crime rate.

Keywords: Covid, Assaults, Crime,Toronto neighbourhoods, residents, open data toronto

Contents

*Code and data are available at: https://github.com/Varun1005473462/final_folder-main.git.

1 Introduction

After the first reported outbreak of Covid in December 2019, the entire world changed. Covid was quickly declared to be a health emergency of global concern by the World health organization in March of 2020. Governments took measures to restrict the spread of the disease such as imposing lockdowns, travel bans, restrictions on public gatherings, closing down businesses and advising people to isolate at home. Citizens of countries were essentially restricted to their homes and rarely left. Covid had a huge impact on the global economy and had restricted people's actions and movements. However, this paper wants to analyse how Covid has affected one of the most common types of crime, assaults, which are often physical in nature. The question this paper seeks to answer is how the Covid pandemic had affected assault rates in various neighbourhoods in Toronto for the year 2020.

In this paper, we use data on Covid cases and crime statistics in Toronto neighbourhoods for the year 2020 to see how the number of Covid cases in a neighbourhood had affected assaults and what factors affected the assault rate the most. We constructed a linear model using R (R Core Team 2020) to model covid's effect on assaults. We also useed the dataset to construct various plots and summary statistics about the dataset to improve the quality of our analysis. The results obtained seem to imply that the Covid pandemic had a positive effect on assaults in 2020 as there is a positive relation between number of covid cases in a neighbourhood and number of assaults. The results obtained contribute to our understanding of how the presence of a common threat that affects everyone such as a pandemic in this case can discourage people from criminal actions. It will allow us to improve our predictions of a person's behavior during such situations. Better predictions of people's behaviors will allow the police to better manage the public order and government to make reasonable and effective policy to reduce the crime rate

The paper first presents an overview of the datasets and the variables we will be using for this study in the data section. The model section presents the regression steps taken to generate the model that shows how Covid affects crime. The results section discusses the various conclusions we can draw from the data and the model and finally the discussion section discusses the strengths and weaknesses of the paper. Any additional details are included in the appendix.

2 Data

2.1 Data Source and collection

The relevant datasets were obtained using opendatatoronto and show statistics related to Covid cases and to crime figures. The first dataset is about Covid cases in Toronto and records all Covid cases reported since January 2020. The dataset includes all cases related to outbreaks and sporadic cases, those that occur in the community. The data was extracted from the provincial Case & Contact Management System and the variables in the dataset include the neighbourhood the case was reported in, the date it was reported, the client gender, Age group and the classification of the case, whether it was probable or improbable. The full dataset **??** can be found in Appendix B.

The second dataset used for this paper was the dataset related to crime statistics in Toronto neighbourhoods. The dataset contains a list of crime related statistics by neighbourhood and includes figures such as number of assaults, number of car thefts, number of shootings, etc. for the years 2014 to 2020. We are only interested in the number of assaults for 2019 and 2020. The full dataset **??** can be found in Appendix B.

The datasets were extracted using R (R Core Team 2020) and R packages opendatatoronto (Gelfand 2020), tidyverse (Wickham et al. 2019a), dplyr (Wickham et al. 2021) and knitr (Xie 2021a). The data was then processed, manipulated and analysed for the purposes of this paper.

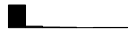
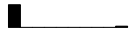
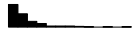
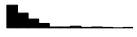
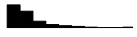
2.2 Data Modification

For the purposes of this paper, we are interested only in covid cases that were reported during 2020 since that was when the pandemic first started. So the only cases selected were those reported during 2020. We then focused on the variables Client Gender and the age group of the infected. These variables are the ones that pertain the most to the total number of covid cases in a neighbourhood and so are the ones I will be focusing on for this dataset as they also more important than the others. New variables using the dataset were created such as number of female covid cases, number of male covid cases, number of transgender covid cases, number of other genders and unknown genders covid cases and variables related to the various age groups. We also created a new variable where we counted the total number of Covid cases in a neighbourhood. However, we will be excluding all age groups greater than or equal to 50 since it is unlikely that anyone of that age will be assaulting people as they are mostly middle aged or old people.

For the crime dataset, we only focused on the neighbourhoods and the number of assaults for 2019 and for 2020 since these pertain directly to my paper. We then joined the two datasets together and obtained the final_dataset. All 140 neighbourhoods are included in this dataset and it is the dataset we will be using for the rest of this paper.

2.3 Dataset for this paper

```
## # A tibble: 140 x 13
##    Neighbourhood     Assault_2019 Assault_2020 FEMALE_Covid_CA~ MALE_Covid_CASES
##    <chr>                    <dbl>        <dbl>            <int>            <int>
##  1 Agincourt North             81           69              110               82
##  2 Agincourt South-~          121          132               55               71
##  3 Alderwood                   36           37               47               44
##  4 Annex                      305          275              128              118
##  5 Banbury-Don Mills           73           83               57               52
##  6 Bathurst Manor              59           58              135              103
##  7 Bay Street Corri~          894          516               57               76
##  8 Bayview Village             67           69               42               43
##  9 Bayview Woods-St~           45           33              103               52
```

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
Assault_2019	114	0	147.3	148.9	16.0	99.5	911.0	
Assault_2020	107	0	127.5	127.5	12.0	87.5	761.0	
FEMALE_Covid_CASES	101	0	103.2	86.5	16.0	74.0	412.0	
MALE_Covid_CASES	98	0	93.7	82.6	15.0	67.5	423.0	
transgender_covid_cases	2	0	0.0	0.2	0.0	0.0	1.0	
unknown_covid_cases	10	0	1.0	2.0	0.0	0.0	14.0	
other_covid_cases	2	0	0.0	0.2	0.0	0.0	1.0	
<=19	49	0	18.8	19.9	0.0	12.0	126.0	
20 to 29	68	0	38.6	38.7	3.0	27.5	228.0	
30 to 39	63	0	32.0	29.0	2.0	23.0	187.0	
40 to 49	61	0	27.2	25.9	2.0	19.0	153.0	
Total_CASES	113	0	197.9	167.9	33.0	140.0	829.0	

```
## 10 Bedford Park-Nor~           57           51              100              106
## # ... with 130 more rows, and 8 more variables: transgender_covid_cases <int>,
## #   unknown_covid_cases <int>, other_covid_cases <int>, `<=19` <int>,
## #   `20 to 29` <int>, `30 to 39` <int>, `40 to 49` <int>, Total_CASES <int>
```

The final dataset @ref(table:dataset_final) consists of 140 observations and 13 variables: Neighbourhood: Name of the neighbourhood Assault_2019: number of assaults in that neighbourhood for the year 2019 Assault_2020: number of assaults in that neighbourhood for the year 2020 FEMALE_Covid_CASES: Number of reported covid cases in that neighbourhood for females MALE_Covid_CASES: Number of reported covid cases in that neighbourhood for males transgender_Covid_CASES: Number of reported covid cases in that neighbourhood for transgenders unknown_covid_cases: Number of reported covid cases in that neighbourhood for people of unknown genders other_covid_cases: Number of reported covid cases in that neighbourhood for people of other genders <=19: number of reported covid cases for people less than 19 years old 20 to 29: number of reported covid cases for people between 20 and 29 30 to 39: number of reported covid cases for people between 30 and 39 40 to 49: number of reported covid cases for people between 40 and 49 Total_CASES: total number of covid cases in a neighbourhood

