

Final Report

Group L: lingyunf, yacheng, yizix

2021-12-08

- ABSTRACT
- INTRODUCTION
- METHODS
 - Data Preprocessing
 - Data Summarizations
 - Further Research: table and figure visualizations
 - Statistical Test
 - Data Summarizations on Merged Data
- MODELS AND ANALYSIS
 - Regression Model 1: regression on all variables
 - Regression Model 2: regression on selected variables
 - Regression Model 3: log calculation on dependent variable
 - Regression Model 4: use selected variables to regress on log(dependent variable)
- RESULTS
- DISCUSSIONS
 - Limitations of the Analysis
 - Inferences from Results
 - Potential Stakeholders
 - Contextualization and Policy
 - Conclusions
 - Future Directions

Code



ABSTRACT

This project is an analysis of The Myanmar coup of 2021. The main focus is on the economic and social characteristics of the area (township level) where the people were arrested and died in the coup. We select the characteristic variables through exploratory descriptive statistics and perform statistics by means of hypothesis testing and regression analysis. We draw the conclusion that the number of detainees is significantly related to the employment status of their area, the food poverty index, the literacy rate, and the total number of conflicts in the past period. As a result of the above analysis, we believe that it is feasible to improve the local economic situation as a starting point in order to stabilize the situation in Burma.

INTRODUCTION

In early February 2021, there was an army mutiny in Burma. This mutiny had a huge impact, with many citizens and opposition forces being detained, imprisoned and even killed. We hope that we can draw some useful conclusions by exploring the characteristics of these affected individuals themselves and the areas where they were located.

In our study the main focus of: Dependent Variable: **Detainees per 1000**

Independent Variable: **Economy** (Poverty Ratio, Food Poverty Ratio, Poverty Gap); **Sex**; **Employment** (Employee Number); **Education** (Female, Male, Total).

We hope that through our analysis, we can give the Burmese authorities or international organizations concerned about the situation in Burma some insights related to ameliorating the chaos and promoting stability in Burma.

METHODS

- Data preprocessing
- Descriptive statistical analysis
- Exploratory statistics
- Hypothesis testing (mainly t-tests)
- Linear regression (including simple linear regression and multiple linear regression)

Data Preprocessing

Data loading and cleaning

Get the main dataset of detainees, imprisoned, fallen. Clean the dataset as needed.

Code

Read the conflicts (ACLED) data.

Code

Read the sector indicators (MIMU) data.

Code

```
## Response [https://www.andrew.cmu.edu/user/jweiss2/21f_r/94842/final_2021/MIMU_BaselineData_AllSectors_Countrywide_18Mar2021_revised.xlsx]
##   Date: 2021-12-10 04:38
##   Status: 200
##   Content-Type: application/vnd.ms-excel.sheet.macroEnabled.12
##   Size: 19.3 MB
## <ON DISK>   /var/folders/s8/65bc3d5x649_gpdcbjfqgzbw0000gn/T//RtmpV3f5R7/file185b2e9aa157.xlsx
```

Code

Code

Data merging

Merge data and calculate the number of detainees per 1000

Code

In the conflicts (ACLED) dataset, total number of conflicts per township

Code

```
## # A tibble: 5 × 2
## # Groups:   Township_Name [5]
##   Township_Name conflicts.num
##   <chr>          <int>
## 1 Ahlone          22
## 2 Amarapura       58
## 3 Ann             12
## 4 Aunglan          20
## 5 Aungmyaythazan 109
```

Data preprocessing on MIMU

In the township level, extract the total employee number per town

Code

```
## # A tibble: 5 × 30
##   State_Region Township_Name Township_Pcode Sector Sub_Sector Indicator_Name
##   <chr>         <chr>         <chr>         <chr> <chr>         <chr>
## 1 Ayeyarwady   Bogale           MMR017024     Economy Usual Acti... Population (10 ...
## 2 Ayeyarwady   Danabyu          MMR017022     Economy Usual Acti... Population (10 ...
## 3 Ayeyarwady   Dedaye           MMR017026     Economy Usual Acti... Population (10 ...
## 4 Ayeyarwady   Einme            MMR017015     Economy Usual Acti... Population (10 ...
## 5 Ayeyarwady   Hinthada         MMR017008     Economy Usual Acti... Population (10 ...
## # ... with 24 more variables: Indicator_Type <chr>, Unit <chr>, 2009-2010 <dbl>,
## #   2010 <dbl>, 2010-2011 <dbl>, 2011 <dbl>, 2011-2012 <dbl>, 2012 <dbl>,
## #   2012-2013 <dbl>, 2013 <dbl>, 2013-2014 <dbl>, 2014 <dbl>, 2014-2015 <dbl>,
## #   2015 <dbl>, 2015-2016 <dbl>, 2016 <dbl>, 2016-2017 <dbl>, 2017 <dbl>,
## #   2017-2018 <dbl>, 2018 <dbl>, 2018-2019 <dbl>, 2019 <dbl>, 2019-2020 <dbl>,
## #   Source_Name <chr>
```

Code

```
## [1] 0
```

For the employment data, there is no missing on the column of “2014”, so we use the data on employee of 2014 for analysis use.

In the township level, extract the education and sex data

Code

```
## State_Region Township_Name fem.literacy male.literacy total.literacy
## 1 Yangon Ahlone 97.9 99.2 98.5
## 2 Mandalay Amarapura 93.8 97.8 95.6
## 3 Rakhine Ann 71.1 88.0 79.0
## 4 Magway Aunglan 90.5 96.9 93.4
## 5 Mandalay Aungmyaythazan 95.8 98.8 97.1
```

Code

```
## # A tibble: 5 × 4
## State_Region Township_Name Indicator_Type `2014`
## <chr> <chr> <chr> <dbl>
## 1 Ayeyarwady Bogale Total 97.5
## 2 Ayeyarwady Danubyu Total 91.7
## 3 Ayeyarwady Dedaye Total 96.4
## 4 Ayeyarwady Einme Total 95.1
## 5 Ayeyarwady Hinthada Total 89.3
```

In the state level, extract the income inequality value: - “Poverty headcount ratio” - “Food poverty headcount index” - “Poverty gap ratio”

Code

```
## # A tibble: 5 × 30
## State_Region SR_Pcode Sector Sub_Sector Indicator_Name Indicator_Type Unit
## <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Ayeyarwady MMR017 Economy Income Poverty headcou... Total Perc...
## 2 Bago (East) MMR007 Economy Income Poverty headcou... Total Perc...
## 3 Bago (West) MMR008 Economy Income Poverty headcou... Total Perc...
## 4 Chin MMR004 Economy Income Poverty headcou... Total Perc...
## 5 Kachin MMR001 Economy Income Poverty headcou... Total Perc...
## # ... with 23 more variables: 2009-2010 <dbl>, 2010 <dbl>, 2010-2011 <dbl>,
## # 2011 <dbl>, 2011-2012 <dbl>, 2012 <dbl>, 2012-2013 <dbl>, 2013 <dbl>,
## # 2013-2014 <dbl>, 2014 <dbl>, 2014-2015 <dbl>, 2015 <dbl>, 2015-2016 <dbl>,
## # 2016 <dbl>, 2016-2017 <dbl>, 2017 <dbl>, 2017-2018 <dbl>, 2018 <dbl>,
## # 2018-2019 <dbl>, 2019 <dbl>, 2019-2020 <dbl>, 2020 <dbl>, Source_Name <chr>
```

Code

```
## [1] 0
```

Code

```
## # A tibble: 5 × 30
## State_Region SR_Pcode Sector Sub_Sector Indicator_Name Indicator_Type Unit
## <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Ayeyarwady MMR017 Economy Income Food poverty he... Total Perc...
## 2 Bago (East) MMR007 Economy Income Food poverty he... Total Perc...
## 3 Bago (West) MMR008 Economy Income Food poverty he... Total Perc...
## 4 Chin MMR004 Economy Income Food poverty he... Total Perc...
## 5 Kachin MMR001 Economy Income Food poverty he... Total Perc...
## # ... with 23 more variables: 2009-2010 <dbl>, 2010 <dbl>, 2010-2011 <dbl>,
## # 2011 <dbl>, 2011-2012 <dbl>, 2012 <dbl>, 2012-2013 <dbl>, 2013 <dbl>,
## # 2013-2014 <dbl>, 2014 <dbl>, 2014-2015 <dbl>, 2015 <dbl>, 2015-2016 <dbl>,
## # 2016 <dbl>, 2016-2017 <dbl>, 2017 <dbl>, 2017-2018 <dbl>, 2018 <dbl>,
## # 2018-2019 <dbl>, 2019 <dbl>, 2019-2020 <dbl>, 2020 <dbl>, Source_Name <chr>
```

