Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

**R-Project: Group E2**

# Myanmar Military Coup of 2021

## Abstract

This report looks at records of individuals who were detained, imprisoned, or killed during the Myanmar Coup of 2021. We conduct a statistical analysis on the datasets to understand detainment better across age, sex and township. We also merge data related to the socioeconomic factors in different township in Myanmar. We then see what factors are responsible to contributing to the detainment in Myanmar during the coup.

## Introduction

1st February 2021 marks the military coup in Myanmar. The military overthrew democratic leader Aung San Suu Kyi (supported by the National League of Democracy),

Since the coup started the there have been violent and non violent protests between the protestors of the coup and the administrative forces. Throughout these clashes, civilians and opposition forces have been detained, imprisoned, and killed. The Military initially did not respond with force, however when civil disobedience and general strikes grew the military became more aggressive as well

This response rose to the level of escalating violence between the two opposing groups when two protestors were killed. Millions of people in Myanmar went out to protest on the streets. Since then, the protests and number of imprisoned, detained and killed continued to rise.

In our research of the Coup in Myanmar we are researching how the coup impacted the number of people who were detained, imprisoned, and killed across the country. We hope to gain knowledge about few key characteristics.

Primarily we are concerned about the in-discriminant mistreatment of the protestors:

1) We believe that for this democratic cause people from all age groups, sexes and townships participaed. However, we are trying to see that did the military show restraint against people who are young and old; or did they detained, imprisoned, and killed individuals across all age groups.
    a. We also wish to see that if there was a bias when the government forces who mistreat or imprison one group over another
    b. How did the distribution of the detained, imprisoned and fallen varied across townships?

2) Given the history of Myanmar army in past, we hypothesize that their reactions to protest are in discriminant. For our purpose we try to hypothesize that Military did not discriminate

in the detainment and imprisonment of the protesters across different ages. We also hypothesis there was no difference in mean age between those in imprisoned and fallen groups.

3) We then get deeper into our analysis to see if there were conditions based on the township which contributed to higher detainment during the coup. We will conduct a regression exercise.

**Methods**

For the purpose of describing our methodology we have divided our project into distinct sections.

1) Data cleaning and merging: In our project we are using 5 datasets.
    a. Detained (information on people who were detained)
    b. Sentenced (information on people who were given sentences)
    c. Fallen (information people who were killed in conflict)
    d. MIMU (information on socio economic and geographical characteristics on Myanmar by region)
    e. Conflicts information on conflicts and protests in Myanmar in 2021

Before we could do our analysis, we had to merge our datasets. Joining datasets on a common township column such as townships is easy. However the problem that we faced was while the MIMU and ACLED data have well aligned township names, the AAPP data townships are based on free text inputs. Some of the datasets have a column for township, while others simply have an address (with a township embedded). To be able to merge the data at the township level, we will take an approximate matching approach.

We first cleaned the name, sex, age, status, date_of_arrest, township, and region state columns. We concatenated them to each other in a new dataset. We then joined this new dataset to MIMU township columns so that we are able to standardize the names of townships

across detainee, imprisoned and fallen datasets. We then were able to aggregate the results of detainee, imprisoned and fallen across townships.

**Data Description and Summarization:**

For data description and summarization, we choose few key variables to display our data on. We found the attributes of :

1) Age
2) Sex
3) state_region
4) township

We identify how detainment was distributed amongst these categories.

To explain the detainment data we summarize:

1) Detainees per township
2) Detainees_per_capita per township
3) Detainee per township separated by sex
4) Median age of detainees per township by each sex category
5) Median age of detainees per region by each sex category
6) Proportion of males and female detained in each township as a percentage of total male and female detained
7) Proportion of males and female detained in each state_region as a percentage of total male and female detained
8) Distribution of Detainee, imprisoned and Fallen by state_region
9) Proportion of detainees per date of event
10) Distribution on detainee by month separated by sex

**Statistical Testing**

For this question we wanted to focus on the age of the people who were detained, imprisoned or killed during the coup protest.

Myanmar has been ruled by a military junta for many of the years since it gained independence from British colonial rule in 1948. In 1988 there were widespread protest the military role in the government. In August of the same year army acted against the protesters which resulted in the death of at least 300 and displacing thousands[1]. We know that the military killed people indiscriminately in 1988. We hypothesize the military takes actions against the protesters

[1] https://www.cfr.org/backgrounder/myanmar-history-coup-military-rule-ethnic-conflict-rohingya

indiscriminately and does not take the age into consideration when sentencing them after they have been detained. We test two separate hypotheses.