These 13 variables are the ones that are most pertinent to the topic of this paper

2.4 Summary statistics

The above table @ref(table: summary 1) shows summary statistics for the dataset. As you can see, on average, the number of assaults per neighbourhood in Toronto decreased from 147.3 to 127.5 from 2019 to 2020. It is possible that this decrease in assaults per neighbourhood is because of Covid. Furthermore, the number of covid cases on average in a neighbourhood is 197.9 with a standard deviation of 167.9 so each neighbourhood has a varying amount of covid cases. Similarly the standard deviation for assaults in 2020 is 127.5 which means that there is a chance we will be able to obtain a relation between covid and assaults.

2.5 Correlation statistics for the final dataset

From the above table @ref(table: summary 2) we can see that the variables with the strongest correlation with assaults in 2020 are number of female covid cases, number of male covid cases, number of covid cases for age groups less than or equal to 19, 20 to 29, 30 to 39 and 40 to 49. It makes for these variables to relate

	Assault_2019	Assault_2020	FEMALE_Covid_CASES	MALE_Covid_CASES	trans
Assault_2019	1	.	.	.	
Assault_2020	0.96	1	.	.	
FEMALE_Covid_CASES	0.45	0.46	1	.	
MALE_Covid_CASES	0.54	0.55	0.95	1	
transgender_covid_cases	−0.04	−0.03	−0.08	−0.06	
unknown_covid_cases	0.25	0.31	0.33	0.32	
other_covid_cases	0.12	0.15	0.10	0.15	
<=19	0.32	0.31	0.84	0.87	
20 to 29	0.62	0.62	0.85	0.92	
30 to 39	0.63	0.66	0.86	0.91	
40 to 49	0.43	0.45	0.91	0.93	
Total_CASES	0.50	0.51	0.99	0.99	

to number of assaults in a neighbourhood since they are the most important ones since age and gender are incredibly important in such situations when an assault is occurring. We also see that age groups and the number of male and female covid cases are strongly related.

Fig @ref(fig:figure 1) shows us how the number of assaults in 2019 compare to the number of assaults in 2020 per neighbourhood. There does not appear to be a significant difference between the values for each neighbourhood, however, it does seem like most of the assaults in 2020 were less than the ones in 2019 which again might imply that covid does affect number of assaults in a neighbourhood.

2.6 Covid cases by neighbourhood

As we can see from figure @ref(fig:figure 2), the number of covid cases per neighbourhood varies and are not similar to each other. So there is variation which means there's a possibility the number of assaults might also vary with the covid cases.

3 Model

The model I will be using for this paper is:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \beta_4 * x_4$$

y= number of assaults in a neighbourhood x_1 = number of female covid cases in a neighbourhood x_2 = number of male covid cases in a neighbourhood x_3 = number of covid cases between ages 30 and 39 x_4 = number of covid cases between ages 40 and 49

Here, the number of assaults is the explanatory variable and the number of female covid cases, number of make cases, number between 30 and 39 and number between 40 and 49 are the independent variables.

The reason I have chosen these 4 variables is because from the correlation table @ref(table: summary 2) in the data section, these variables have some of the strongest correlation with the number of assaults in 2020.

Additionally, most crimes would not be committed by people older than 50 and would normally be committed by those in the range of 30 to 49 as those who are younger than 30 are still trying to figure out their lives and might not turn to crime yet and those older than 50 are middle aged and old people who would not normally commit crimes. Additionally, gender is one of the most important factors in today's world and is an important variable to include. It is always important to take gender into account as it can often be a confounding variable if you do not include it in the model. I have excluded the transgender and unknown variable from the model because those entries were far less than the number of entries for male and female.