Code

```
## [1] 0
```

Code

```
## # A tibble: 5 × 30
##   State_Region SR_Pcode Sector Sub_Sector Indicator_Name Indicator_Type Unit
##   <chr>      <chr>   <chr>  <chr>      <chr>      <chr>      <chr>
## 1 Ayeyarwady MMR017 Economy Income Poverty gap rat... Total Perc...
## 2 Bago (East) MMR007 Economy Income Poverty gap rat... Total Perc...
## 3 Bago (West) MMR008 Economy Income Poverty gap rat... Total Perc...
## 4 Chin       MMR004 Economy Income Poverty gap rat... Total Perc...
## 5 Kachin     MMR001 Economy Income Poverty gap rat... Total Perc...
## # ... with 23 more variables: 2009-2010 <dbl>, 2010 <dbl>, 2010-2011 <dbl>,
## #   2011 <dbl>, 2011-2012 <dbl>, 2012 <dbl>, 2012-2013 <dbl>, 2013 <dbl>,
## #   2013-2014 <dbl>, 2014 <dbl>, 2014-2015 <dbl>, 2015 <dbl>, 2015-2016 <dbl>,
## #   2016 <dbl>, 2016-2017 <dbl>, 2017 <dbl>, 2017-2018 <dbl>, 2018 <dbl>,
## #   2018-2019 <dbl>, 2019 <dbl>, 2019-2020 <dbl>, 2020 <dbl>, Source_Name <chr>
```

Code

```
## [1] 0
```

For the income data, there is no missing on the column of “2010”, so we use the data on income of 2010 for analysis use.

Merge all the data into 1 tibble

Code

```
## # A tibble: 330 × 30
##   State_Region Township_Name Township_Pcode Sector Sub_Sector Indicator_Name
##   <chr>      <chr>      <chr>      <chr>  <chr>      <chr>
## 1 Yangon    Ahlone      MMR013037 Economy Usual Acti... Population (1...
## 2 Mandalay  Amarapura   MMR010006 Economy Usual Acti... Population (1...
## 3 Rakhine   Ann         MMR012014 Economy Usual Acti... Population (1...
## 4 Magway    Aunglan     MMR009016 Economy Usual Acti... Population (1...
## 5 Mandalay  Aungmyaythazan MMR010001 Economy Usual Acti... Population (1...
## 6 Sagaing   Ayadaw      MMR005014 Economy Usual Acti... Population (1...
## 7 Bago (East) Bago        MMR007001 Economy Usual Acti... Population (1...
## 8 Yangon    Bahan       MMR013044 Economy Usual Acti... Population (1...
## 9 Sagaing   Banmauk     MMR005023 Economy Usual Acti... Population (1...
## 10 Kayah    Bawlake     MMR002005 Economy Usual Acti... Population (1...
## # ... with 320 more rows, and 24 more variables: Indicator_Type <chr>,
## #   Unit <chr>, 2009-2010 <dbl>, 2010 <dbl>, 2010-2011 <dbl>, 2011 <dbl>,
## #   2011-2012 <dbl>, 2012 <dbl>, 2012-2013 <dbl>, 2013 <dbl>, 2013-2014 <dbl>,
## #   2014 <dbl>, 2014-2015 <dbl>, 2015 <dbl>, 2015-2016 <dbl>, 2016 <dbl>,
## #   2016-2017 <dbl>, 2017 <dbl>, 2017-2018 <dbl>, 2018 <dbl>, 2018-2019 <dbl>,
## #   2019 <dbl>, 2019-2020 <dbl>, Source_Name <chr>
```

Code

```
## # A tibble: 172 × 3
##   MIMU_township detainees detainees.per.1000
##   <chr>      <dbl>      <dbl>
## 1 Ahlone      1          0.0180
## 2 Aunglan     1          0.00425
## 3 Ayadaw     10          0.0642
## 4 Bago        44          0.0895
## 5 Bahan       17          0.176
## 6 Bilin       1          0.00552
## 7 Bogale      2          0.00620
## 8 Botahtaung  7          0.171
## 9 Budalin     5          0.0405
## 10 Chauk      1          0.00540
## # ... with 162 more rows
```

Code

```
## # A tibble: 169 × 13
##   MIMU_township State_Region detainees.per.1000 detainees employee.num
##   <chr>         <chr>         <dbl>     <dbl>     <dbl>
## 1 Ahlone       Yangon             0.0180         1       1115
## 2 Aunglan      Magway             0.00425        1       4481
## 3 Ayadaw       Sagaing            0.0642        10       2021
## 4 Bago         Bago (East)        0.0895        44       9162
## 5 Bahan       Yangon             0.176         17       2641
## 6 Bilin        Mon                0.00552        1       3424
## 7 Bogale       Ayeyarwady         0.00620        2      10602
## 8 Botahtaung  Yangon             0.171         7        1001
## 9 Budalin     Sagaing            0.0405         5       2767
## 10 Chauk      Magway             0.00540        1       3390
## # ... with 159 more rows, and 8 more variables: poverty.ratio <dbl>,
## #   food.poverty.index <dbl>, poverty.gap.ratio <dbl>, fem.literacy <dbl>,
## #   male.literacy <dbl>, total.literacy <dbl>, sex.ratio <dbl>,
## #   conflicts.num <int>
```

Data Summarizations

data summarizations on Detainees dataset and other variables




Code

```
##           No           Name           Sex           Age
## Min.      : 1   Length:6709   Length:6709   Length:6709
## 1st Qu.:2066   Class :character   Class :character   Class :character
## Median :3771   Mode  :character   Mode  :character   Mode  :character
## Mean      :3717
## 3rd Qu.:5447
## Max.      :7194
## NA's      :4
## Father's Name      Status      Date of Arrest      Section of Law
## Length:6709      Length:6709      Length:6709      Length:6709
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## Plaintiff      Current Condition      Address      Region/State
## Length:6709      Length:6709      Length:6709      Length:6709
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## Remark      Township
## Length:6709      Length:6709
## Class :character   Class :character
## Mode  :character   Mode  :character
##
##
##
##
```

Code

```
## spec_tbl_df [6,709 × 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ No : num [1:6709] 1 2 3 4 5 6 7 8 9 10 ...
## $ Name : chr [1:6709] "Aung San Suu Kyi" "Win Myint" "Tun Tun Hein" "Dr. Zaw Myint\rMaung" ...
## $ Sex : chr [1:6709] "F" "M" "M" "M" ...
## $ Age : chr [1:6709] NA NA NA NA ...
## $ Father's Name : chr [1:6709] "General Aung San" "U Tun Kyin" "U Kya Hein" "U Chit Maung" ...
## $ Status : chr [1:6709] "MP (Pyithu\rHluttaw, Kawthmu\rTownship, Yangon\rRegion), \rGovernment (State\rCounsello
r), NLD\r(Chairperson)" "MP (Pyithu\rHluttaw, Tamwe\rTownship, Yangon\rRegion), \rGovernment\r(President), NLD\r(Vice Chairman
1)" "MP (Pyithu\rHluttaw, Naungcho\rTownship, Shan\rState), Pyithu\rHluttaw Deputy\rSpeaker, NLD\r(Central Executive" "MP (Regi
onal\rHluttaw, \rAmarapura\rTownship, \rMandalay Region), \rGovernment (Chief\rMinister of" ...
## $ Date of Arrest : chr [1:6709] "1-Feb-21" "1-Feb-21" "1-Feb-21 and 10-\rFeb-21" "1-Feb-21 and 7-\rFeb-21" ...
## $ Section of Law : chr [1:6709] "Export and Import\rLaw S: 8 and\rNatural Disaster\rManagement law S:\r25, Penal Code
S:\r505 (b), \rTelecommunic" | __truncated__ "Natural Disaster\rManagement law S:\r25, Penal Code S:\r505 (b)" NA "Penal Code S:
505\r(b), Natural\rDisaster\rManagement law S:\r25, 30, Anti-\rCorruption Law S:\r55" ...
## $ Plaintiff : chr [1:6709] "Superintendent\rKyι Linn of\rSpecial Branch, \rDekkhina District\rAdministrator (S:\r8 an
d 67), \rSuperintendent\r" | __truncated__ "Superintendent\rMyint Naing, \rDekkhina District\rAdministrator" NA "Deputy Director\rN
weni Khine of\rTownship\rGeneral\rAdministration\rDepartment" ...
## $ Current Condition: chr [1:6709] "House Arrest" "House Arrest" "Detained" "Detained in\rMandalay\rPrison" ...
## $ Address : chr [1:6709] "Naypyitaw" "Naypyitaw" "Naypyitaw" "Mandalay" ...
## $ Region/State : chr [1:6709] "Naypyitaw" "Naypyitaw" "Naypyitaw" "Mandalay" ...
## $ Remark : chr [1:6709] "Myanmar Military Seizes Power and\rSenior NLD leaders including Daw\rAung San Suu Kyι and
President U\rWin Myin" | __truncated__ "Senior NLD leaders including Daw\rAung San Suu Kyι and President U\rWin Myint were det
ained. The\rNLDâ€™s chief" | __truncated__ "Senior NLD leaders including Daw\rAung San Suu Kyι and President U\rWin Myint were
detained. The\rNLDâ€™s chief" | __truncated__ "Senior NLD leaders including Daw\rAung San Suu Kyι and President U\rWin Myint we
re detained. The\rNLDâ€™s chief" | __truncated__ ...
## $ Township : chr [1:6709] "Naypyitaw" "Naypyitaw" "Naypyitaw" "Mandalay" ...
## - attr(*, "spec")=
## .. cols(
## .. No = col_double(),
## .. Name = col_character(),
## .. Sex = col_character(),
## .. Age = col_character(),
## .. `Father's Name` = col_character(),
## .. Status = col_character(),
## .. `Date of Arrest` = col_character(),
## .. `Section of Law` = col_character(),
## .. Plaintiff = col_character(),
## .. `Current Condition` = col_character(),
## .. Address = col_character(),
## .. `Region/State` = col_character(),
## .. Remark = col_character()
## .. )
## - attr(*, "problems")=<externalptr>
```