**Hypothesis 1:**

H0:There is no difference between the age averages of the detained and imprisoned

Versus

Ha:There is a difference between the age averages of the detained and imprisoned

**Hypothesis 2:**

H0:There is no difference between the age averages of the detained and killed

Versus

Ha:There is a difference between the age averages of the detained and killed

**Regression:**

We have created our linear regression model to check if predictors from the MIMU data frame have any effect whatsoever to predict the number of instances of events (detainees, fallen and imprisoned) per capita on a township level.

For our first plot we will be trying to plot how the attributes of number of employed people by township (Employees), total number of households with internet access (Households_Internet) by township, the number of people 4ategorized as vulnerable population per township (Vulnerable_Population) and the population per township.

Methodology:

We will first check if these predictors are statistically relevant by checking their P values. And then check if the predictors have any collinearity between each other to ensure our results are not asked. We will also check if any of these variables have high correlation amongst each other. And drop the variables which lead to high correlation with other variables or are statistically insignificant.

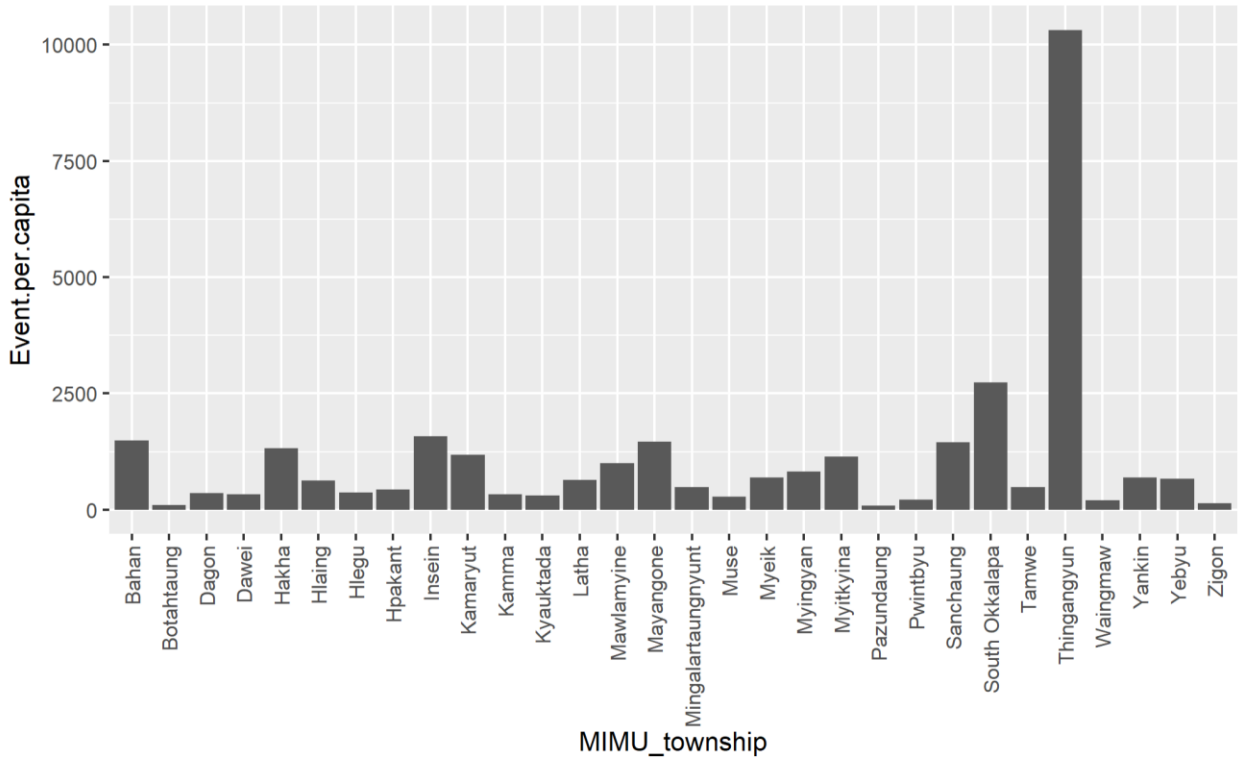
We will then plot another regression to check if the reduced number of statistically significant variables are able to improve the results for our regression.