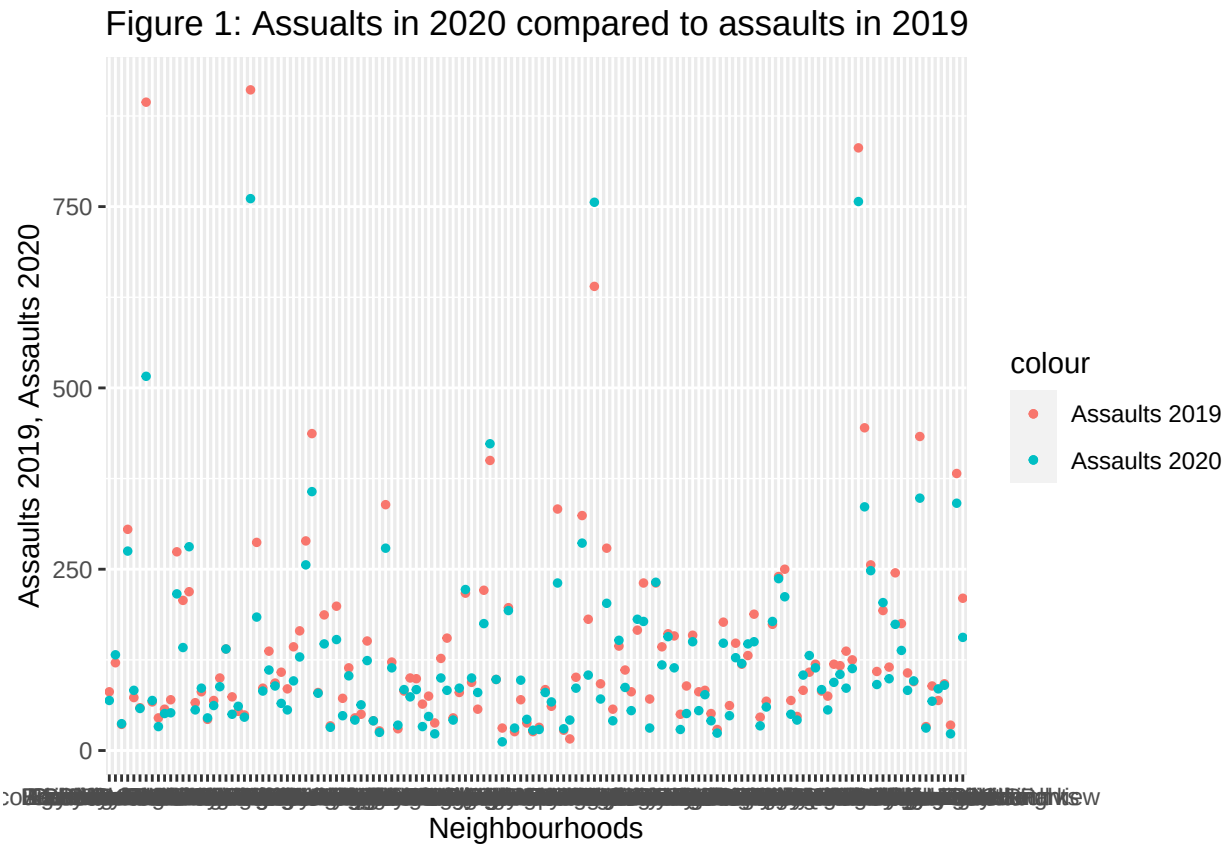


Figure 1: (#fig:figure 1)Assualts in 2020 compared to assaults in 2019

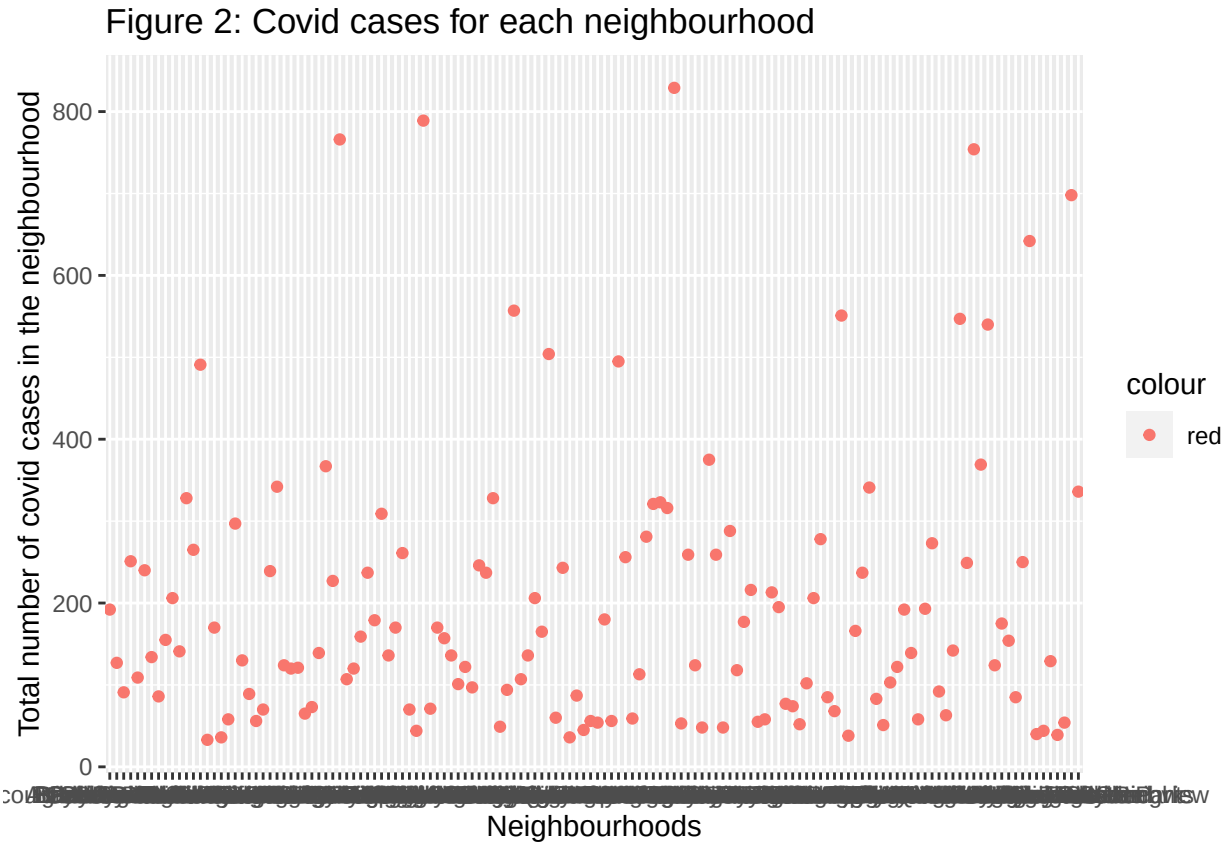


Figure 2: (#fig:figure 2)Covid cases for each neighborhood in 2020

4 Results

Using modelsummary (Arel-Bundock et al. 2022) and AIC selection process, I was able to obtain my model and the estimates for the model such as intercept, R^2, estimates for independent variables etc which are included in the table **??**table: model estimates).

The AIC process and the code for the estimates are located in tables **??**model). By the process of forwards AIC, I was able to obtain the final model by removing variables from the model to see if it reduced the AIC value. Once I obtained the lowest AIC value I could obtain, I had obtained my final model.

4.1 Table of estimates

```
##                          Terms   values
## 1                    Intercept   47.418
## 2  Female_covid_cases estimate   -0.858
## 3    Male_covid_cases estimate    1.059
## 4            30 to 39 estimate   13.834
## 5            40 to 49 estimate   -1.964
## 6        Number of observation  140.000
## 7                          R^2    0.508
## 8                 R^2 adjusted    0.493
## 9                          AIC 1666.500
## 10                         BIC 1684.100
## 11                    Log.Lik -827.243
## 12                          F   34.810
```

4.2 Final model

So, using the estimates obtained, the model we have is:

$$y = 47.418 - 0.858 * x_1 + 1.059 * x_2 + 13.834 * x_3 - 1.964 * x_4$$

This shows that there appears to be more of a positice relation between number of assaults and covid cases in a neighbourhood. When there is a unit increase in number of females who have covid, there is 0.858 decrease in the number of assaults in that neighbourhood. However, when there is a unit increase in the number of males with covid, there is a 1.059 increase in the number of assaults in the neighbourhood. When there is an unit increase in the number of people between 30 and 39 who get covid, there is a 13.834 increase in the number of assaults. However, when there is an unit increase in the number of people between 40 and 49 who get covid, there is only a 1.964 decrease in the number of assaults.

4.3 Final model graphs

From figure 3 @ref(fig:figure 3), we can see that there appears to be a sort of positive relation between the total number of cases in a neighbourhood and the number of assaults in that neighbourhood for the year 2020. Except for a few outliers, it seems that when cases are low, assaults are low and when cases increase, assaults also increase.

