Code

A study of the socio-economic and geographic aspects of detainment since the Myanmar Coup in 2021

Project group D2: Abhilash Biswas (abhilasb), Jaison Jose (jjose), Vidisha Chowdhury (vchowdhu)

09/12/2021

Abstract

On 1 February 2021, the quasi-democratic government of Myanmar was overthrown by the Tatmadaw - the Burmese military in a coup d'état which sparked a series of protests in the country. Owing to the brutal response of the military to the peaceful demonstrations, people have started resorting to more violent forms of resistance such as paramilitary training to oppose the Tatmadaw. In response, the military has been attempting to aggressively eliminate resistance in the country. According to the Assistance Association for Political Prisoners (AAPP), nearly 1,300 people have been killed by the junta and more than 10,000 have been arrested since the coup. This analysis is an attempt to characterise detainment in Myanmar. We study individual level attributes such as gender and age distribution of detainees, look at the association between detainment, imprisonment and deaths. In addition, our township level analysis tries to examine if detainment is associated with some socio-economic, demographic or geographic patterns in Myanmar.

Introduction

Since the military coup in Myanmar, many civilians and opposition forces have been detained, imprisoned, and killed. The Assistance Association for Political Prisoners (AAPP) records these occurrences along with the names, gender, designation and additional attributes of people involved in these incidents. Since February 2021, as the Burmese military adopted more violent ways to curb public resistance, there is a growing consensus among people that the military needs to be countered using more aggressive forms of opposition. In certain parts of the country, civilians are training with firearms and hand grenades making it evident that the country is on the brink of a civil war.

Media reports suggest that the military has been targeting areas that are home to armed civilians known as the People's Defense Force. It has been reported that there has been a large buildup of troops in the northwest of Myanmar. According to residents, fire rockets have been launched, homes burnt down, food supplies cut and civilians attempting to escape have been shot.

Using data recorded by the AAPP and township level Census data, we try to analyze if detainment in Myanmar is correlated with socio-economic and geographic patterns. We perform multiple tabulations to assess whether detainment varies across individual characteristics such as gender and the position/designation. Additionally, we try to visualize if there is any pattern in detainment across socio economic factors like literacy rate, proportion of urban population etc. Finally, we conduct 2 regression models where we assess a bunch of factors and it's association with detainment. Apart from some of the tabulations at an individual level, all other tests and analysis is at a township level. This exercise will help us investigate if there are particular patterns in detainment and analyze reports of targeted military operations.

Methods

The three data sets we use and their sources are outlined below.

1. Data on detainment, imprisonment and death of political prisoners - AAPP
2. Historical data on Myanmar indicators at the township level - Myanmar Information Management Unit (MIMU)
3. A derived data set on conflicts and protests in Myanmar in 2021 - Armed Conflict Location & Event Data Project (ACLED)

1. Cleaning and merging

a. Cleaning all the individual datasets

Code

In this section we clean all the individual datasets, namely:

- Detainees
- Imprisoned

- Fallen
- Conflicts

For each individual datasets, the following broad cleaning steps were followed:

1. Drop rows containing no names for individuals
2. Clean the Sex column (to be used for analysis later)
3. Clean the Status column to extract the designation of individuals
4. Extract the month of the incident that took place
5. Convert each dataset to a nested tibble at a township level

Based on the cleaning performed, below are some summary numbers:

Detainees dataset

a. There are a total of 1967 unique townships in the detainees dataset. However, some of them may be random text and not actual township names.
b. Out of 6689 observations in the detainees dataset, 17 had missing Sex values
c. Out of 6689 observations in the detainees dataset, 4755 had missing age values

Imprisoned dataset

a. There are a total of 96 unique townships in the imprisoned dataset. However, some of them may be random text and not actual township names.
b. Out of 306 observations in the imprisoned dataset, 0 had missing Sex values
c. Out of 306 observations in the imprisoned dataset, 267 had missing age values

Fallen dataset

a. There are a total of 214 unique townships in the fallen dataset. However, some of them may be random text and not actual township names.
b. Out of 1205 observations in the fallen dataset, 36 had missing Sex values
c. Out of 1205 observations in the fallen dataset, 303 had missing age values

Conflicts dataset

a. There are a total of 314 unique townships in the conflicts dataset. However, some of them may be random text and not actual township names.
b. There are a total of 14188 reported in this dataset.
c. There are a total of 8 unique actors involved in these conflicts.

b. Getting the sector indicators and merging

Code

```
## Response [https://www.andrew.cmu.edu/user/jweiss2/21f_r/94842/final_2021/MIMU_BaselineData_AllSectors_Count
rywide_18Mar2021_revised.xlsm]
##   Date: 2021-12-10 04:40
##   Status: 200
##   Content-Type: application/vnd.ms-excel.sheet.macroEnabled.12
##   Size: 19.3 MB
## <ON DISK>  /var/folders/h_/mljy2_zs24d70d80xqt9v9_w0000gn/T//Rtmp1BdsJD/file30fe75d1cd0e.xlsm
```

Code

In this section, we have picked certain indicators from the MIMU dataset and merged it with township level detainee, fallen, imprisoned and conflict counts for our analysis.

Key datasets

merged_data_det = detainees merged with MIMU indicators

a. Number of unique townships = 170
b. We have a total of 1831 detainees in this dataset as compared to 6689 in the individual detainees dataset.

merged_data_det_conf = detainees merged with MIMU indicators merged with conflict data

a. Number of unique townships = 166
b. We have a total of 1814 detainees in this dataset as compared to 6689 in the individual detainees dataset.
c. We have a total of 8764 conflicts in this dataset as compared to 14188 in the individual detainees dataset.
d. This is the dataset we will use primarily for our analysis

This suggests that in the merging process, we have lost quite a bit of individual counts of detainees and fallen. For our subsequent analysis, we have assumed that if a town is not present in the above datasets, then the count of the appropriate variable (for example total detained) for that township is 0.

Socio-economic variables chosen

Below is a list of the socio economic indicators chosen and a brief rationale:

indicators	rationale
Age dependency ratio	To see if number of dependent population in a township affects detainment
Child dependency ratio	Same as above
Population size	An essential control while comparing detainee counts across townships
Proportion of female-headed households	Interesting indicator of social dynamics in a town
Sex ratio	Since male and female may have different detainment rates
Urban Population as Percentage of Total Population	To see if detainments happen more in urban or rural areas
Adult literacy rate	To see if more literate population are detained in higher numbers
Number of Households by source of lighting	Indicator of development status of the town
Percentage of households with safe sanitation	Indicator of development status of the town
Percentage of Households' use of drinking water by source	Indicator of development status of the town
Approximate Vulnerable Population	Similar rationale as for Age dependency ratio
Non-Dependent population	Same as above
Wealth Ranking Index	Indicator of wealth
Infant mortality rate	Important health indicator for a town
Township hospital	Same as above
Children Immunized (BCG)	Same as above
Malnutrition under three years	Same as above
Battles	To see if more battles lead to more detainment
Fatalities	To see if fatalities and detainment is correlated