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

**Results and Inferences of result:**

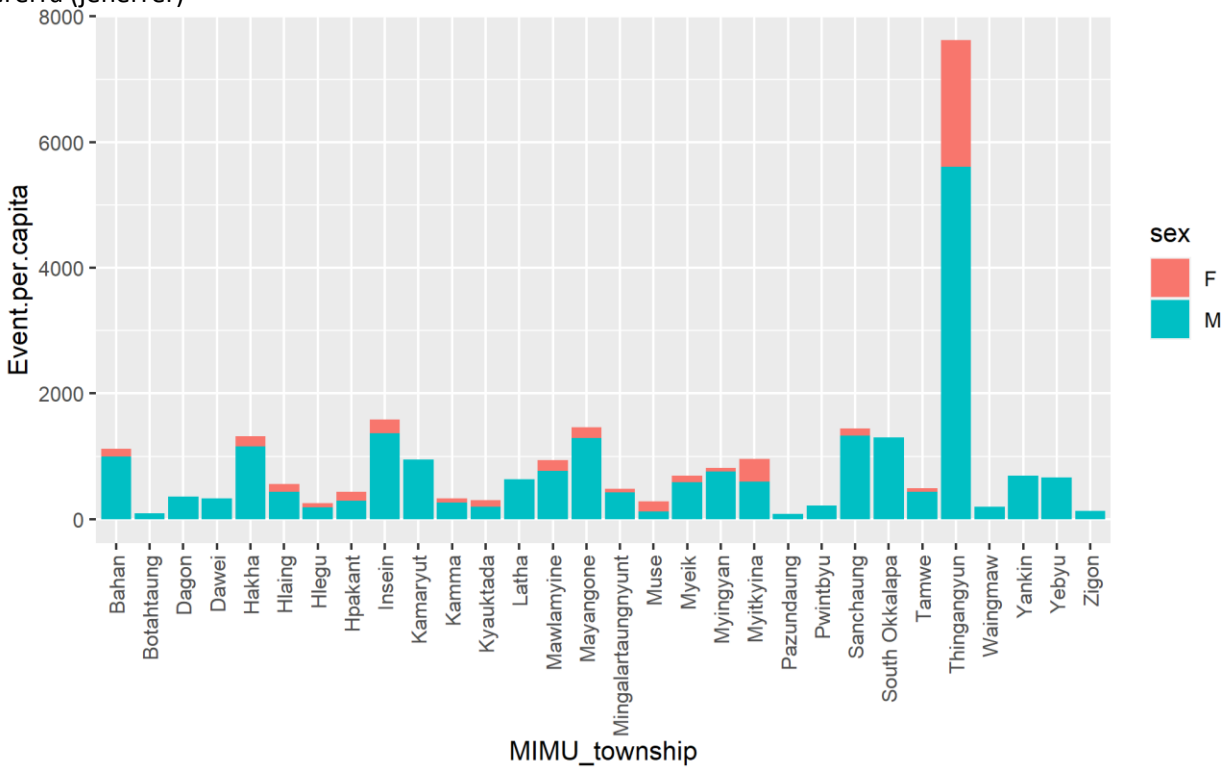
**Data description results:**

From our initial summation we try to see how detainment per capital were distributed across different townships. We see that the minimum number of detainment per capita was of were found to be in Thanbyuzayat where the min detainee per capita were 2.954. The highest figure were found to be in Thingangyun where the number was 224.35




- Thingangyun did in fact had the highest number of detainees across all townships. We also see that the greatest number of detainment happened here suggesting two things. Either the violence was substantially more prevalent here compared to other townships or it could also mean that the forces here had lower tolerance for any display of protest and would detain protesters right away.

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)




In all townships more males have been imprisoned than females. Again, Thingangyun has the highest number of both male and female events per capita.

From our results we see that minimum median age is 11 and it belongs to Pathein. The highest median age of detainee is of 70 and it belongs to Tedim. The highest median age by sex of detainee is of 49.5 and it belongs to females in Bago.

We also found out that Thingangyun has the highest number of detained by sex as a proportion of total male and females who were detained.