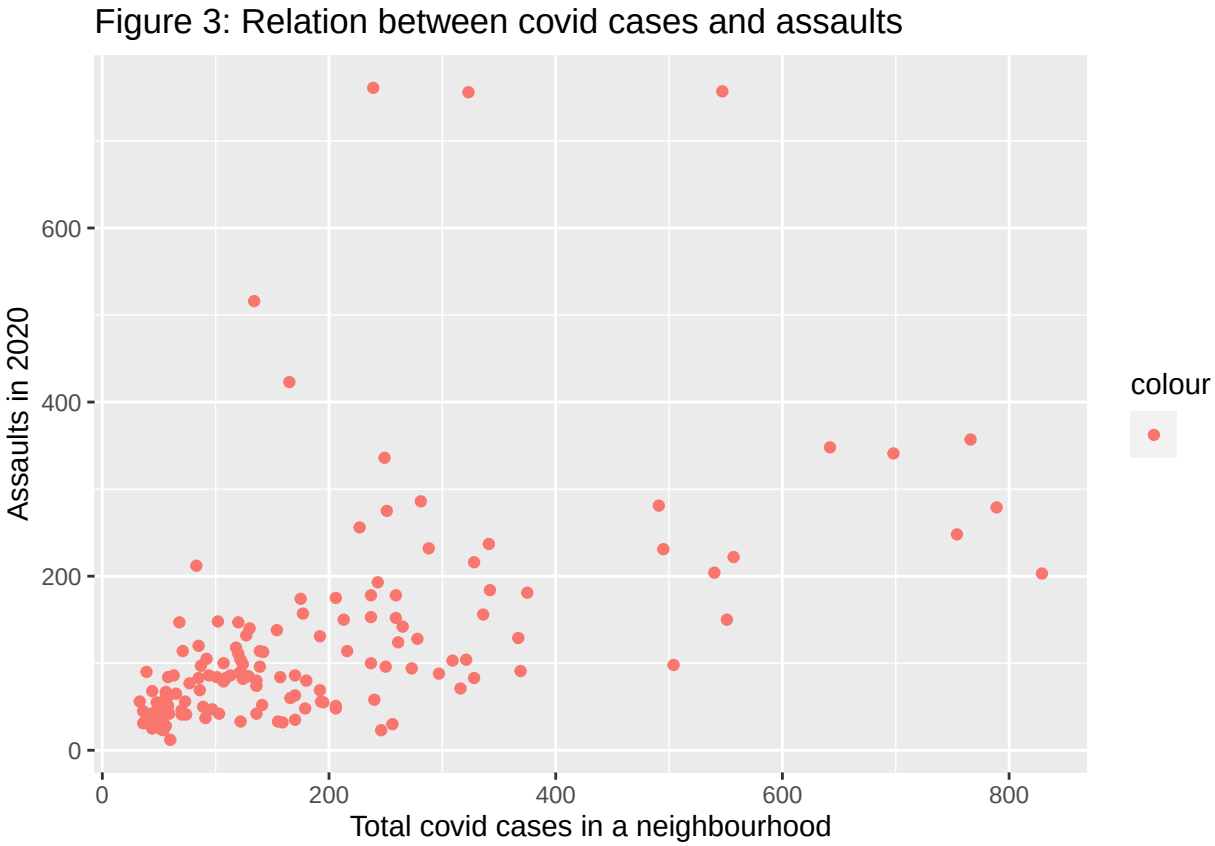


Figure 3: (#fig:figure 3)Relationship between Covid and Assault rates in 2020

Figure 4: Relation between number of assaults in 2020 and the number of f

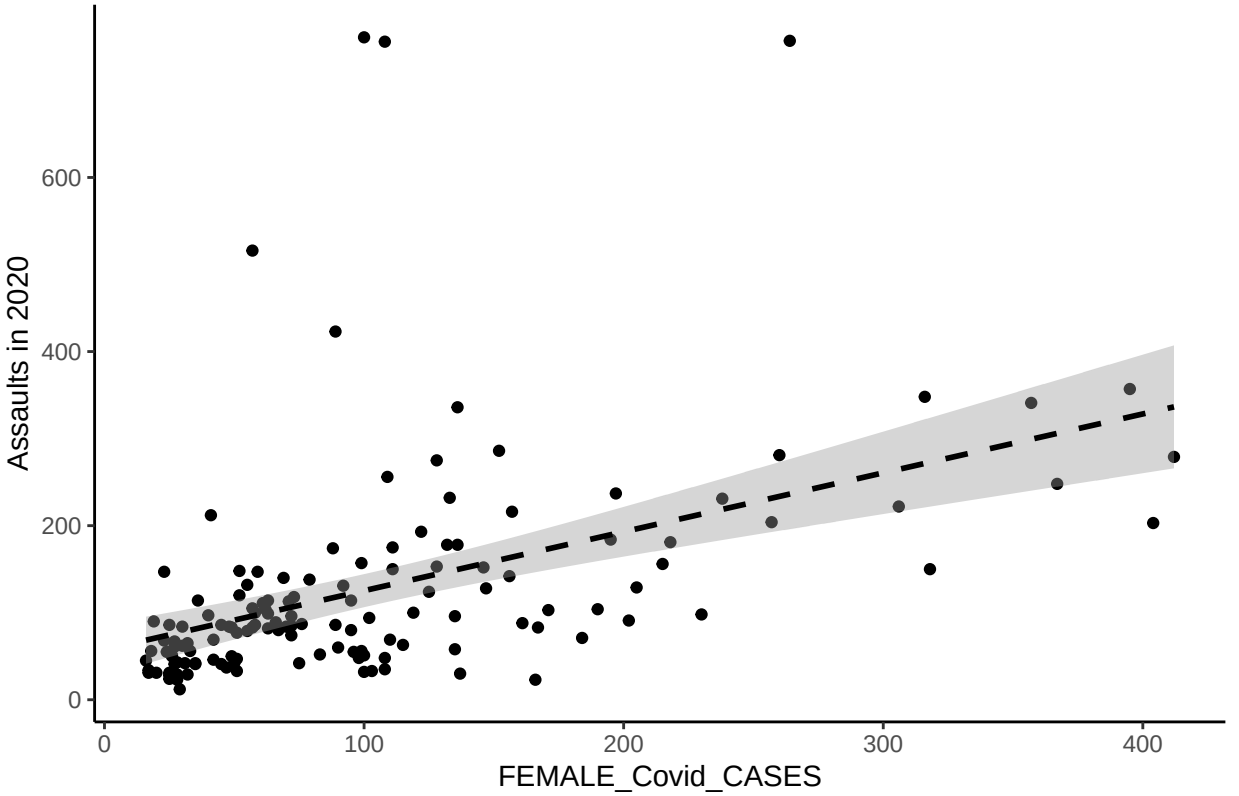