Code

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
detainees.per.1000	171	0	0.1	0.1	0.0	0.0	0.9	 _
detainees	40	0	10.6	15.5	1.0	5.0	97.0	 _ _
employee.num	170	0	4111.1	3174.0	33.0	3351.0	18922.0	 _ _

Code

```
## MIMU_township      detainees.per.1000  detainees      State_Region
## Length:173        Min. :0.001454      Min. : 1.00      Length:173
## Class :character   1st Qu.:0.013888      1st Qu.: 2.00      Class :character
## Mode :character    Median :0.035709      Median : 5.00      Mode :character
##                   Mean :0.065149      Mean :10.62
##                   3rd Qu.:0.083757      3rd Qu.:11.00
##                   Max. :0.868666      Max. :97.00
##
## employee.num
## Min. : 33
## 1st Qu.: 1884
## Median : 3351
## Mean : 4111
## 3rd Qu.: 5637
## Max. :18922
```

Code

```
## tibble [173 × 5] (S3: tbl_df/tbl/data.frame)
## $ MIMU_township      : chr [1:173] "Monywa" "Myeik" "Thingangyun" "Insein" ...
## $ detainees.per.1000: num [1:173] 0.261 0.327 0.353 0.18 0.17 ...
## $ detainees          : num [1:173] 97 93 74 55 54 50 48 45 44 43 ...
## $ State_Region       : chr [1:173] "Sagaing" "Tanintharyi" "Yangon" "Yangon" ...
## $ employee.num       : num [1:173] 9411 2889 3438 3461 4132 ...
```

Data summarizations on the fallen and imprisoned dataset


Code

```
##      No.           Name           Sex           Age
## Length:895      Length:895      Length:895      Min.   : 1.00
## Class :character Class :character Class :character 1st Qu.:22.00
## Mode  :character Mode  :character Mode  :character Median :30.00
##                                     Mean  :31.81
##                                     3rd Qu.:39.00
##                                     Max.   :90.00
## Father's name     Date of Incident Deceased Date  Organization
## Length:895        Length:895        Length:895      Length:895
## Class :character  Class :character  Class :character Class :character
## Mode  :character  Mode  :character  Mode  :character Mode  :character
##
##
## Place of Incidents Home Address      Township      States/Regions
## Length:895          Length:895          Length:895      Length:895
## Class :character    Class :character    Class :character Class :character
## Mode  :character    Mode  :character    Mode  :character Mode  :character
##
##
## Remarks
## Length:895
## Class :character
## Mode  :character
##
##
```

Code

```
## spec_tbl_df [895 × 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ No. : chr [1:895] "1" "2" "3" "4" ...
## $ Name : chr [1:895] "Na Pwar (aka) Ko\rNyi Nyi Oo" "Mya Thwate Thwate\rKhaing" "Nay Nay Win Htet" "Thet Naing Win\r(aka) Min Min" ...
## $ Sex : chr [1:895] "M" "F" "M" "M" ...
## $ Age : num [1:895] 32 19 18 37 16 48 30 26 30 35 ...
## $ Father's name : chr [1:895] "U Hla Ngwe" "U Min Lwin" "Unknown" "U Maung\rSan" ...
## $ Date of Incident : chr [1:895] "08-Feb-21" "09-Feb-21" "15-Feb-21" "20-Feb-21" ...
## $ Deceased Date : chr [1:895] "8-Feb-21" "19-Feb-21" "15-Feb-21" "20-Feb-21" ...
## $ Organization : chr [1:895] "Civilian" "Student" "Civilian" "Civilian" ...
## $ Place of Incidents: chr [1:895] "Mandalay" "Naypyitaw" "Myeik,\rTanintharyi\rRegion" "Kannar Road,\rMandalay City" ...
## $ Home Address : chr [1:895] "75 Street,\rbetween 37 and\r38 Street" "Hlaykhwintaung,\rLower\rPaunglaung\rHydro Power \rProject" "Toe Chal Ward" "Near 41 Street" ...
## $ Township : chr [1:895] "Maha Aung\rMyay" "Zeyathiri" "Myeik" "Maha Aung\rMyay" ...
## $ States/Regions : chr [1:895] "Mandalay" "Naypyitaw" "Tanintharyi" "Mandalay" ...
## $ Remarks : chr [1:895] "In another incident, 32 year old Ko\rNa Pwar ((aka) Ko Ko Oo), died after\ra car intenti onally hit him at night\rin Mandalay." "On February 9, peaceful anti-coup\rprotests in Naypyitaw were\r suppressed using a water cannon,\rubber bullets"| __truncated__ "On 15 February evening, 18-year old\rMaung Nay Nay Win Htet was\rbeaten on his head to death while\rguarding a "| __truncated__ "In Mandalay, a shipyard raid\rturned violent on Saturday when\rsecurity forces open ed fire on\r demonstrators "| __truncated__ ...
## - attr(*, "spec")=
## .. cols(
## .. No. = col_character(),
## .. Name = col_character(),
## .. Sex = col_character(),
## .. Age = col_character(),
## .. `Father's name` = col_character(),
## .. `Date of Incident` = col_character(),
## .. `Deceased Date` = col_character(),
## .. Organization = col_character(),
## .. `Place of Incidents` = col_character(),
## .. `Home Address` = col_character(),
## .. Township = col_character(),
## .. `States/Regions` = col_character(),
## .. Remarks = col_character()
## .. )
## - attr(*, "problems")=<externalptr>
```

Code

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
Age	72	0	31.8	13.1	1.0	30.0	90.0	

Code

```
## # A tibble: 5 × 13
## No. Name Sex Age `Father's name` `Date of Incide...` `Deceased Date`
## <chr> <chr> <chr> <dbl> <chr> <chr> <chr>
## 1 1 "Na Pwar (... M 32 "U Hla Ngwe" 08-Feb-21 8-Feb-21
## 2 2 "Mya Thwat... F 19 "U Min Lwin" 09-Feb-21 19-Feb-21
## 3 3 "Nay Nay W... M 18 "Unknown" 15-Feb-21 15-Feb-21
## 4 4 "Thet Nain... M 37 "U Maung\rSan" 20-Feb-21 20-Feb-21
## 5 5 "Wai Yan T... M 16 "Unknown" 20-Feb-21 20-Feb-21
## # ... with 6 more variables: Organization <chr>, Place of Incidents <chr>,
## # Home Address <chr>, Township <chr>, States/Regions <chr>, Remarks <chr>
```