2. Tabulations

We start our analysis by tabulating some individual and socio-economic characteristics and seeing how detainee counts differ across these characteristics groups.

a. Distribution of detainees, imprisoned and fallen by gender (Univariate)

Distribution of detainees by Sex

Code

Sex	Count
F	1245
M	5427

Distribution of imprisoned by Sex

Code

Sex	Count

Sex	Count
F	56
M	250

Distribution of fallen by Sex

Code

Sex	Count
F	84
M	1085

From these tables, we can see that male citizens are detained, imprisoned and fallen at a disproportionately higher count. However, there could be a lot of biases in play here. Firstly, the sex ratio may be skewed towards men in the overall population. Further, if males are disproportionately represented in government service (which was a source of a lot of protests against this coup), it would lead to higher counts of men for detainment as well as other categories.

b. Distribution of detainees by designation (Univariate)

Code

designation	Count
Academic professional	110
Activist	33
Administrative staff (any level of state)	55
Civilian	5335
Government Minister	24
Media Person	43
Medical Professional	67
MP or Former MP	49
NLD	207
Other Misc	370
Public Servant	21
Religious person	44
Student	331

We see that civilians are the most detained category. However, this could be a categorisation problem where a bunch of categories are classified as civilians. These categories are not necessary mutually exclusive. Apart from civilians, we see that members of the National League for Democracy were the most detained category. This makes sense as the coup removed the government formed by this party and hence a large number of members from this party would be detained. It is also interesting to note that academic professionals are a significant detainee category. This could be because of government employees in state run academic institutions came out in a civil disobedience against the coup.

c. Top 5 and bottom 5 detainee townships and the number of imprisoned in each (Multivariate)

Code

MIMU_township	imprisoned	detainees
Monywa	1	97
Myeik	3	93

MIMU_township	imprisoned	detainees
North Okkalapa	15	48
Bago	2	44
Dawei	18	43
Thayetchaung	6	8
Tachileik	4	6
Kawthoung	2	5
Kyangin	2	5
Tanintharyi	2	1

We can see that the top 5 towns in terms of detainees also have a higher count of imprisoned people. Likewise the bottom 5 towns have a lower count of imprisoned people. This may suggest that detainment could be followed by imprisonment as a natural next step. It could also indicate that the difference between what is called a detainment and what is called imprisonment may not be very well defined.

d. Top 5 and bottom 5 detainee townships and their socio economic indicators (Multivariate)

Code

MIMU_township	detainees	electricity_access	IMR	under_3_malnutrition	improved_drinking_water_access	Township hospital
Monywa	97	51104	19.5	12.6	93.30	0
Myeik	93	5721	6.8	1.9	86.58	0
Thingangyun	74	42137	14.8	1.2	99.36	0
Insein	55	55409	31.0	1.7	98.97	1
Myitkyina	54	26777	7.2	3.0	90.96	0
Shwedaung	1	9512	16.5	1.4	88.36	1
Shwegu	1	3571	34.6	1.2	87.79	1
Tanintharyi	1	1214	1.7	2.6	44.59	1
Tatkon	1	9394	9.2	0.2	85.12	1
Wundwin	1	12862	14.6	1.6	92.11	1

This table provides some interesting observations. Higher detainment towns have higher electricity access. This may indicate that more developed towns also had higher detainment. We will see later how urban population in a town affects detainment. It is important to note that both detainment and counts of electricity access can just be positively correlated with population size.

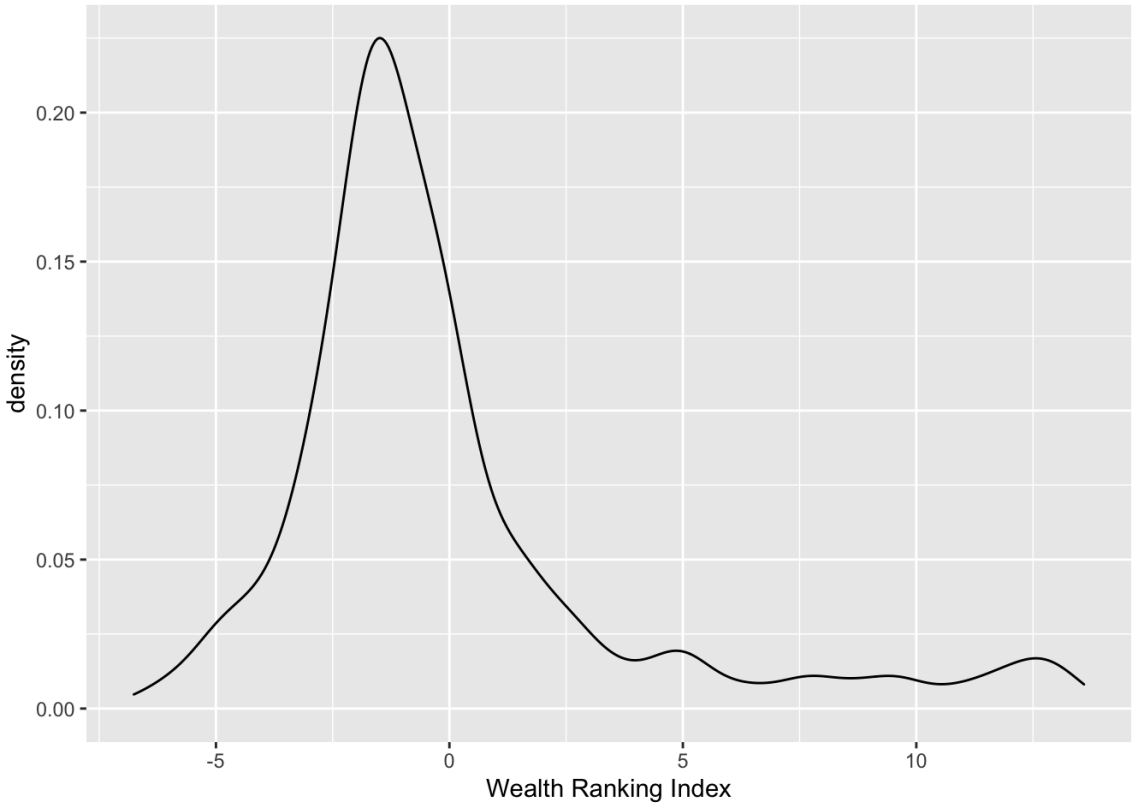
We do not see much of a difference in Infant mortality rate, malnutrition rate and access to improved drinking water rate to be associated with detainment counts. However, townships that had higher detainment also seem to not have township hospitals existing in them.