Figure 5: Relation between number of assaults in 2020 and number of male

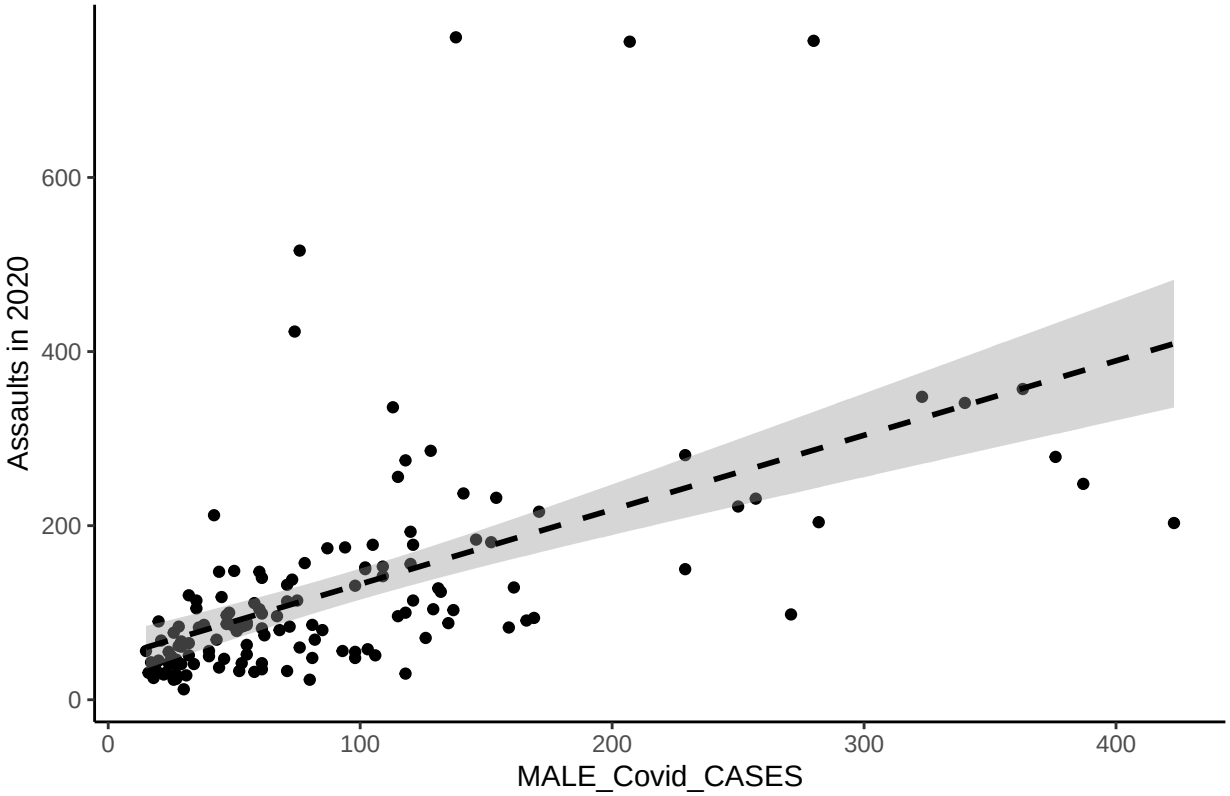


Figure 6: Relation between number of assaults in 2020 and number of covi

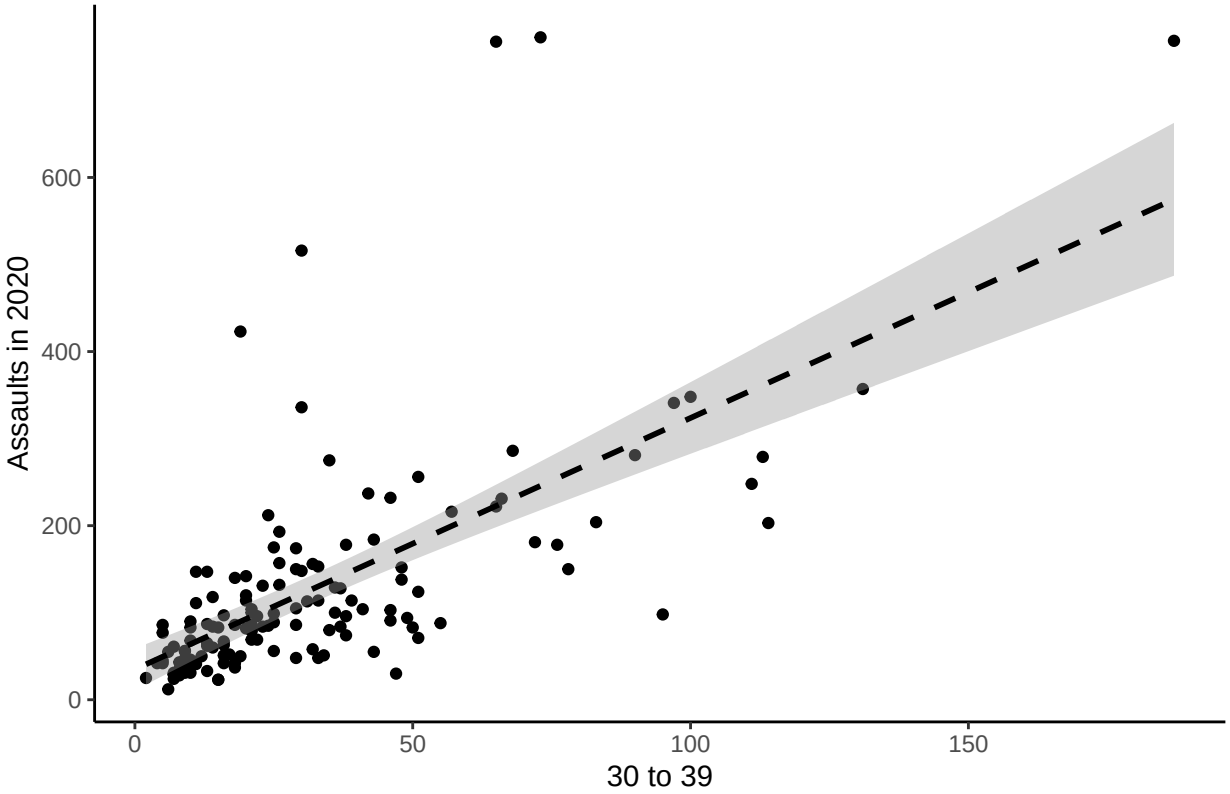