Code

##	No	Name	Sex /Age	Age
##	Min. : 1.00	Length:308	Length:308	Length:308
##	1st Qu. : 77.25	Class :character	Class :character	Class :character
##	Median :153.50	Mode :character	Mode :character	Mode :character
##	Mean :153.50			
##	3rd Qu. :229.75			
##	Max. :306.00			
##	NA's :2			
##	Father's Name	Status	Date of Arrest	Section of Law
##	Length:308	Length:308	Length:308	Length:308
##	Class :character	Class :character	Class :character	Class :character
##	Mode :character	Mode :character	Mode :character	Mode :character
##				
##				
##				
##	Plaintiff	Current Condition	Prison	Address
##	Length:308	Length:308	Length:308	Length:308
##	Class :character	Class :character	Class :character	Class :character
##	Mode :character	Mode :character	Mode :character	Mode :character
##				
##				
##				
##	Region/State	Remark		
##	Length:308	Length:308		
##	Class :character	Class :character		
##	Mode :character	Mode :character		
##				
##				
##				
##				

Code

```
## spec_tbl_df [308 × 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ No : num [1:308] 1 2 3 4 5 6 7 8 9 10 ...
## $ Name : chr [1:308] "Dr. Aung Moe\rNyo" "Nan Khin Htwe\rMyint" "Nyi Pu" "Bo Bo Wai\rMaung" ...
## $ Sex /Age : chr [1:308] "M" "F" "M" "M" ...
## $ Age : chr [1:308] "62" NA NA NA ...
## $ Father's Name : chr [1:308] "U Nyo" "U Saw Hla Tun" NA NA ...
## $ Status : chr [1:308] "MP (Regional Hluttaw,\rPwintbyu Township, Magwe\rRegion), Government (Chief\rMinister of
Magwe Region),\rNLD (Secretary)," "MP (State Hluttaw, Hpa-an\rTownship, Karen State),\rGovernment (Chief Minister\rrof Karen Sta
te), NLD (Central\r"| __truncated__ "MP (State Hluttaw, Gwa\rTownship, Rakhine State),\rGovernment (Chief Minister\rrof Rakhine
State), NLD\r(Central"| __truncated__ "MP (State Hluttaw, No (1)\rConstituency of Kyarinn\rSeikkyee Township, Karen\rState), Go
vernment (Minister\rrof " | __truncated__ ...
## $ Date of Arrest : chr [1:308] "1-Feb-21" "1-Feb-21 and\r8-Feb-21" "1-Feb-21 and\r10-Feb-21" "1-Feb-21" ...
## $ Section of Law : chr [1:308] "Penal Code S: 505\r(b), Natural\rDisaster\rManagement law\rS: 25" "Anti-Corruption\rLaw
S: 55, Penal\rCode S: 505 (b)" "Penal Code S: 505\r(b)" "Penal Code S: 505\r(b)" ...
## $ Plaintiff : chr [1:308] NA NA "Deputy\rTownship\rAdministrat\ror Kyaw\rThein" NA ...
## $ Current Condition: chr [1:308] "Sentenced to 2\ryears" "Sentenced to 2\ryears and 75 Years" "Sentenced to 2\ryears with
hard\rlabour" "Sentenced to 2\ryears" ...
## $ Prison : chr [1:308] NA "Hpa-An Prison" NA NA ...
## $ Address : chr [1:308] "Magwe" "Kayin" "Rakhine" "Kayin" ...
## $ Region/Sta te : chr [1:308] "Magwe" "Kayin" "Kayin" "Kayin" ...
## $ Remark : chr [1:308] "Myanmar Military Seizes Power and\rSenior NLD leaders including Daw Aung\rSan Suu Kyi and
President U Win Myint"| __truncated__ "Myanmar Military Seizes Power and\rSenior NLD leaders including Daw Aung\rSan Suu Kyi an
d President U Win Myint"| __truncated__ "Myanmar Military Seizes Power and\rSenior NLD leaders including Daw Aung\rSan Suu Kyi
and President U Win Myint"| __truncated__ "Myanmar Military Seizes Power and\rSenior NLD leaders including Daw Aung\rSan Suu Ky
i and President U Win Myint"| __truncated__ ...
## - attr(*, "spec")=
## .. cols(
## .. No = col_double(),
## .. Name = col_character(),
## .. `Sex /Age` = col_character(),
## .. Age = col_character(),
## .. `Father's Name` = col_character(),
## .. Status = col_character(),
## .. `Date of Arrest` = col_character(),
## .. `Section of Law` = col_character(),
## .. Plaintiff = col_character(),
## .. `Current Condition` = col_character(),
## .. Prison = col_character(),
## .. Address = col_character(),
## .. `Region/Sta te` = col_character(),
## .. Remark = col_character()
## .. )
## - attr(*, "problems")=<externalptr>
```

Code

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
No	307	1	153.5	88.5	1.0	153.5	306.0

Code

```
## # A tibble: 5 × 14
##   No Name `Sex /Age` Age `Father's Name` Status `Date of Arrest`
##   <dbl> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 1 "Dr. A... M 62 U Nyo "MP (Regional... "1-Feb-21"
## 2 2 "Nan K... F <NA> U Saw Hla Tun "MP (State Hl... "1-Feb-21 and\r...
## 3 3 "Nyi P... M <NA> <NA> "MP (State Hl... "1-Feb-21 and\r...
## 4 4 "Bo Bo... M <NA> <NA> "MP (State Hl... "1-Feb-21"
## 5 5 "Min K... M <NA> <NA> "MP (State Hl... "1-Feb-21"
## # ... with 7 more variables: Section of Law <chr>, Plaintiff <chr>,
## # Current Condition <chr>, Prison <chr>, Address <chr>, Region/Sta te <chr>,
## # Remark <chr>
```

Further Research: table and figure visualizations

We would like to research the relationship between some sector indicators, the conflicts numbers and the number of detainees in the township level.

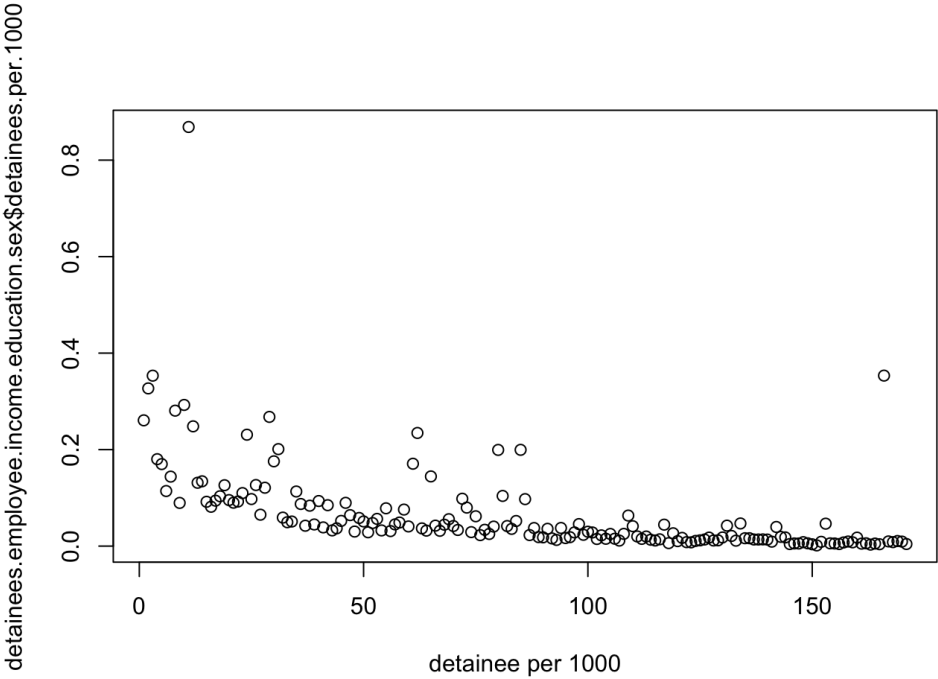
First, we plot the dependent variable for analysis.