3. Data descriptions and visualizations

To take our analysis further and assess important variables for statistical tests, in this section we show visualizations of interesting inferences that we were able to clean from the data set.

a. Wealth Ranking index distribution (Univariate)

Code



(Due the absence of an exact definition of wealth index in the MIMU codebook, we assume that the wealth index is a composite measure of a household's cumulative living standard calculated using data on a household's ownership of selected assets. https://dhsprogram.com/topics/wealth-index/ (https://dhsprogram.com/topics/wealth-index/))

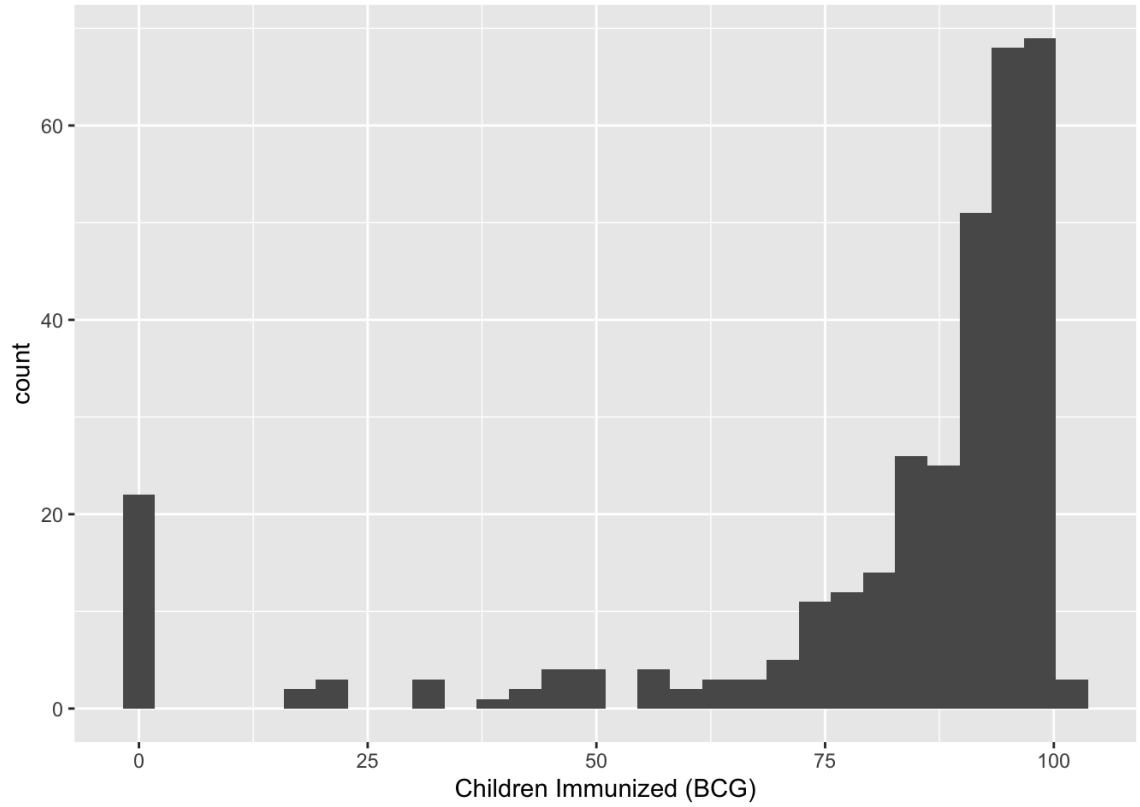
The sample distribution of the wealth index across townships shows that the variable is likely to have an asymmetric distribution since its distribution is positively skewed.

The density graph indicates that a high fraction of townships have a lower wealth ranking ranking (below 5)

b. Distribution of Children immunised with BCG (Univariate)

Code

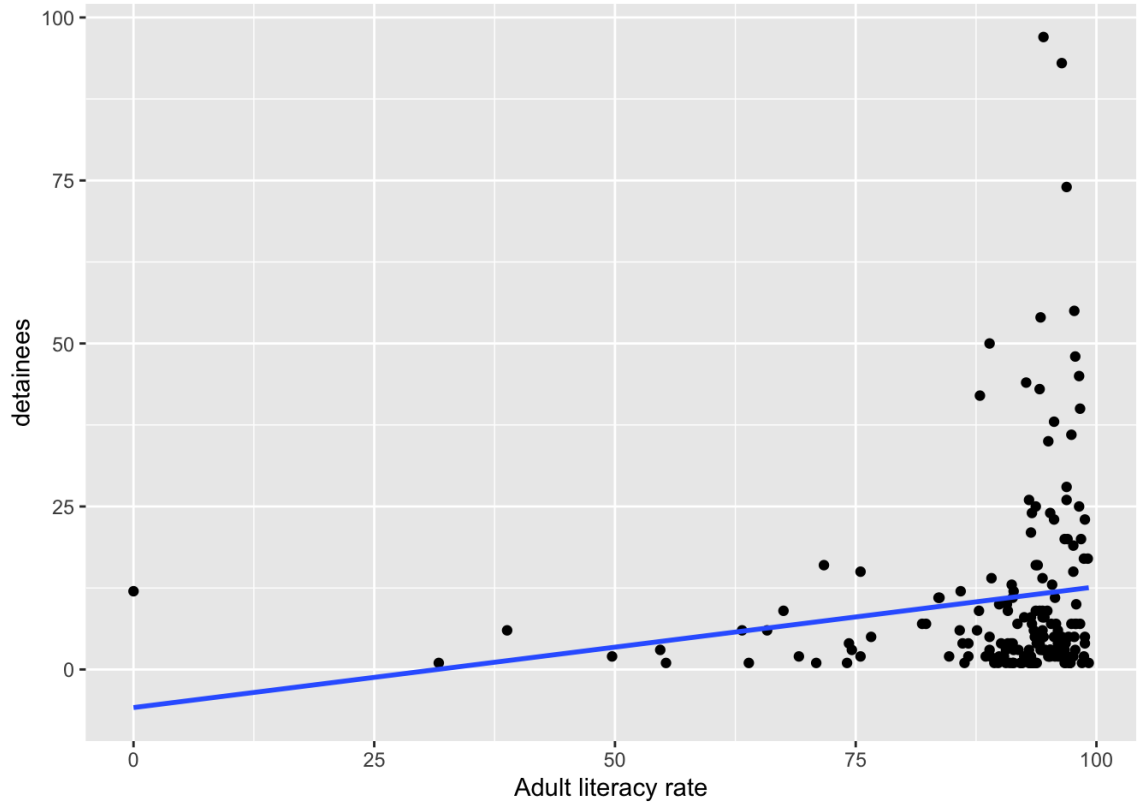
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The distribution of child immunization rates suggests that while a significant chunk of the townships have achieved above 80% or complete BCG immunization, there are about 21 townships with immunization rates close to zero.

c. Relationship between township detainee count and literacy rate (Bivariate)