Figure 7: Relation between number of assaults in 2020 and number of covi

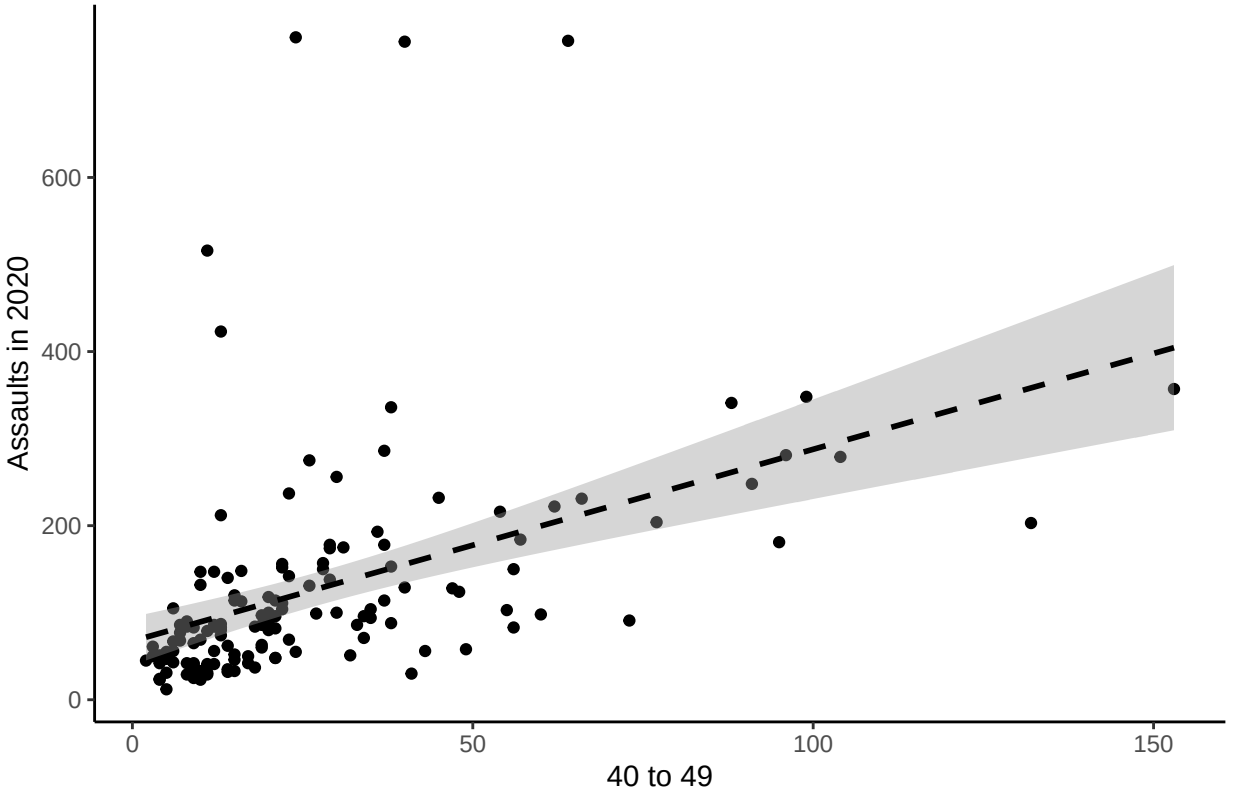


Figure 8: Relation between number of assaults in 2020 and the final model

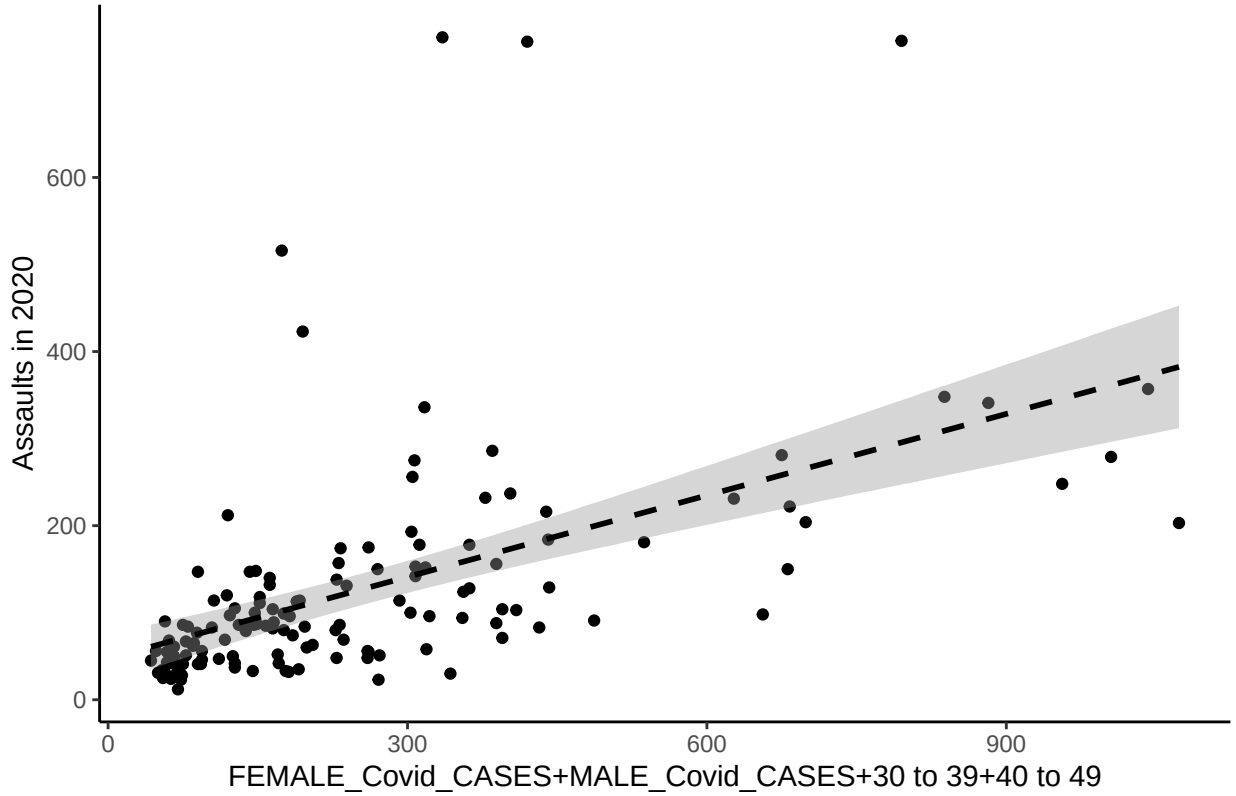
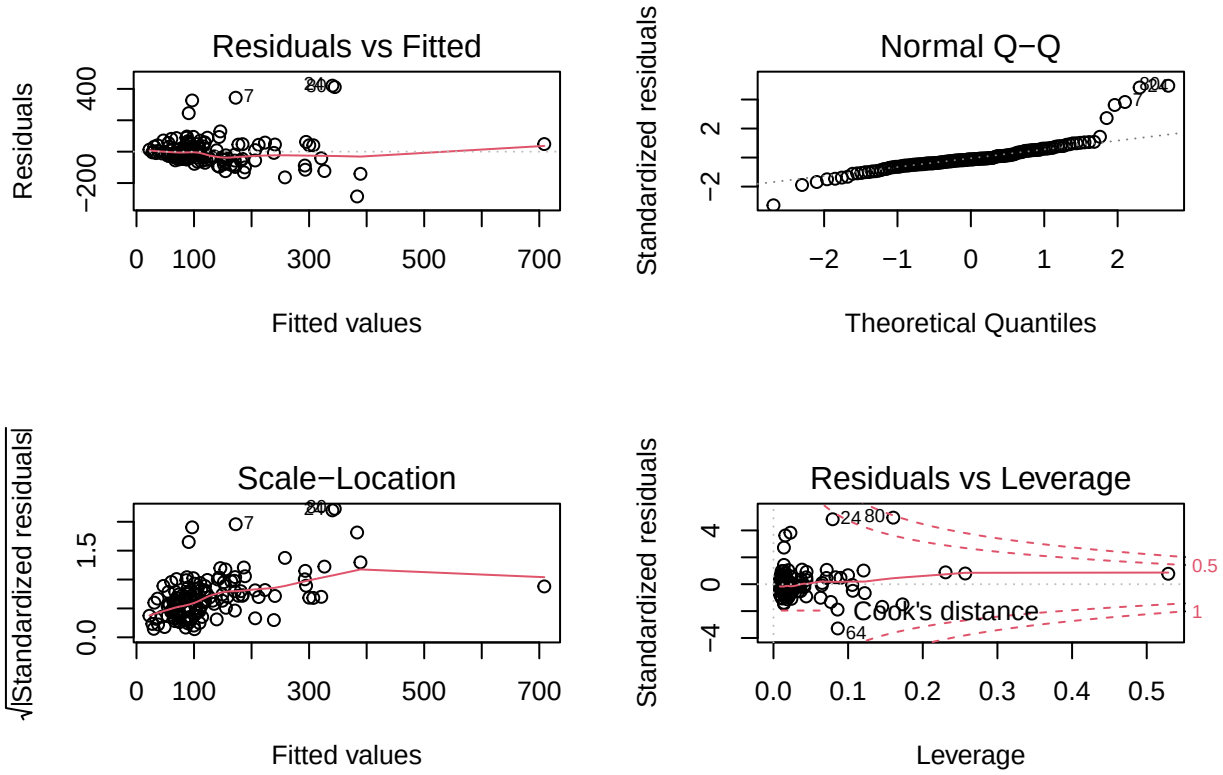