Table and visualizations on dependent variable

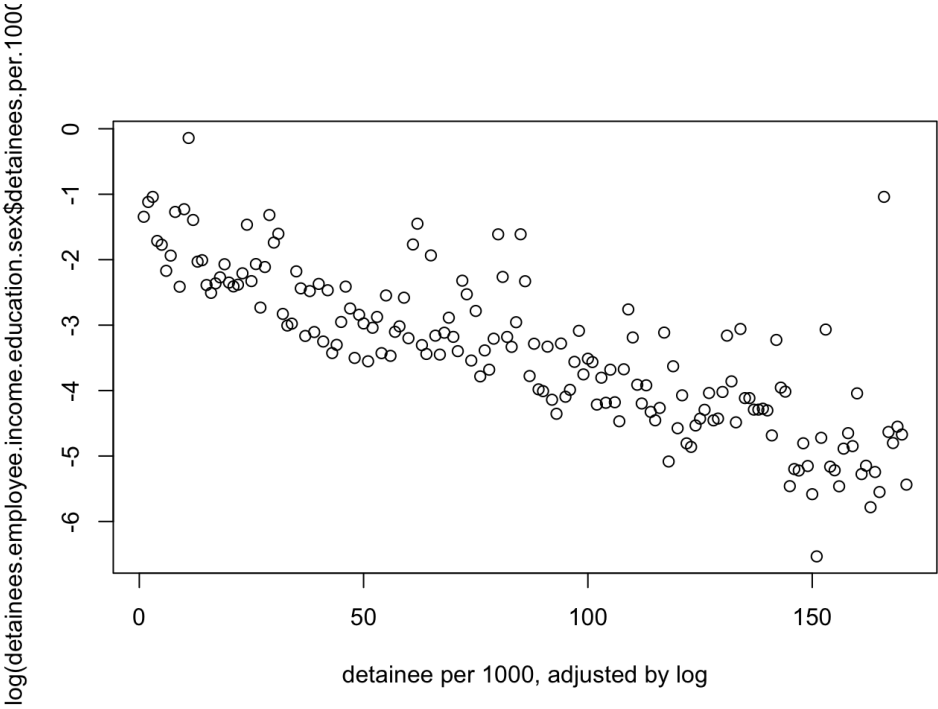
Code

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.001454	0.013967	0.035709	0.065479	0.084287	0.868666

Code



Code



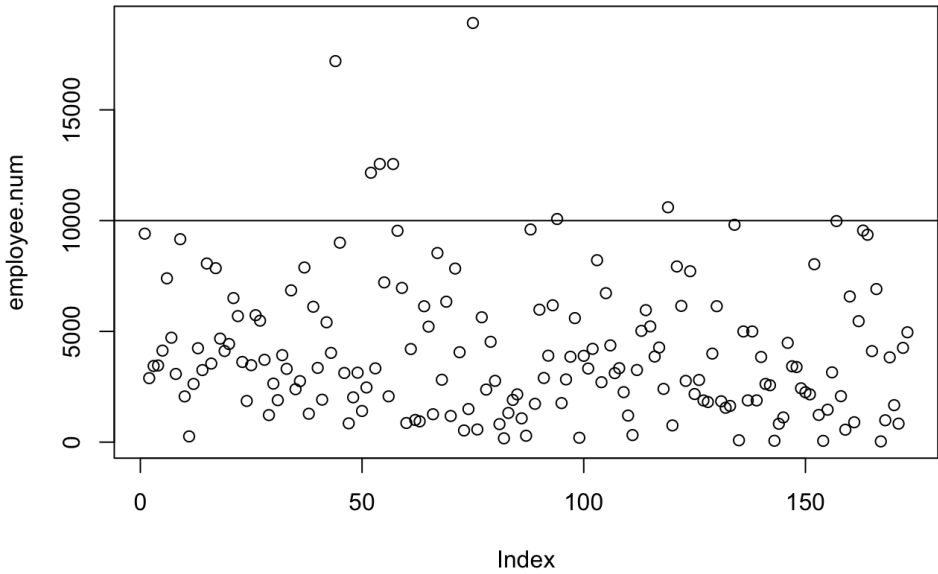
The detainee per 1000 is mainly concentrated below 1, the absolute number is small, and the gap between towns is more obvious after taking the log, showing a certain trend, which needs to be dismantled by the regression analysis afterwards.

Table and visualizations on univariate

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	33	1884	3351	4111	5637	18922

Code

Code



```
## integer(0)
```

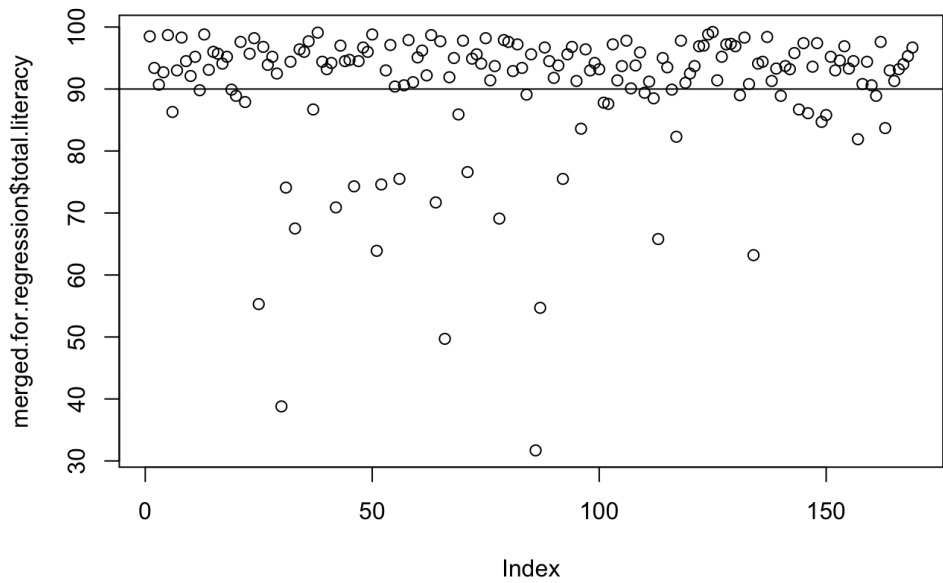
Code

The scatter plot shows that the majority of towns still employ less than 10,000 people and have a relatively limited level of economic development. The percent of having employee number > 10,000 is 4.05%

Code

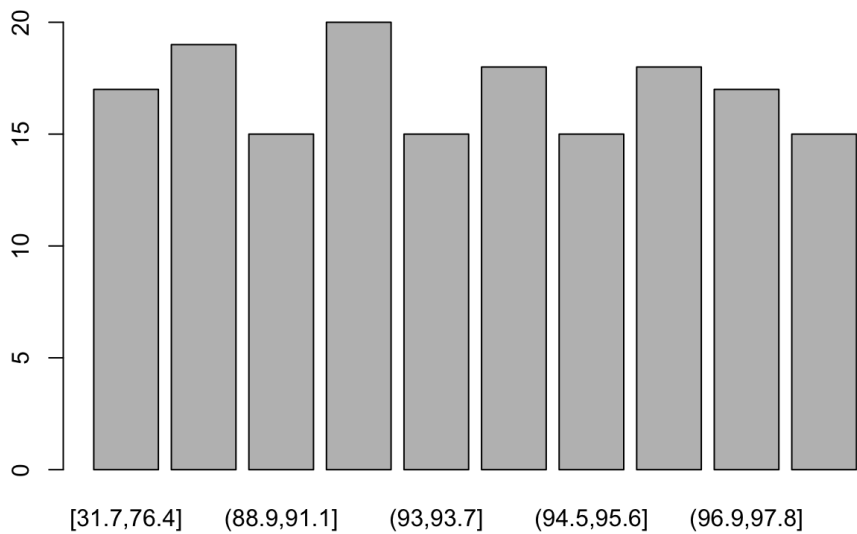
```
## # A tibble: 10 × 2
##   literacy_bin    n
##   <fct>         <int>
## 1 [31.7, 76.4]    17
## 2 (76.4, 88.9]   19
## 3 (88.9, 91.1]   15
## 4 (91.1, 93]     20
## 5 (93, 93.7]     15
## 6 (93.7, 94.5]   18
## 7 (94.5, 95.6]   15
## 8 (95.6, 96.9]   18
## 9 (96.9, 97.8]   17
## 10 (97.8, 99.2]  15
```

Code



```
## integer(0)
```

Code



High literacy rate overall.The percent of having literacy rate > 90% is 75.15%

Code

```
## # A tibble: 2 × 2
##   sex_cate    mean_detainees
##   <chr>      <dbl>
## 1 More Female 0.0675
## 2 More Male  0.0549
```