When comparing detainees with imprisoned and fallen, Yangon had the most number of detainee (240), imprisoned (171) and fallen (24). This is very important information as it shows Yangon had been the center of military action against the protesters. When we do the analysis for township level data, we see that highest number of fallen happened in Thingangyun and had a total number of 11.

Our analysis also showed that the maximum number of people detained were in March. Since the coup began in February, the military was taking less actions against the protesters. We Know that protests increased after two protesters were killed in March. This results accurately capture that most people were arrested in March.

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

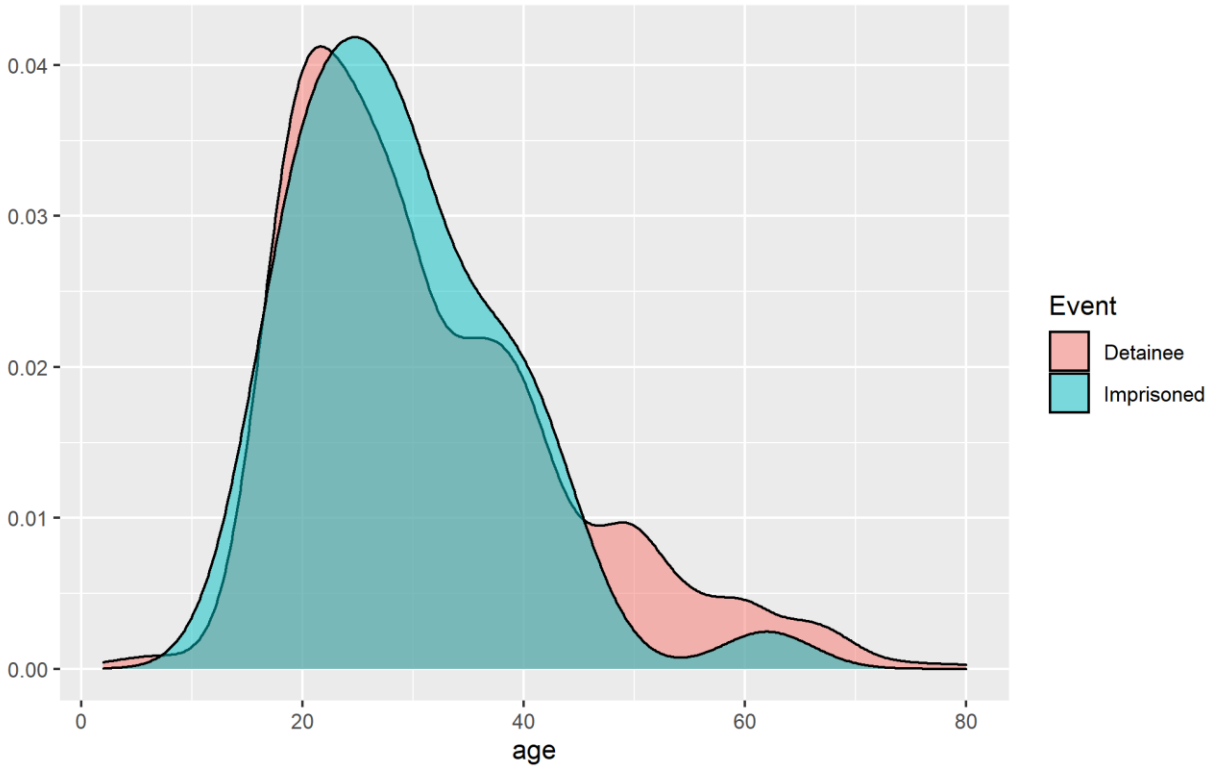
**Statistical testing.**

Hypothesis 1: We believe that there will be no difference in mean age of those that were detained and imprisoned.