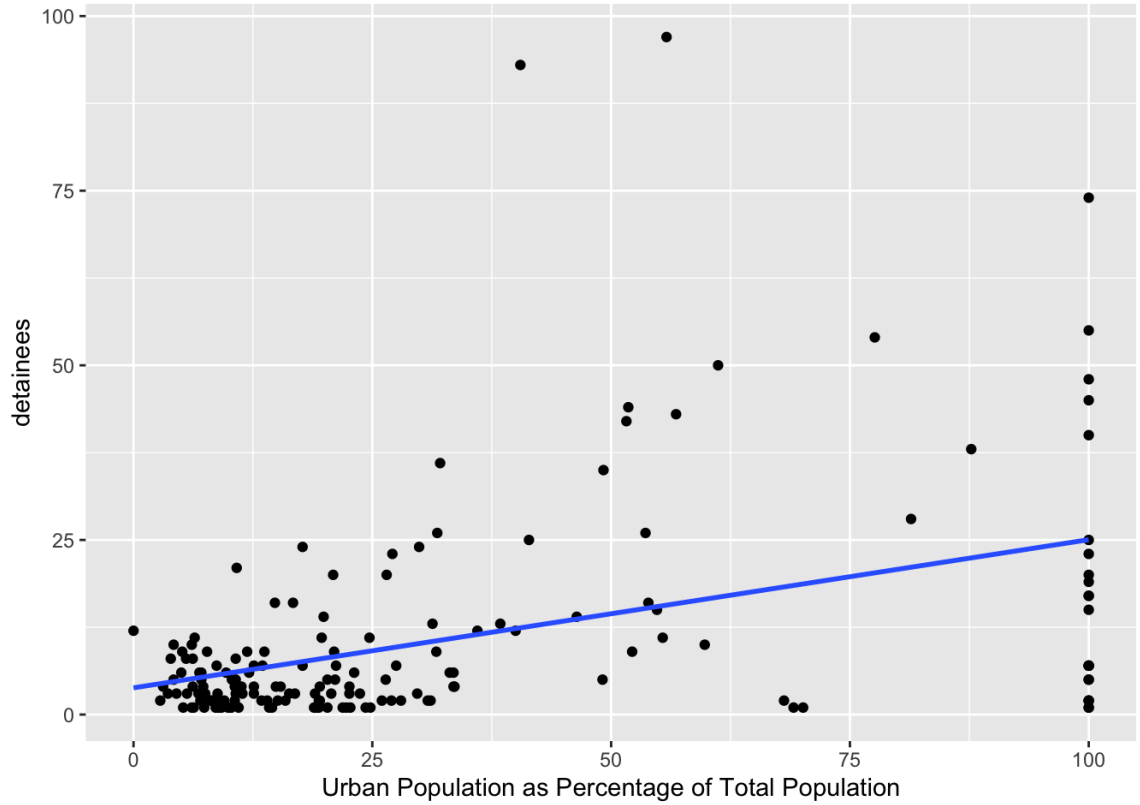
Code



We can see that townships with higher literacy rate had more number of detainees. There could be several possibilities. Firstly, more data might be captured on detainees in towns that have a higher educated population. Secondly, we may have higher resistance against the coup from more literate areas, although we can only say that after some statistical tests.

d. Relationship between township detainee count and proportion of urban population (Bivariate)

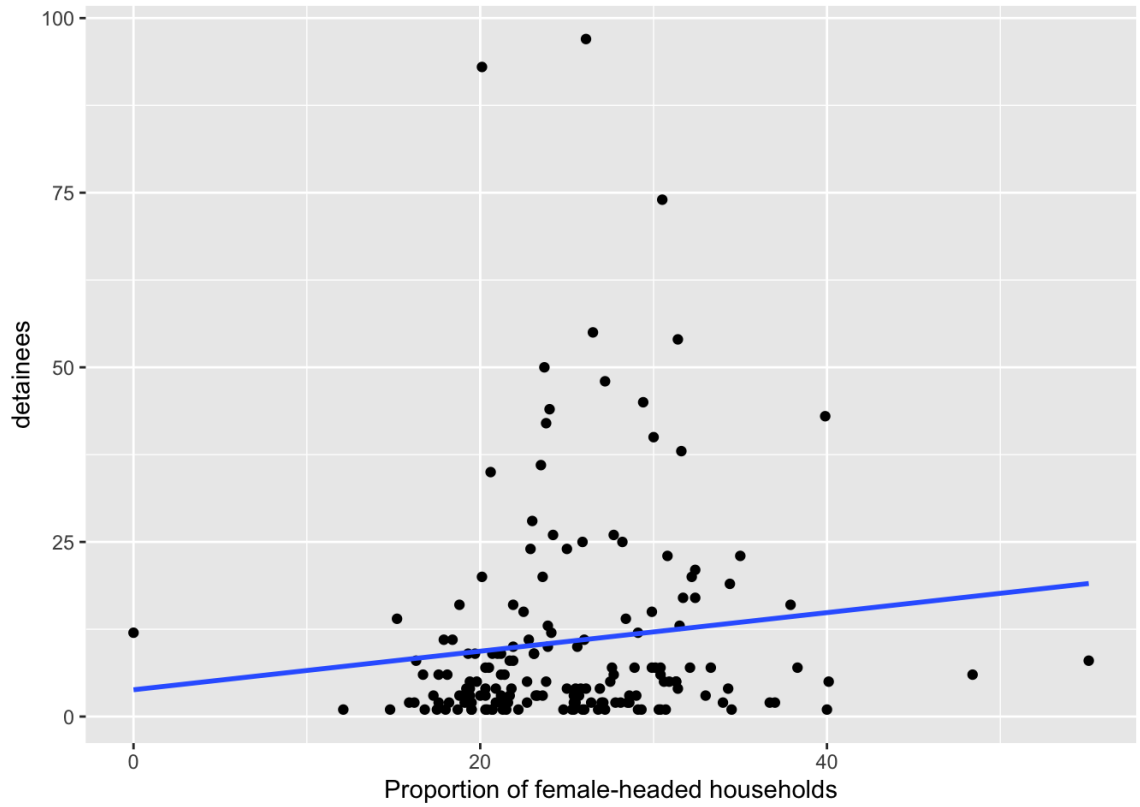
Code



This graph suggests a positive association between count of detainees and proportion of urban population, at a township level. We later investigate if this relation is statistically significant.

e. Relationship between township detainee count and proportion of female headed households (Bivariate)