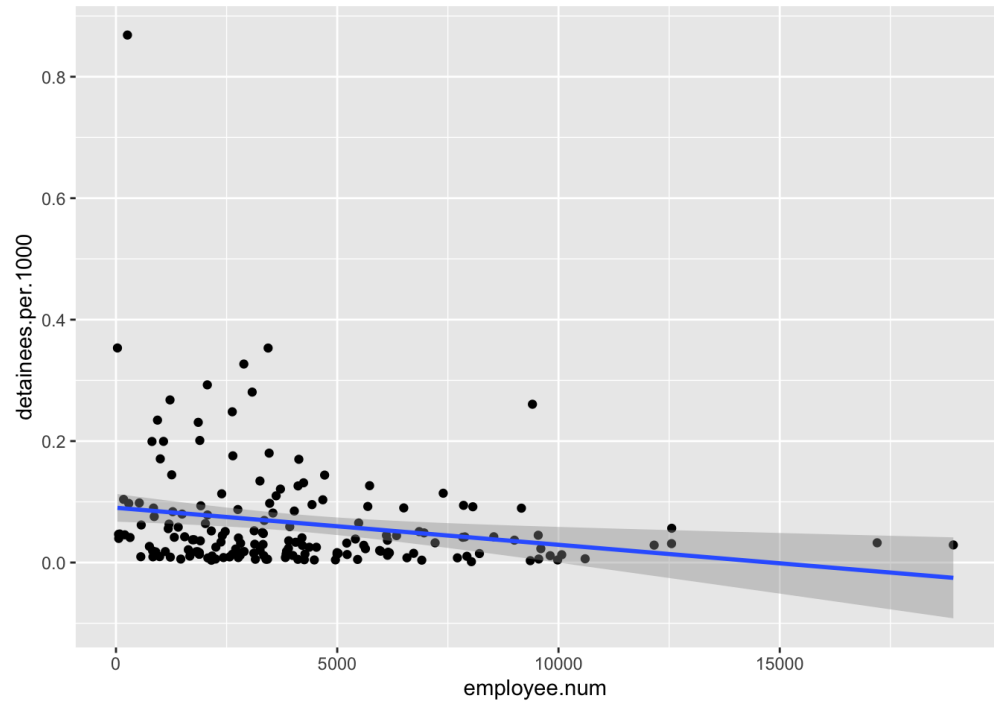
On average, the township with more women, which is defined as sex.ratio > 100 will turns to have 0.0125821 more detainees.ratio.1000, it may comes from woman turns to support Aung San Suu Kyi more in political campaigns.`

Next, we plot to learn the relationship between an important variable `employee.num` and `detainees.per.1000`

Table and visualizations on bivariate - employment and detainees.

Code

```
## `geom_smooth()` using formula 'y ~ x'
```



The correlation between the number of employees and the number of detainees.per.1000 is low, the corr coefficient is 0.0919718, overall the more employees in the area, the lower the number of arrests per 1,000 people.

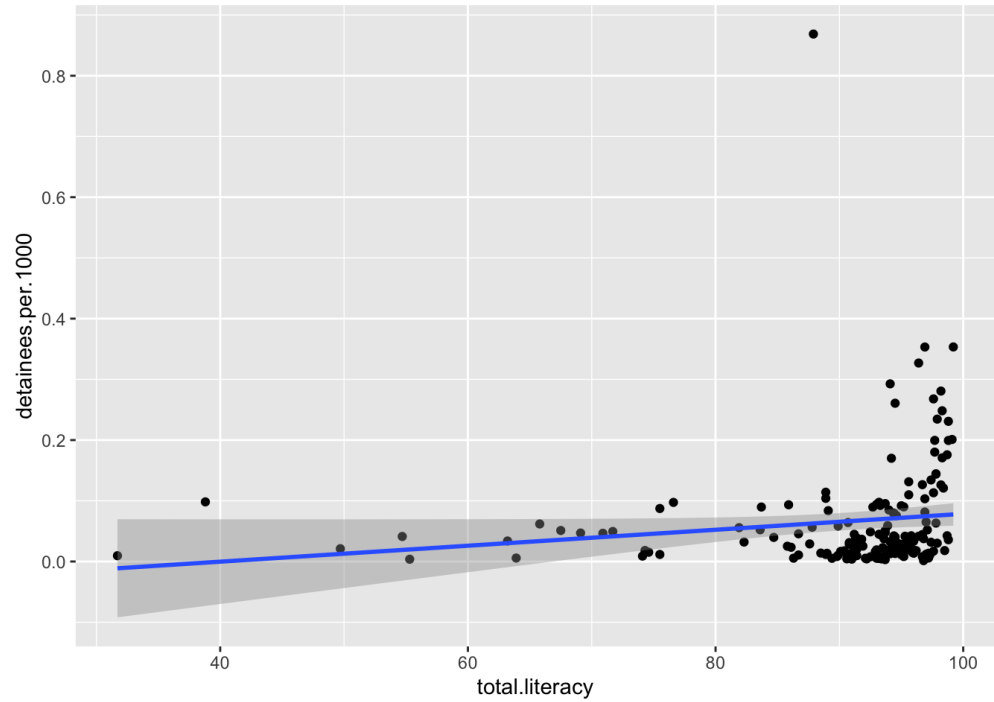
Table and visualizations on bivariate - education and detainees.

Code

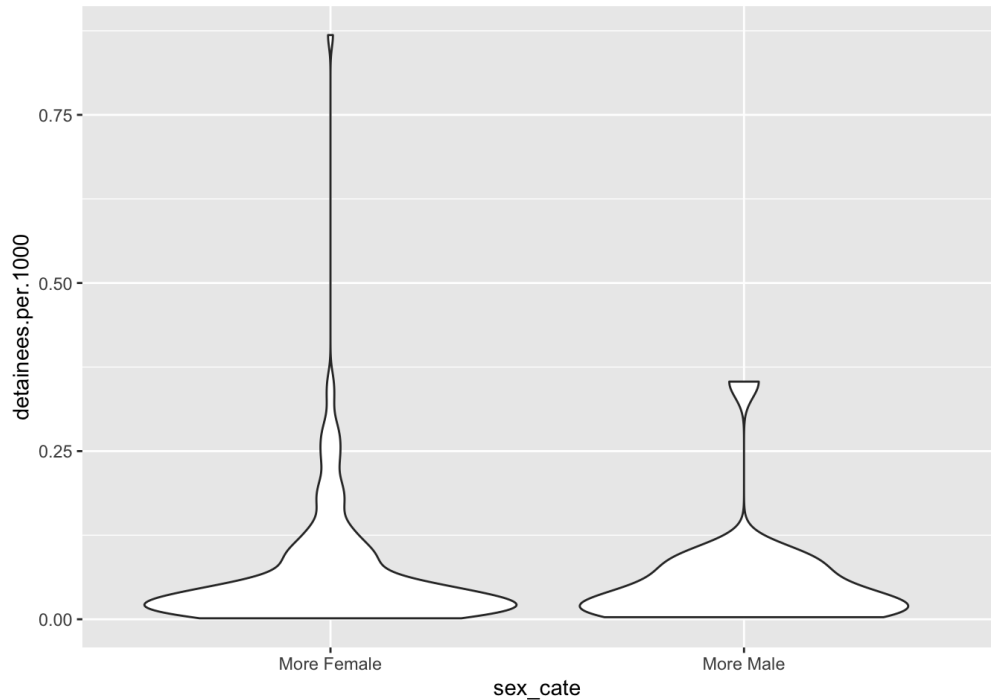
```
## # A tibble: 1 × 1
##   corr
##   <dbl>
## 1 0.181
```