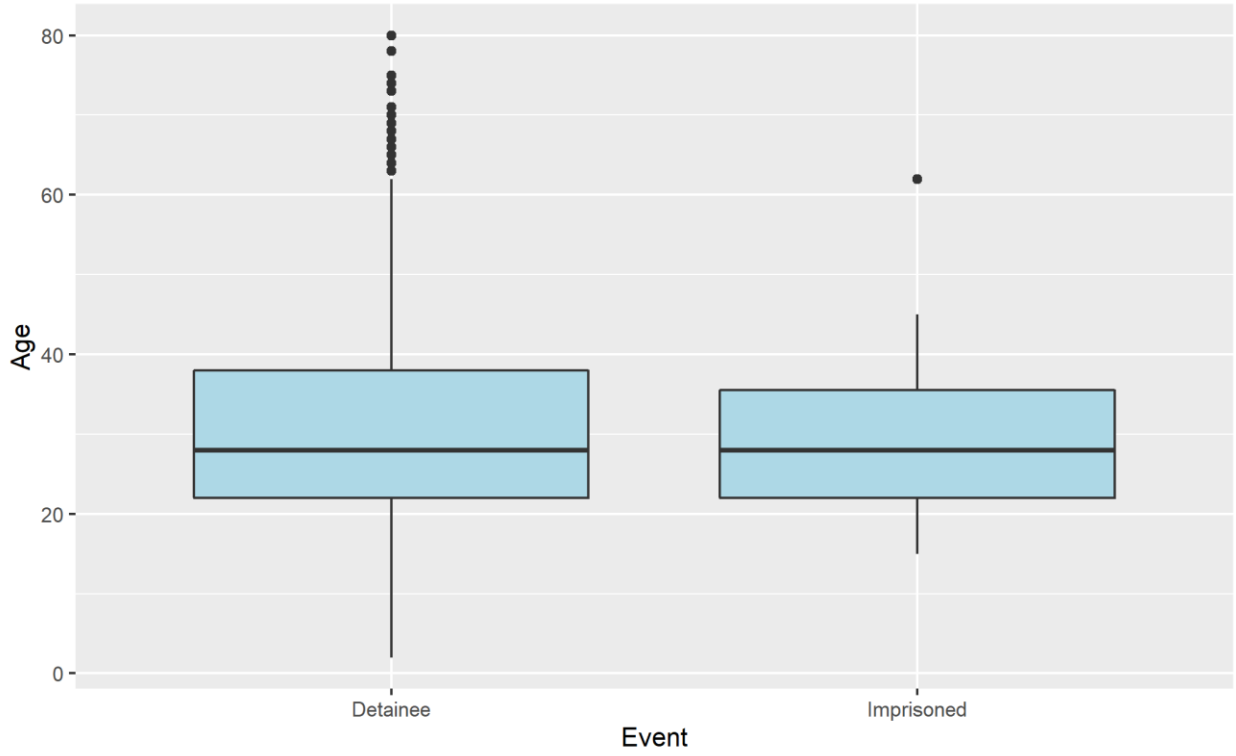
H0:There is no difference between the age averages of the detained and imprisoned

Versus

Ha:There is a difference between the age averages of the detained and imprisoned



Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)




The distribution shows that both detainees and imprisoned are distributed similarly over age. We see that both distributions have a right skew. From first looking at it we see that there is not a substantial difference in the mean of two categories. However, we will continue to use our testing methodology to get the answer

We also see that the mean age of the detained and imprisoned is about the same. We also see that detainees are slightly more varied than the imprisoned over the age group. This shows that there are no people in certain age group who have a higher likelyhood to get sentenced an imprisonment

We first perform the T-test

Welch Two Sample t-test

data:  age by Event

t = 1.4408, df = 40.732, p-value = 0.1573

alternative hypothesis: true difference in means between group Detainee and group Imprisoned is not equal to 0

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)
95 percent confidence interval:

-0.908723 5.429926

sample estimates:

mean in group Detainee mean in group Imprisoned

31.31188 29.05128

The results show that there is a slight difference in mean age of detained and imprisoned. However, the p-value associated with the test is 0.1572911 so we cannot reject the null hypothesis (H0) of no difference between the (true) averages of the two groups since the p-value is greater than the usual significance level of 0.05. Based on these data, we conclude that there is not enough evidence of a difference between the (true) averages of the two groups at the usual significance level of alpha = 0.05.

We then perform Wilcox Test

Wilcoxon rank sum test with continuity correction

data: age by Event

W = 40030, p-value = 0.4845

alternative hypothesis: true location shift is not equal to 0

The results show that there is a slight difference in mean age of detained and imprisoned. However, the p-value associated with the test is so we cannot reject the null hypothesis (H0) of no difference between the (true) averages of the two groups since the p-value is greater than the usual significance level alpha = 0.05.