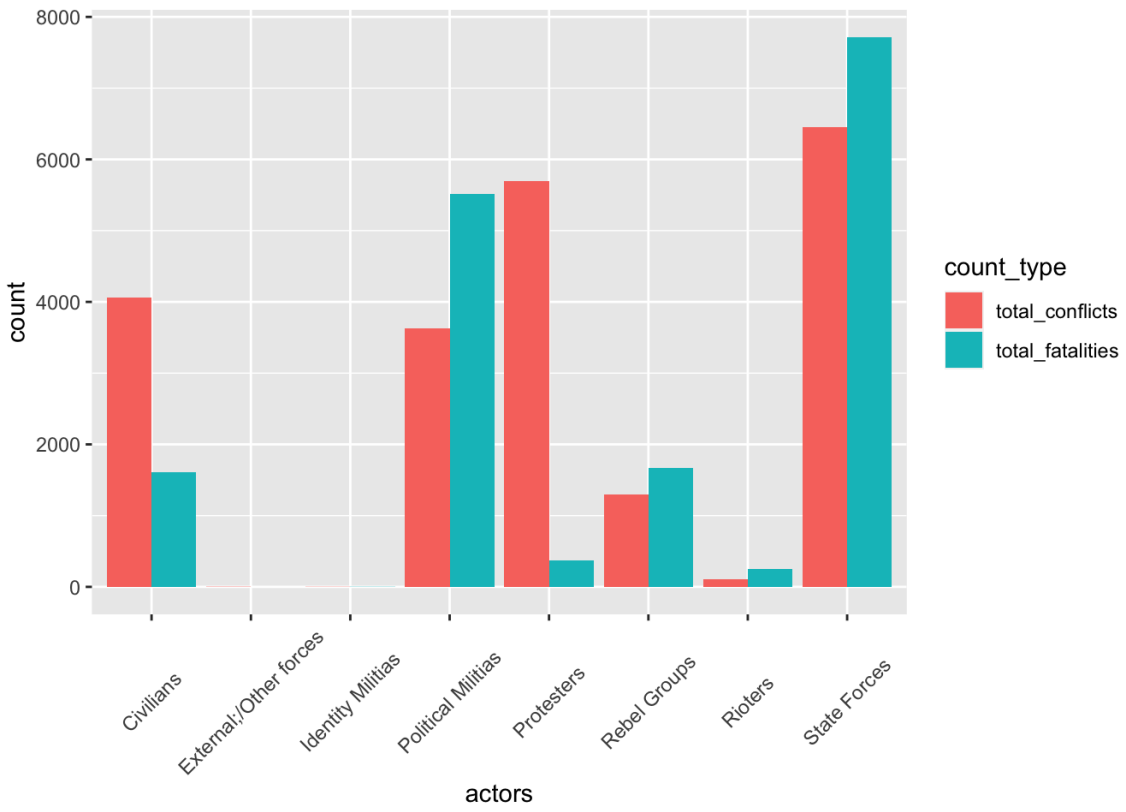
Code



While no clear trend seems to emerge, one can see a lot of variation in detainee counts in townships which had 20-40% of female headed households.

f. Number of conflicts and fatalities by actor (Trivariate)

Code

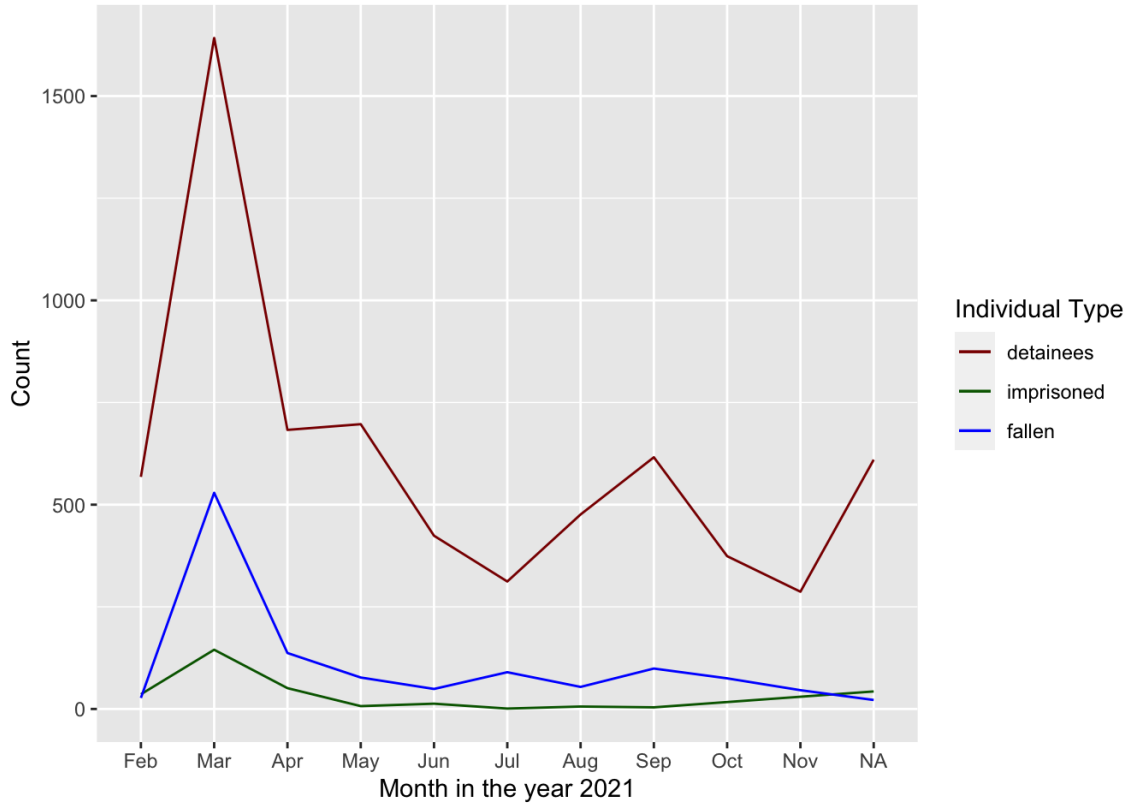


We can see that State forces (Burmese Military) and Political Militias are involved in the most conflicts (not necessarily with each other, but possibly) and account for the highest number of fatalities. It is useful to note that if there is a conflict between 2 parties, each of the parties get 1 added to their conflict count.

For protesters, the reason why they have such high number of conflicts but no fatalities is because most of their activities are peaceful protests (with no involved 2nd actor) and the source dataset categorises peaceful protests as a type of conflict.

g. Detainees, fallen and imprisoned over time (Trivariate)

Code



We can see that there is a spike in detainee, imprisoned and fallen counts during march, followed by frequent bursts of detainment. This is aligned with the general reports of increased detainment following the coup. Imprisonment and fatalities have generally reduced over time, except for two spikes in July and September.

4. Statistical tests

In this section we shall drawing inferences based on the data exploration that we have done. We are forming hypothesis on variables by what we have analyzed above and testing them for statistical significance

a. Checking if proportion of urban population is different in towns which haven't had any detainment as compared to those which have had atleast 1 detainment