Code

```
## `geom_smooth()` using formula 'y ~ x'
```



Code

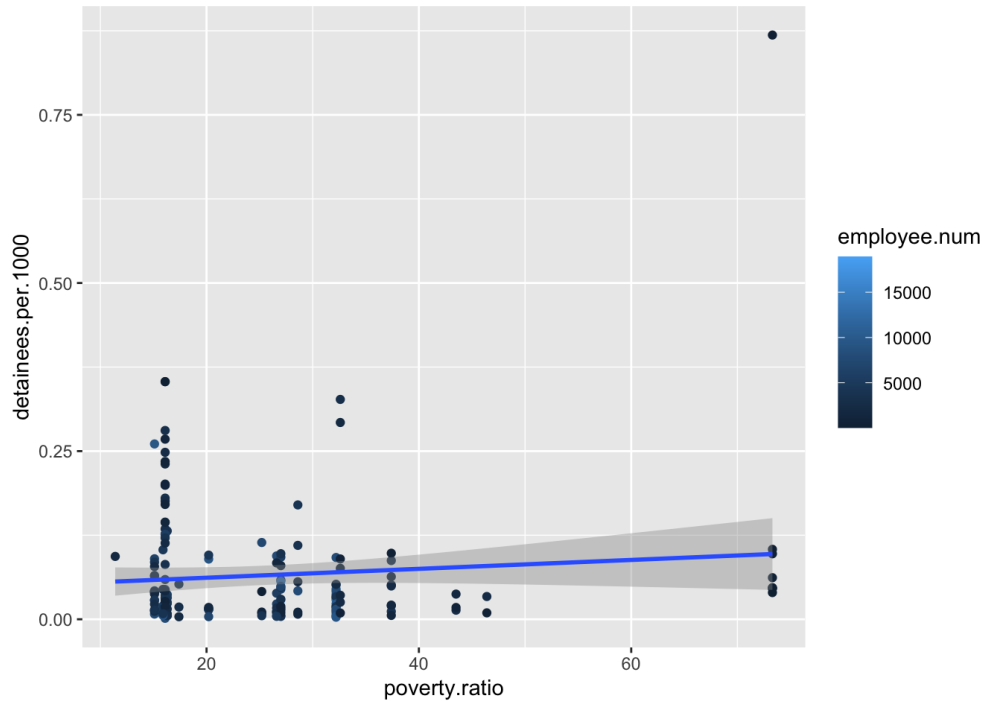


For the more female areas, the vast majority of towns have lower numbers of arrests per 1,000 and lower concentrations overall, but also receive extreme values.

Table and visualizations on trivariate and multivariate

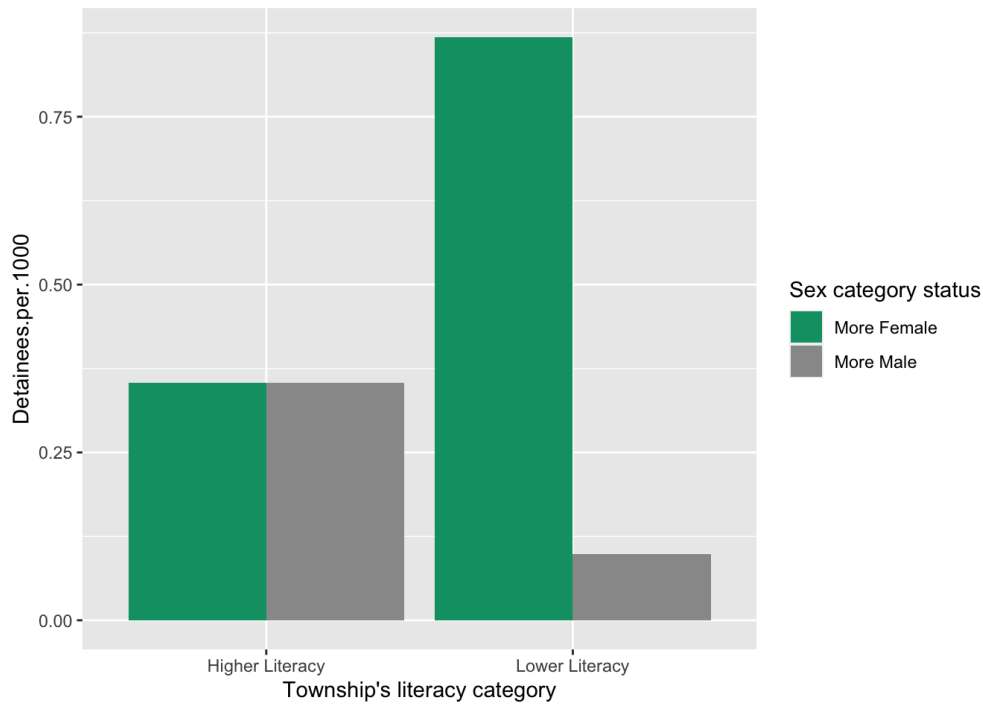
Code

```
## `geom_smooth()` using formula 'y ~ x'
```



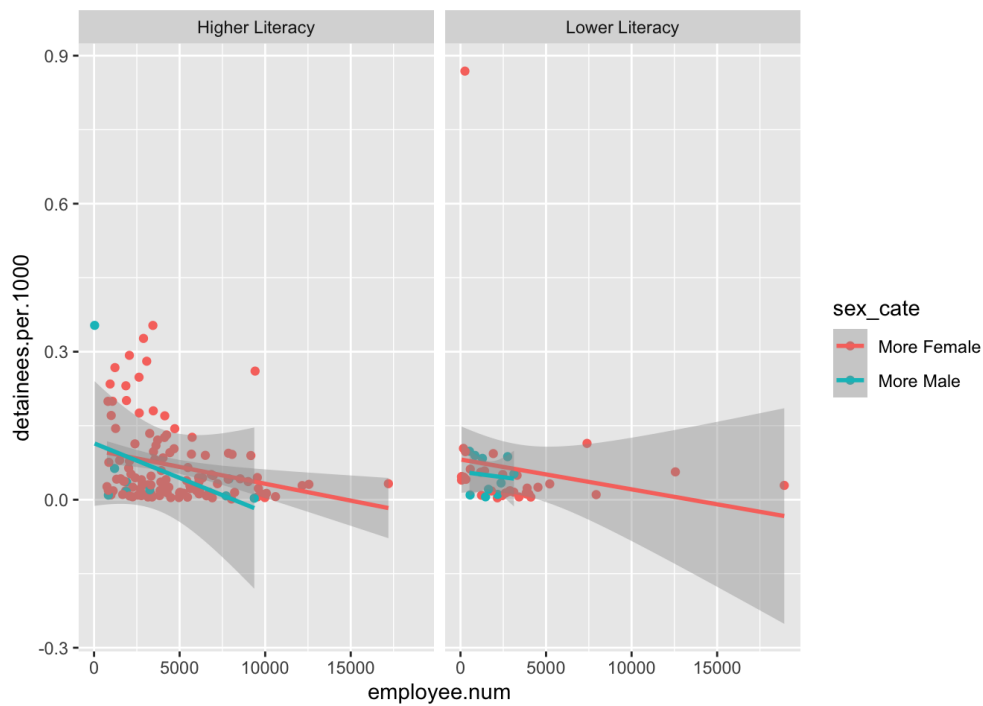
From the figure above, we found the positive correlation between the poverty ratio and the number of detainees.

Code



Code

```
## `geom_smooth()` using formula 'y ~ x'
```



The difference in detainees.per.1000 between the high and low literacy groups is brought about by the gap between the different gender groups. It is similarly influenced by the extreme values. Therefore, in the subsequent analysis, we will use detainees per 1000 after taking logs to eliminate the effect of extreme values.

A final figure summarizes the results in the multivariate descriptive analysis, where the difference in detainee.per.1000 between literacy rates was not significant, but the detainees.per.1000 were more dispersed in the high literacy rate group. An overall negative trend was observed.

Table on fallen data: bivariate and trivariate

Code


```
## # A tibble: 151 × 2
##   `Region/State`      `n()`
##   <chr>              <int>
## 1 ,B Sahgwoegyin Town      2
## 2 ,K Wacahiningmaw To      3
## 3 ,M Paguwke Township      2
## 4 ,S Kagaalainy gTownshi    1
## 5 1M) aWgwared, Taungd      1
## 6 A Vyiellyaagrew, aTdhdya ou 1
## 7 aAuybeiyina Trrowadndsyh ip 1
## 8 Aipyeyarwaddy           5
## 9 ALyaebyuatrtwa aTdodwyn sh 1
## 10 aMdaynadra Tlaoywnship     2
## # ... with 141 more rows
```

We would like to research the relation between the number of people detained per town and the number of people fallen per town.

- First, summarize the number of detainees per town.

Code

```
## # A tibble: 5 × 2
## # Groups:   Region/State [5]
##   `Region/State`      detained.num
##   <chr>              <int>
## 1 ,B Sahgwoegyin Town      2
## 2 ,K Wacahiningmaw To      3
## 3 ,M Paguwke Township      2
## 4 ,S Kagaalainy gTownshi    1
## 5 1M) aWgwared, Taungd      1
```

- Second, summarize the number of fallens per town.

Code

```
## # A tibble: 5 × 2
## # Groups:   States/Regions [5]
##   `States/Regions` fallen.num
##   <chr>              <int>
## 1 Ayeyarwady          15
## 2 Bago                 50
## 3 Chin                 16
## 4 Kachin                17
## 5 Kayah                 15
```

We would like to briefly analyze the detainee and the fallen dataset.