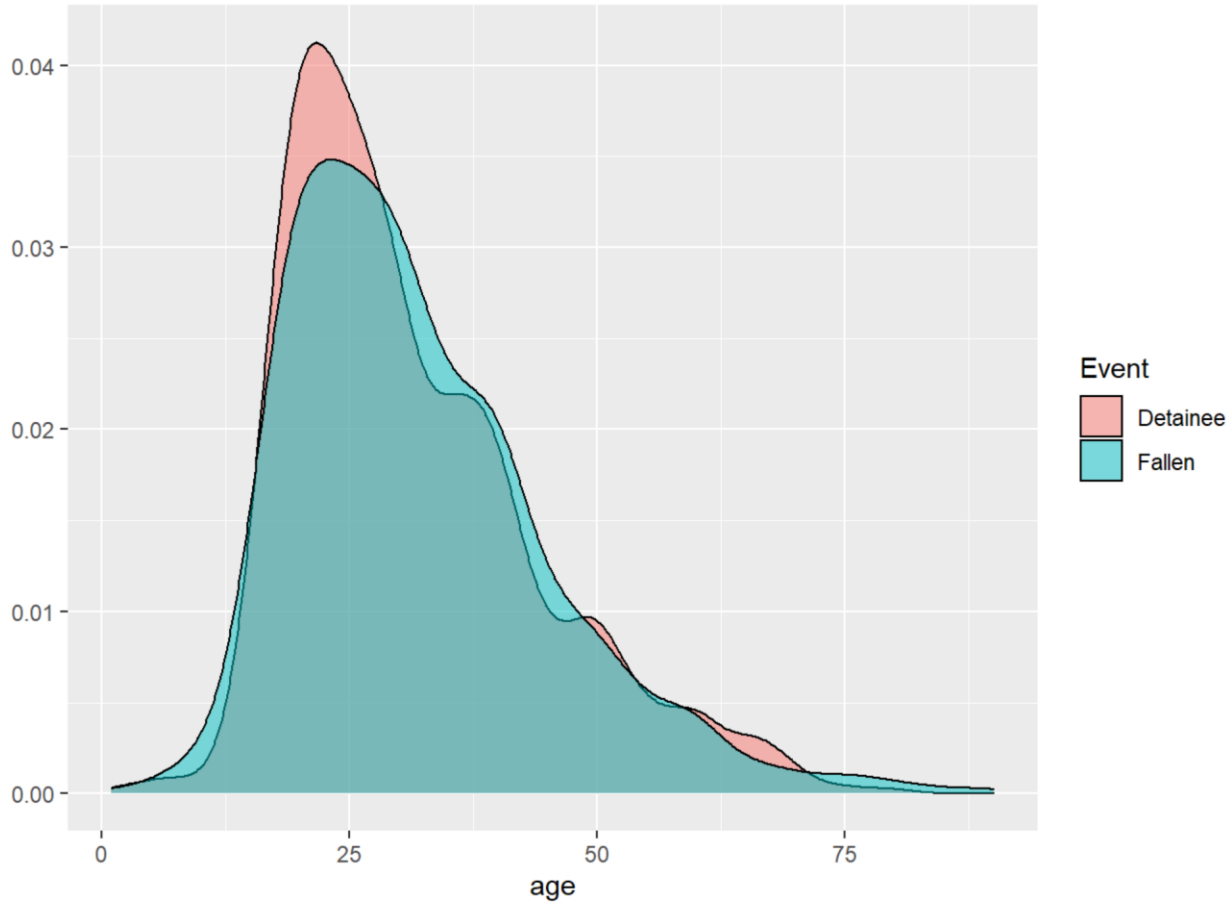
This is similar result to what we found in Welch two sample t-test. Hence we can conclude that there is no statistically significant difference in the median age of the two groups.

Hypothesis 2: We believe that there will be no difference in mean age of those that were detained and fallen.

H0:There is no difference between the age averages of the detained and imprisoned

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)
Versus

Ha:There is a difference between the age averages of the detained and imprisoned




The results show that there is a slight difference in mean age of detained and fallen. However, the p-value associated with the test is 0.2804472 so we cannot reject the null hypothesis (H0) of no difference between the (true) averages of the two groups since the p-value is greater than the usual significance level 0.05. Based on these data, we conclude that there is not enough evidence of a difference between the average ages of the two groups at the usual significance level of 0.05.