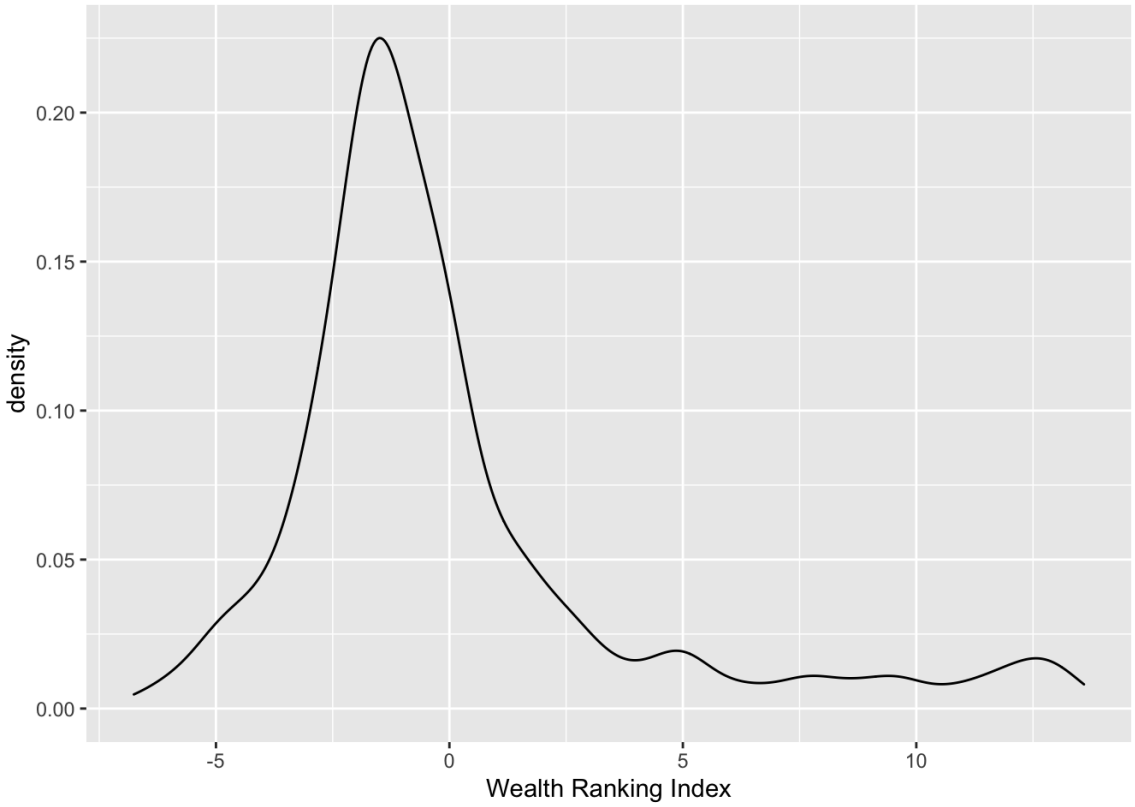
Code

```
##
##  Welch Two Sample t-test
##
## data:  Urban Population as Percentage of Total Population.x by any_detainment
## t = -3.4267, df = 321.77, p-value = 0.0006902
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -16.770450  -4.537264
## sample estimates:
## mean in group 0 mean in group 1
##        22.08144        32.73529
```

Townships having >0 detainment have higher proportion of urban population (on average). This difference is statistically significant at 1% level of alpha.

b. Checking how the wealth ranking index varies between townships having no detainees vs those having atleast 1 detainee

Code



Since the wealth ranking index is not quite normally distributed, we will perform a Wilcoxon rank sum test on it (across township detainment binary variable)

Code

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Wealth Ranking Index.x by any_detainment
## W = 10385, p-value = 2.043e-05
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -1.671132 -0.575284
## sample estimates:
## difference in location
##              -1.094134
```

Based on this test, we can see that there is a statistically significant difference on wealth ranking index between detainee and not detainee townships.

5. Results (regressions)

Our descriptive analysis raises a number of questions.

- Is there a significant relationship between the detainee count of a township and its state of development as measured by health, electricity access and literacy indicators?
- Are more urbanized townships associated with higher detainee counts?
- Since there is some indication of difference in detainee counts by age and gender at the individual level, are detainee counts significantly associated with the age dependency ratio and sex ratio at the township level?

In our next step, we try to answer the above questions by estimating a linear regression. A linear regression tries to check for a linear relation between an independent variable and a dependent one keeping other factors unchanged.

a. Regressing detainees on all the chosen indicators from MIMU (including population size) at a township level

Code

```
##
## Call:
## lm(formula = detainees ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.091  -5.396  -1.117   4.120  63.928
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -3.041e+01  2.483e+01  -1.225  0.22258
## conflicts                      9.061e-02  2.233e-02   4.057 8.11e-05 ***
## `Age dependency ratio`         8.640e-01  5.823e-01   1.484  0.14002
## `Child dependency ratio`      -3.495e-01  5.386e-01  -0.649  0.51743
## `Population size`              6.469e-02  4.983e-02   1.298  0.19628
## prop_female_headed_hh         -1.803e-01  1.853e-01  -0.973  0.33224
## sex_ratio                     -1.528e-01  1.647e-01  -0.928  0.35492
## prop_urban_population          2.744e-01  9.228e-02   2.974  0.00345 **
## `Adult literacy rate`         -3.850e-02  1.344e-01  -0.286  0.77498
## electricity_access            -3.118e-04  1.895e-04  -1.645  0.10217
## percent_safe_sanitation        2.536e-02  7.872e-02   0.322  0.74781
## improved_drinking_water_access -3.419e-02  4.799e-02  -0.712  0.47745
## approx_vulnerable_population  -1.940e-04  6.733e-05  -2.882  0.00456 **
## `Non-Dependent population`     2.649e-04  9.680e-05   2.737  0.00698 **
## `Wealth Ranking Index`        -5.577e-01  8.384e-01  -0.665  0.50698
## IMR                           -1.762e-01  1.266e-01  -1.391  0.16633
## `Township hospital`           -4.913e+00  2.375e+00  -2.068  0.04040 *
## `Children Immunized (BCG)`     1.602e-01  6.834e-02   2.344  0.02044 *
## under_3_malnutrition           1.759e-01  4.538e-01   0.388  0.69882
## Battles                       -2.618e-01  2.130e-01  -1.229  0.22102
## Fatalities                     5.584e-02  3.710e-02   1.505  0.13449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.9 on 144 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.5829, Adjusted R-squared:  0.5249
## F-statistic: 10.06 on 20 and 144 DF,  p-value: < 2.2e-16
```

We are controlling for population size as we expect that townships with higher populations may also have a higher detainment count.

We see that conflicts, proportion of urban population, vulnerable population and non-dependent population, existence of township hospitals and child immunization rate against BCG are all statistically significantly associated with detainee counts in a township. For instance, a higher proportion of urban population is associated with significantly higher detainee counts.

However, many indicators such as age dependency ratio, population size and sex ratio are statistically insignificant in explaining variation in detainee counts. While it's possible that detainee counts are not associated with these variables, it is also likely that many of these are linearly correlated with each other due to which their estimated coefficients are insignificant in the model. Hence, it is worth checking for correlation between our independent variables.

b. Assessing collinearity in the above regression

Code