Figure 4,5,6,7 and 8 all show that there seems to be a positive relation between the number of covid cases in a neighbourhood and the number of assaults in that neighbourhood. The strongest positive relation is shown to be between the number of assaults in a neighbourhood and the age group 30 to 39 since figure 6 has the steepest slope.



We can analyse the goodness of fit for the model from the above 4 plots **??**fig:figure 5). From the residuals vs fitted plot, we see that linearity for the model does hold and that the independent variables do enter in a linear way. However, from the spread of the residuals we see that variance of our residuals is not constant which means that the model has heteroscedasticity. From the Normal Q-Q plot, we see that our data is normally distributed since the points mostly lie on the dotted line except for a few outliers.

4.4 Conclusion of results

The results obtained from our model and analysis of the dataset seem to imply that the number of covid cases in a neighbourhood had a positive relationship with the number of assaults in that same neighbourhood in the year 2020. The final model we obtained is:

$$y = 47.418 - 0.858 * x_1 + 1.059 * x_2 + 13.834 * x_3 - 1.964 * x_4$$

where y= number of assaults in a neighbourhood x_1 = number of female covid cases in a neighbourhood x_2 = number of male covid cases in a neighbourhood x_3 = number of covid cases between ages 30 and 39 x_4 = number of covid cases between ages 40 and 49

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

	Model 1
(Intercept)	127.457
	(10.773)
Num.Obs.	140
R2	0.000
R2 Adj.	0.000
AIC	1757.7
BIC	1763.6
Log.Lik.	−876.854
F	

	Model 1
(Intercept)	57.625
	(14.982)
FEMALE_Covid_CASES	0.677
	(0.111)
Num.Obs.	140
R2	0.211
R2 Adj.	0.205
AIC	1726.5
BIC	1735.4
Log.Lik.	−860.274
F	36.883

Appendix

A Appendix A

A.1 AIC selection

For this paper, AIC selection using modelsummary was used to obtain the final model. Several models were constructed and then using modelsummary, the model with the lowest AIC value was selected.

```
## the first model constructed using lm with no variables
first_model <-
  lm(Assault_2020 ~1,
     data = final_dataset)
modelsummary(first_model)
```

The AIC obtained from this model is 1757.7, so we will add a variable to see if we can reduce it.

```
second_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES,data = final_dataset)
modelsummary(second_model)
```

The AIC obtained from this model is 1726.5 which is less so we can keep the female variable.

	Model 1
(Intercept)	58.410
	(13.510)
FEMALE_Covid_CASES	−1.174
	(0.339)
MALE_Covid_CASES	2.029
	(0.355)
Num.Obs.	140
R2	0.363
R2 Adj.	0.354
AIC	1698.6
BIC	1710.3
Log.Lik.	−845.280
F	39.044

	Model 1
(Intercept)	58.992
	(13.795)
FEMALE_Covid_CASES	−1.181
	(0.341)
MALE_Covid_CASES	2.035
	(0.357)
transgender_covid_cases	−10.752
	(47.089)
Num.Obs.	140
R2	0.363
R2 Adj.	0.349
AIC	1700.5
BIC	1715.2
Log.Lik.	−845.253
F	25.867

```
third_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES,data = final_dataset)
modelsummary(third_model)
```

AIC obtained is 1698.6 so we can keep going.

```
fourth_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+transgender_covid_cases,data = final_dataset)
modelsummary(fourth_model)
```

The AIC obtained is 1700.5 so we don't include transgender covid cases

```
fifth_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`20 to 29`,data = final_dataset)
modelsummary(fifth_model)
```

We keep the 20 to 29 variable since the AIC is lower.

	Model 1
(Intercept)	59.460
	(12.957)
FEMALE_Covid_CASES	−0.892
	(0.334)
MALE_Covid_CASES	0.870
	(0.468)
‘20 to 29‘	2.033
	(0.563)
Num.Obs.	140
R2	0.419
R2 Adj.	0.406
AIC	1687.7
BIC	1702.5
Log.Lik.	−838.874
F	32.660

```
sixth_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`20 to 29`+`30 to 39`,data = final_dataset)
modelsummary(sixth_model)
```

So we move on with this model.

```
seventh_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`20 to 29`+`30 to 39`+`40 to 49`,data = final_da
modelsummary(seventh_model)
```

We proceed on with this model and will try to remove the 20 to 29 to see if we get a lower AIC.

```
eight_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`30 to 39`+`40 to 49`,data = final_dataset)
modelsummary(eight_model)
```

The AIC is even lower so we remove 20 to 29. We try with another variable

```
seventh_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`40 to 49`,data = final_dataset)
modelsummary(seventh_model)
```

This time the AIC is higher.

```
seventh_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`30 to 39`,data = final_dataset)
modelsummary(seventh_model)
```

The AIC is again higher than the AIC value for the eight model. So the eight model is our final model for this paper.

Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`30 to 39`+`40 to 49`

These are the estimates obtained for the final model using modelsummary.