**Is there any association between the events and gender?**

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

We will run a chi-square analysis to see if there is any association between the Events and Gender, in other words we aim to explain if there is any uneven distribution on how the type of event affects the population based on gender.

Our null hypothesis state that there is no significant difference between gender and type of event.

Our alternative hypothesis state that the gender is related to the type of events that affect the population.

Our significance level is 95%, alpha = 0.05. When performing the chisq test, we find that the p-value is less than 0.05. We reject the null hypothesis, in Myanmmar the coup events are related to gender.

| | Sex | Detainee | Fallen | Imprisoned |
|---|---|---|---|---|
| | <chr> | <int> | <int> | <int> |
| 1 | F | 264 | 80 | 53 |
| 2 | M | 1613 | 1055 | 248 |

Pearson's Chi-squared test

data: event_gender

X-squared = 42.947, df = 2, p-value = 4.722e-10

We will also conduct a fisher test between event and gender, considering only two events: detainee and fallen. We will run a fisher test analysis to see if there is any association between Events and Gender, in other words we aim to explain if there is any uneven distribution on how the type of events detainment and fallen affect the population based on gender. Our null hypothesis state that there is no significant difference between gender and these type of events. Our alternative hypothesis state that the gender is related to the type of events that affect the population. Our significance level is 95%, alpha = 0.05.

When performing the fisher test, we find that the pvalue 1.8527866^{-9} is less than 0.05. We reject the null hypothesis, in Myanmmar the proportions of detainee and fallen are associated to gender.

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

## Results for Regression

1> First regression model:

We conducted our regression to predict events per capita on the basis of picking 5 variables from our dataframe that we felt would be the most relevant for our regression

```
lm(formula = Event.per.capita ~ Vulnerable_Population.per.1000 +
    Female_Leaded_Houses.per.1000 + Population_completed_Middle_school.per.1000 +
    Households_Internet.per.1000 + Employees.per.1000, data = df_conflicts_det)

Residuals:
    Min      1Q  Median      3Q     Max
-55.244 -17.316  -8.322   9.574 141.077

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                             140.81599   58.13310   2.422   0.0176 *
Vulnerable_Population.per.1000           -0.15353    0.06622  -2.319   0.0229 *
Female_Leaded_Houses.per.1000            -0.21014    0.29965  -0.701   0.4851
Population_completed_Middle_school.per.1000  -0.09135    0.18078  -0.505   0.6147
Households_Internet.per.1000              0.43922    0.27188   1.615   0.1100
Employees.per.1000                       -0.11822    0.08367  -1.413   0.1614
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 32.99 on 82 degrees of freedom
Multiple R-squared:  0.4312,    Adjusted R-squared:  0.3966
F-statistic: 12.43 on 5 and 82 DF,  p-value: 5.493e-09
```

After conducting the regression we find out that the preidictors – numberof people employed per 1000 (Employees.per.1000), no of houses with internet connection (Households_Internet.per.1000) , number of female led houses (Female_Leaded_Houses.per.1000), number of people who completed middle school per thousand (Population_completed_Middle_school.per.1000) are statistically insignificant because of their p values being lower than 0.05.

We also performed a collinearity condition between the five chosen predictors only to find out that the independent variables have very little corelation amongst them

Vulnerable_population.per.1000 is the only statistically significant predictor. we will be dropping the predictors from our regression which are statistically insignificant.

2> Second regression model:

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

```
Call:
lm(formula = Event.per.capita ~ Vulnerable_Population.per.1000,
    data = df_conflicts_det)

Residuals:
    Min      1Q  Median      3Q     Max
-60.098 -18.763  -6.163  13.288 142.433

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 124.95490   12.43190  10.051 3.59e-16 ***
Vulnerable_Population.per.1000  -0.18317    0.02467  -7.424 7.63e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.34 on 86 degrees of freedom
Multiple R-squared:  0.3906,    Adjusted R-squared:  0.3835
F-statistic: 55.11 on 1 and 86 DF,  p-value: 7.629e-11
```

By dropping the statistically insignificant variables we see that our R squared values has decreased from 0.431 to 0.39

We can only conclude from this that there are most certainly other variables that we haven't included, that could better explain the regression.

We havent tried other regression models. we need to check if a ridge, lasso or polynomial regression could better explain the regression.