```
##                            row                       column       cor p
## 1         Age dependency ratio       Child dependency ratio 0.9801850 0
## 2              Population size approx_vulnerable_population 0.8801481 0
## 3              Population size     Non-Dependent population 0.9143501 0
## 4           electricity_access     Non-Dependent population 0.7460249 0
## 5 approx_vulnerable_population     Non-Dependent population 0.7305859 0
## 6        prop_urban_population         Wealth Ranking Index 0.9324241 0
```

We see that some of the variables are highly correlated with each other. This would affect their interpretation in the regression as it increases their standard errors. Hence to make our regression model more robust, we will drop some ofthe highly correlated variables.

c. Running a 2nd model after dropping the highly correlated independent variables and comparing the 2 models

Code

	full.model	reduced.model
(Intercept)	−30.414	−29.317
conflicts	0.091***	0.102***
Age dependency ratio	0.864	0.279*
Child dependency ratio	−0.349	
Population size	0.065	0.051**
prop_female_headed_hh	−0.180	−0.339+
sex_ratio	−0.153	−0.099
prop_urban_population	0.274**	0.206***
Adult literacy rate	−0.039	0.172
electricity_access	0.000	0.000
percent_safe_sanitation	0.025	0.008
improved_drinking_water_access	−0.034	0.006
approx_vulnerable_population	0.000**	
Non-Dependent population	0.000**	
Wealth Ranking Index	−0.558	
IMR	−0.176	−0.097
Township hospital	−4.913*	−3.799
Children Immunized (BCG)	0.160*	0.083
under_3_malnutrition	0.176	0.137
Battles	−0.262	−0.306
Fatalities	0.056	0.049
Num.Obs.	165	165
R2	0.583	0.538
R2 Adj.	0.525	0.488
AIC	1278.1	1287.0
BIC	1346.4	1342.9

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

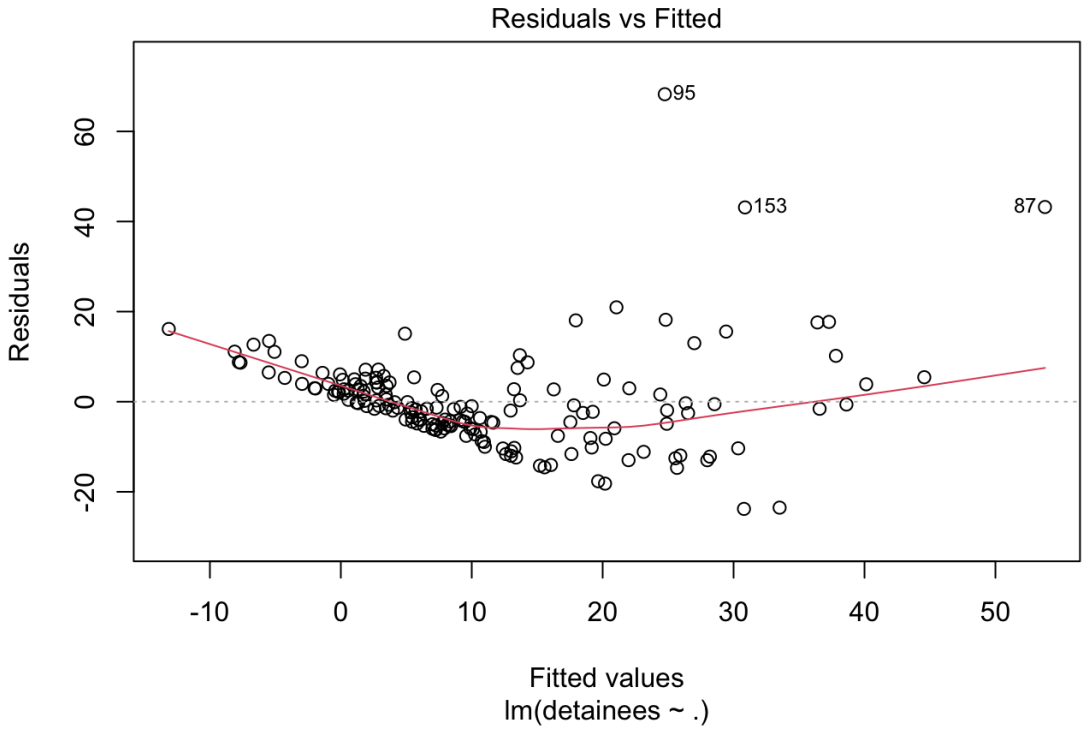
	full.model	reduced.model
Log.Lik.	−617.042	−625.494
F	10.061	10.766

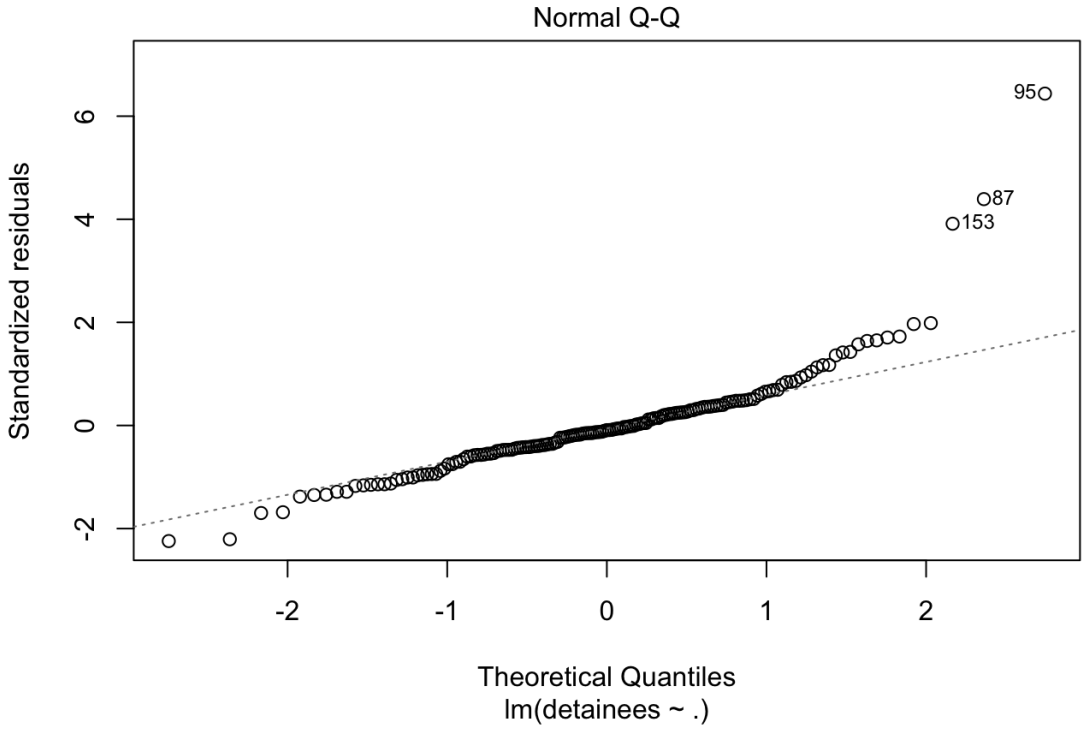
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

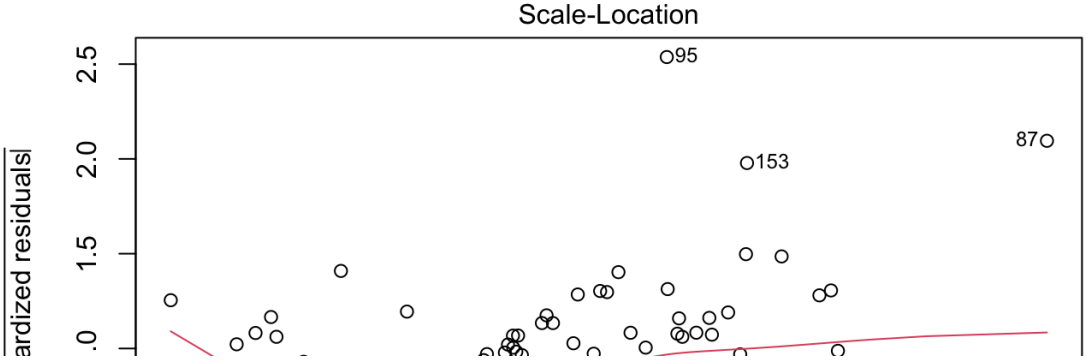
After dropping the correlated variables, we see that:

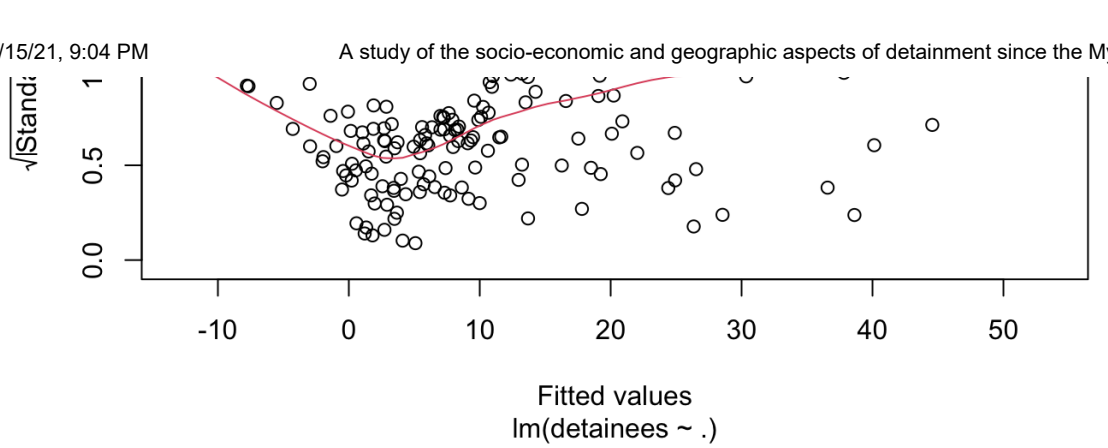
- Conflicts remain highly statistically significant
- Age dependency ratio and population size become significant, indicating that they were getting affected by collinearity
- Township hospitals and children immunized were significant earlier but now become insignificant after dropping the correlated variables.

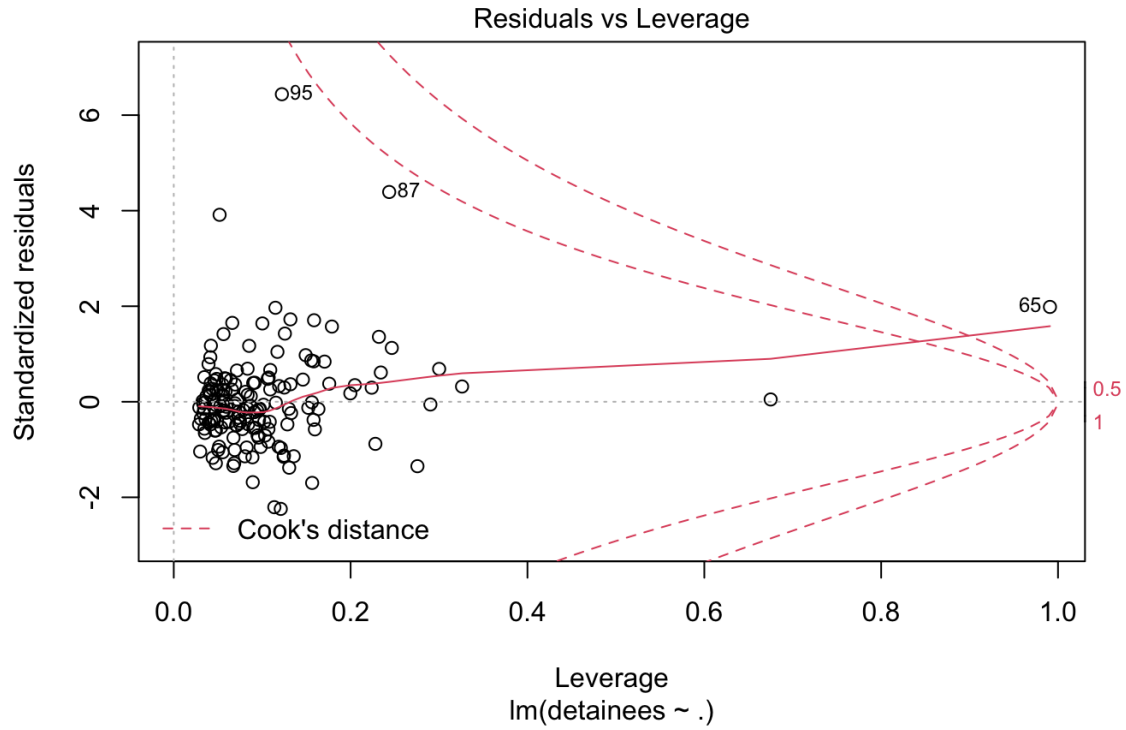
Code


Residuals vs Fitted
Residuals
Fitted values
lm(detainees ~ .)
95