	Model 1
(Intercept)	49.826
	(12.435)
FEMALE_Covid_CASES	−1.009
	(0.316)
MALE_Covid_CASES	0.638
	(0.445)
‘20 to 29‘	0.205
	(0.685)
‘30 to 39‘	3.561
	(0.844)
Num.Obs.	140
R2	0.486
R2 Adj.	0.471
AIC	1672.4
BIC	1690.1
Log.Lik.	−830.202
F	31.973

	Model 1
(Intercept)	47.521
	(12.258)
FEMALE_Covid_CASES	−0.854
	(0.318)
MALE_Covid_CASES	1.046
	(0.469)
‘20 to 29‘	0.040
	(0.677)
‘30 to 39‘	3.802
	(0.835)
‘40 to 49‘	−1.960
	(0.814)
Num.Obs.	140
R2	0.508
R2 Adj.	0.489
AIC	1668.5
BIC	1689.1
Log.Lik.	−827.241
F	27.643

	Model 1
(Intercept)	47.418
	(12.092)
FEMALE_Covid_CASES	−0.858
	(0.309)
MALE_Covid_CASES	1.059
	(0.416)
‘30 to 39‘	3.834
	(0.642)
‘40 to 49‘	−1.965
	(0.807)
Num.Obs.	140
R2	0.508
R2 Adj.	0.493
AIC	1666.5
BIC	1684.1
Log.Lik.	−827.243
F	34.810

	Model 1
(Intercept)	57.065
	(13.424)
FEMALE_Covid_CASES	−1.035
	(0.345)
MALE_Covid_CASES	2.360
	(0.397)
‘40 to 49‘	−1.615
	(0.902)
Num.Obs.	140
R2	0.378
R2 Adj.	0.364
AIC	1697.3
BIC	1712.0
Log.Lik.	−843.649
F	27.517

	Model 1
(Intercept)	49.331
	(12.283)
FEMALE_Covid_CASES	−1.031
	(0.306)
MALE_Covid_CASES	0.698
	(0.396)
‘30 to 39‘	3.720
	(0.652)
Num.Obs.	140
R2	0.486
R2 Adj.	0.475
AIC	1670.5
BIC	1685.2
Log.Lik.	−830.248
F	42.887

```
final_model<-
  lm(Assault_2020 ~FEMALE_Covid_CASES+MALE_Covid_CASES+`30 to 39`+`40 to 49`,
     data = final_dataset)
modelsummary(final_model)
```

B Appendix B

B.1 Covid Cases

This is a preview of the covid_cases dataset obtained from opendatatoronto (Gelfand 2020)

```
## # A tibble: 32,000 x 19
##     ‘_id‘ Assigned_ID ‘Outbreak Associated‘ ‘Age Group‘   ‘Neighbourhood~‘ FSA
##     <dbl>       <dbl> <chr>                 <chr>         <chr>            <chr>
##  1 310277      318475 Sporadic              40 to 49 Yea~ Humewood-Cedarv~ M6C
##  2 310278      318476 Sporadic              30 to 39 Yea~ Bay Street Corr~ M5S
##  3 310279      318477 Sporadic              20 to 29 Yea~ Waterfront Comm~ M5J
##  4 310280      318478 Sporadic              19 and young~ Roncesvalles     M6R
##  5 310281      318479 Sporadic              19 and young~ <NA>             M5H
##  6 310282      318480 Sporadic              19 and young~ Don Valley Vill~ M2J
##  7 310283      318481 Sporadic              19 and young~ Eglinton East    M1J
##  8 310284      318482 Sporadic              30 to 39 Yea~ Morningside      M1E
##  9 310285      318483 Sporadic              60 to 69 Yea~ Roncesvalles     M6K
## 10 310286      318484 Sporadic              19 and young~ Bedford Park-No~ M5M
## # ... with 31,990 more rows, and 13 more variables:
## #   ‘Source of Infection‘ <chr>, Classification <chr>, ‘Episode Date‘ <date>,
## #   ‘Reported Date‘ <date>, ‘Client Gender‘ <chr>, Outcome <chr>,
## #   ‘Currently Hospitalized‘ <chr>, ‘Currently in ICU‘ <chr>,
## #   ‘Currently Intubated‘ <chr>, ‘Ever Hospitalized‘ <chr>,
## #   ‘Ever in ICU‘ <chr>, ‘Ever Intubated‘ <chr>, year <chr>
```

	Model 1
(Intercept)	47.418
	(12.092)
FEMALE_Covid_CASES	−0.858
	(0.309)
MALE_Covid_CASES	1.059
	(0.416)
‘30 to 39‘	3.834
	(0.642)
‘40 to 49‘	−1.965
	(0.807)
Num.Obs.	140
R2	0.508
R2 Adj.	0.493
AIC	1666.5
BIC	1684.1
Log.Lik.	−827.243
F	34.810

B.2 Neighbourhood crime rates

This is a preview of the neighbourhood crimes dataset obtained from opendatatoronto (Gelfand 2020)

```
## # A tibble: 140 x 104
##    `_id` OBJECTID Neighbourhood            Hood_ID F2020_Populatio~ Assault_2014
##    <dbl>    <dbl> <chr>                    <chr>              <dbl>        <dbl>
##  1     1        1 Yonge-St.Clair           097                14083           16
##  2     2        2 York University Heights  027                30277          273
##  3     3        3 Lansing-Westgate         038                18146           42
##  4     4        4 Yorkdale-Glen Park       031                17560          106
##  5     5        5 Stonegate-Queensway      016                27410           91
##  6     6        6 Tam O'Shanter-Sullivan   118                29970          103
##  7     7        7 The Beaches              063                23364           88
##  8     8        8 Thistletown-Beaumond He~ 003                10948           61
##  9     9        9 Thorncliffe Park         055                23518           86
## 10    10       10 Danforth East York       059                18427           68
## # ... with 130 more rows, and 98 more variables: Assault_2015 <dbl>,
## #   Assault_2016 <dbl>, Assault_2017 <dbl>, Assault_2018 <dbl>,
## #   Assault_2019 <dbl>, Assault_2020 <dbl>, Assault_Rate2014 <dbl>,
## #   Assault_Rate2015 <dbl>, Assault_Rate2016 <dbl>, Assault_Rate2017 <dbl>,
## #   Assault_Rate2018 <dbl>, Assault_Rate2019 <dbl>, Assault_Rate2020 <dbl>,
## #   AutoTheft_2014 <dbl>, AutoTheft_2015 <dbl>, AutoTheft_2016 <dbl>,
## #   AutoTheft_2017 <dbl>, AutoTheft_2018 <dbl>, AutoTheft_2019 <dbl>, ...
```

B.3 Appendix C

C References

Arel-Bundock, Vincent, Joachim Gassen, Nathan Eastwood, Nick Huntington-Klein, Moritz Schwarz, Benjamin Elbers, and Grant McDermott. 2022. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready.* https://cran.r-project.org/package=modelsummary.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://cran.r-project.org/package=janitor.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto, Open Data. 2022. "About COVID-19 Cases in Toronto." City of Toronto Open Data Portal. https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2022. *Tidyr: Tidy Messy Data.* https://tidyr.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019b. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———, et al. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, Romain Francois, Jennifer Bryan, Shelby Bearrows, Rstudio, Jukka Jylänki, and Mikkel Jørgensen. 2022. *Readr: Read Rectangular Text Data.* https://cran.r-project.org/package=readr.

Xie, Yihui. 2021a. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.

———. 2021b. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents.* https://github.com/yihui/tinytex.

(Wickham et al. 2019b)

(Wickham 2016)

(Wickham et al. 2021)

(Wickham 2022)

(Firke 2021)

(Wickham 2007)

(Müller 2020)

(Xie 2021b)

(Wickham et al. 2022)

(Toronto 2022)