**Limitations of your analysis**:

1) We used linear regression in our model. We found out that the linear regression model was not a good for the dataset we were working on. The linear model was assuming the regression as linear when this would not be the case. We believe that for this purpose we should try using polynomial regression so that it is better able to fit the dataset. Ridge and Lasso can also be used to see if they fit the model better. A non-linear model will probably provide better adjusted R square, smaller MSE and better coefficients for the independent variables.

2) We were unable to identify features based on mathematical tools which would help us decide if a particular variable should be included in the regression model. Feature selection should be performed so that variables are not chosen on intuition but on mathematical calculations

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

3) Township columns in the dataset were extremely messy. We used a function so that we can extract the name of the town from the address field. However even after doing so the names were very messy. This had a substantial impact on our project as after doing the join, the number of observations reduced dramatically.

4) Information in the MIMU dataset was not available for the year 2021. This adds bias to us as we are comparing 2021 observations with non 2021 data

**Contextualization to policy and potential stakeholders:**

This research report will be very useful both for the administrative authority in Myanmar and human rights organizations like the UN and Amnesty.

**Administrative authority:**

The central administration is broken down into smaller entities or individuals who are responsible for maintaining the administration in different areas. In our case since the administrative authority is the Myanmar Military it is very possible that the administrative duty falls under a particular commander. These commanders can unanimously decide security policy

for the area they are in charge. Our analysis shows that the highest number of deaths and detainment occurred in Thingangyun. It is possible that the administrator in this township take im-proportionate degree of action against the protestors. For this purpose, the central administrative authority should look into to see if the actions of the local administrator is causing more violence.

The results highlighting where most violence takes place can be an indicator of the fact the needs of the local population are not being met. Maybe the protests are not completely related to imposition of the Coup, but that the coup could have introduced social issues like increased unemployment, slowing of economic activity or reduction in mobility. This report can urge the military to loosen its grip on the township or region where most of the incidents have taken place. This could also urge the military to introduce welfare projects in such area like Thingangyun, Kyauktada, Sanchaung, Hakha, etc. to subside the population dissatisfaction with the military.

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)

**Human Rights organizations like the UN and Amnesty:**

Such organizations are mandated to collect data on human conditions and analyze human rights violation. Our data is able to provide key facts about how the imprisonment and killings are taking place in different regions of the country. This can also help them identify which people to provide ai to first.

This report can help these organizations understand how the socially vulnerable segments of population like women and children are being treated during the coup. This information will be very useful if the results are to show that the military is taking part in internationally frowned practices of indiscriminate killing and imprisonment. The report suggests that since the average minimum age of imprisonment is very low the military has overstepped its limitations. This data can be used to encourage foreign governments to pressurize the Myanmar Military over its immoral practices.

**Conclusion:**

In conclusion, as a result of our study we have found some key characteristics about the detainment, imprisonment and killings during the coup.

We found that minimum number of detainment per capita was of were found to be in Thanbyuzayat where the min detainee per capita were 2.954. The highest figure was found to be in Thingangyun where the number was 224.35. Thingangyun did in fact had the highest number of detainees across all townships. We also see that the greatest number of detainment happened here suggesting two things. Either the violence was substantially more here compared to other townships or it could also mean that the forces here had lower tolerance for any display of protest and would detain protesters right away.

From our results we see that minimum median age is 11 and it belongs to Pathein. The highest median age of detainee is of 70 and it belongs to Tedim. The highest median age by sex of detainee is of 49.5 and it belongs to females in Bago. that there is a slight difference in mean age of detained and fallen.

**Future Directions:**

Abdullah Ahmad(aahamd1)
Shantanu Samant(sssamant)
Jesus Hererra (jeherrer)
For teams who will be working to analyze these data sets further we encourage them to:

- Find a better method of identifying the township individuals who were detained, imprisoned, and killed belong to. This proved the most problematic for us since the only common column between the datasets on which we could join was extremely messy. We suggest that future teams should use the latitude and longitude coordinates to make joins between the different datasets.

- Future work should conduct better methods of feature selection for regression exercises. This will get rid of the bias that was introduced by our intuition.

- Regression models other than a multivariable regression should be used so that true correlation between the dependent and independent variables can be identified. As a result of which better predictions can be made.

- Better supporting datasets should be used which provide the accurate socio economic attributes to be included in the analysis.