153
87


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(detainees ~ .)
95
87
153


Scale-Location
95
87
153





6. Discussion

Having estimated a more robust regression model (model 2), we see the following results. - More urbanized townships are associated with higher detainee counts - Townships which have experienced more conflicts have higher detainee counts - Larger townships are associated with higher detainee

More urbanized townships could be associated with higher detainee counts due to multiple reasons. It's possible that there is more detainment reporting in urban centres. It is also possible that since urban centres often have a more diverse population consisting of academicians, people belonging to the civil society etc. it could represent a more liberal minded population and therefore be associated with more detainment.

It is expected that a higher number of conflicts will be associated with higher detainment and that larger populations will be associated with higher detainment.

However, there are a number of caveats which may affect our results. Given the data set we used we faced constraints such as a large number of missing values primarily due to errors in records of township names which made it difficult to merge various data sets. So we went ahead and assumed that townships with missing detainee counts have not had any detainment. As a part of data cleaning, we made several assumptions while parsing data on age and sex. For instance, age 'under 18' was considered '18'.

In addition, issues such as a small sample size make our results less reliable. Our residual versus fitted plot shows that there is some evidence of heteroskedasticity in the data and the qq plot indicates that the errors are not normally distributed which again affects the precision of our estimates.

7. References

- https://themimu.info/sites/themimu.info/files/documents/MIMU_BaselineData_IndicatorList_18Mar2021.pdf (https://themimu.info/sites/themimu.info/files/documents/MIMU_BaselineData_IndicatorList_18Mar2021.pdf)

- https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Codebook_v1_January-2021.pdf (https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Codebook_v1_January-2021.pdf)
- https://dhsprogram.com/topics/wealth-index/ (https://dhsprogram.com/topics/wealth-index/)
- https://www.sharpsightlabs.com/blog/mutate-in-r/ (https://www.sharpsightlabs.com/blog/mutate-in-r/)
- https://dplyr.tidyverse.org/reference/mutate.html (https://dplyr.tidyverse.org/reference/mutate.html)