Table on fallen data: bivariate

Code

Sex	Mean
F	34.52
LGBT	37.00
M	31.61

Table on fallen data: trivariate

Code

	F	LGBT	M
States/Regions	Mean	Mean	Mean
Ayeyarwady			27.47
Bago	38.00		31.44
Chin	33.50		31.36
Kachin	40.00		35.57
Kayah	15.00		30.77

	F	LGBT	M
States/Regions	Mean	Mean	Mean
Magway	52.20		36.23
Mandalay	26.60	37.00	31.29
Mon	19.00		28.09
Naypyitaw	19.00		38.20
Sagaing	30.50		32.35
Shan	68.33		26.77
Tanintharyi	31.00		30.69
Yangon	40.40		30.64

Statistical Test

Code

```
##
##  Welch Two Sample t-test
##
## data:  Age by Sex
## t = 1.1558, df = 64.761, p-value = 0.252
## alternative hypothesis: true difference in means between group F and group M is not equal to 0
## 95 percent confidence interval:
## -2.118779  7.939498
## sample estimates:
## mean in group F mean in group M
##      34.51613      31.60577
```

Code

```
## [1] 0.2519998
```

According to the t test above, the p value is 0.2519998, which is much larger than 0.05, so we could conclude that the age of fallen people do not have signifiant difference within male and female.

Data Summarizations on Merged Data

As aboved steps show, we finally select some matrixes from both MIMU and ACLED dataset. Now, based on the data, we want to do some summarization and visualization.

Code

```
## # A tibble: 5 × 14
##   MIMU_township State_Region detainees.per.1000 detainees employee.num
##   <chr>         <chr>          <dbl>      <dbl>      <dbl>
## 1 Ahlone       Yangon             0.0180        1        1115
## 2 Aunglan      Magway             0.00425       1        4481
## 3 Ayadaw       Sagaing            0.0642       10        2021
## 4 Bago         Bago (East)        0.0895       44        9162
## 5 Bahan        Yangon             0.176        17        2641
## # ... with 9 more variables: poverty.ratio <dbl>, food.poverty.index <dbl>,
## #   poverty.gap.ratio <dbl>, fem.literacy <dbl>, male.literacy <dbl>,
## #   total.literacy <dbl>, sex.ratio <dbl>, conflicts.num <int>,
## #   literacy_bin <fct>
```

Code



```
## tibble [169 × 14] (S3: tbl_df/tbl/data.frame)
## $ MIMU_township      : chr [1:169] "Ahlone" "Aunglan" "Ayadaw" "Bago" ...
## $ State_Region      : chr [1:169] "Yangon" "Magway" "Sagaing" "Bago (East)" ...
## $ detainees.per.1000: num [1:169] 0.01802 0.00425 0.0642 0.08953 0.17575 ...
## $ detainees         : num [1:169] 1 1 10 44 17 1 2 7 5 1 ...
## $ employee.num      : num [1:169] 1115 4481 2021 9162 2641 ...
## $ poverty.ratio     : num [1:169] 16.1 27 15.1 20.2 16.1 16.3 32.2 16.1 15.1 27 ...
## $ food.poverty.index: num [1:169] 2.4 3.6 1.3 2.8 2.4 3.6 6.1 2.4 1.3 3.6 ...
## $ poverty.gap.ratio : num [1:169] 0.02 0.04 0.02 0.03 0.02 0.03 0.05 0.02 0.02 0.04 ...
## $ fem.literacy      : num [1:169] 97.9 90.5 86.7 90.3 98.1 83.5 90.4 97.6 92.1 88.6 ...
## $ male.literacy     : num [1:169] 99.2 96.9 95.9 95.6 99.4 89.8 95.8 99.2 97.9 96.9 ...
## $ total.literacy    : num [1:169] 98.5 93.4 90.7 92.7 98.7 86.3 93 98.3 94.5 92.1 ...
## $ sex.ratio        : num [1:169] 85.4 90.5 83.2 92 88.9 94.4 97.5 90 80.5 80.1 ...
## $ conflicts.num     : int [1:169] 22 20 42 97 51 36 11 17 50 37 ...
## $ literacy_bin      : Factor w/ 10 levels " [31.7,76.4]",...: 10 5 3 4 10 2 4 10 6 4 ...
## - attr(*, "na.action")= 'omit' Named int [1:2] 40 172
## ..- attr(*, "names")= chr [1:2] "40" "172"
```

Code

```
## MIMU_township      State_Region      detainees.per.1000  detainees
## Length:169        Length:169        Min.      :0.001454  Min.      : 1.00
## Class :character   Class :character   1st Qu.:0.014046  1st Qu.: 2.00
## Mode  :character   Mode  :character   Median :0.035833  Median : 5.00
##                                     Mean  :0.066076  Mean  :10.76
##                                     3rd Qu.:0.084818  3rd Qu.:11.00
##                                     Max.   :0.868666  Max.   :97.00
##
## employee.num      poverty.ratio      food.poverty.index poverty.gap.ratio
## Min.      : 33      Min.      :11.40      Min.      : 0.300      Min.      :0.01000
## 1st Qu.: 1881      1st Qu.:16.10      1st Qu.: 2.400      1st Qu.:0.02000
## Median : 3338      Median :20.20      Median : 3.600      Median :0.03000
## Mean   : 4123      Mean   :25.67      Mean   : 5.149      Mean   :0.04296
## 3rd Qu.: 5690      3rd Qu.:32.20      3rd Qu.: 6.100      3rd Qu.:0.05000
## Max.   :18922      Max.   :73.30      Max.   :25.000      Max.   :0.17000
##
## fem.literacy      male.literacy      total.literacy      sex.ratio
## Min.      :27.40      Min.      :35.80      Min.      :31.7      Min.      : 74.80
## 1st Qu.:86.50      1st Qu.:94.30      1st Qu.:90.1      1st Qu.: 88.00
## Median :91.50      Median :96.60      Median :93.7      Median : 91.50
## Mean   :87.89      Mean   :93.63      Mean   :90.5      Mean   : 92.49
## 3rd Qu.:94.70      3rd Qu.:98.10      3rd Qu.:96.2      3rd Qu.: 95.60
## Max.   :99.00      Max.   :99.50      Max.   :99.2      Max.   :174.20
##
## conflicts.num      literacy_bin
## Min.      : 2.00      (91.1,93] :20
## 1st Qu.: 15.00      (76.4,88.9]:19
## Median : 36.00      (93.7,94.5]:18
## Mean   : 54.32      (95.6,96.9]:18
## 3rd Qu.: 77.00      [31.7,76.4]:17
## Max.   :289.00      (96.9,97.8]:17
## NA's    :3          (Other)   :60
```

Code

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max		
detainees.per.1000	169	0	0.1	0.1	0.0	0.0	0.9		-
detainees	40	0	10.8	15.7	1.0	5.0	97.0		-
employee.num	168	0	4122.6	3205.7	33.0	3338.0	18922.0		-
poverty.ratio	17	0	25.7	13.4	11.4	20.2	73.3		-
food.poverty.index	16	0	5.1	5.2	0.3	3.6	25.0		-
poverty.gap.ratio	8	0	0.0	0.0	0.0	0.0	0.2		-
fem.literacy	113	0	87.9	12.1	27.4	91.5	99.0		
male.literacy	85	0	93.6	9.4	35.8	96.6	99.5		
total.literacy	103	0	90.5	10.7	31.7	93.7	99.2		












	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max		
sex.ratio	124	0	92.5	10.3	74.8	91.5	174.2		--
conflicts.num	99	2	54.3	56.2	2.0	36.0	289.0		

When we run regression, we don't want use missing data. So all the missing data would be dropped in the final model regression process.

Code

```
## # A tibble: 5 × 14
##   MIMU_township State_Region detainees.per.1000 detainees employee.num
##   <chr>          <chr>          <dbl>      <dbl>      <dbl>
## 1 Ahlone        Yangon            0.0180        1        1115
## 2 Aunglan       Magway            0.00425       1        4481
## 3 Ayadaw        Sagaing           0.0642       10        2021
## 4 Bago          Bago (East)       0.0895       44        9162
## 5 Bahan        Yangon            0.176        17        2641
## # ... with 9 more variables: poverty.ratio <dbl>, food.poverty.index <dbl>,
## #   poverty.gap.ratio <dbl>, fem.literacy <dbl>, male.literacy <dbl>,
## #   total.literacy <dbl>, sex.ratio <dbl>, conflicts.num <int>,
## #   literacy_bin <fct>
```

Code

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max		
detainees.per.1000	166	0	0.1	0.1	0.0	0.0	0.9		-
detainees	40	0	10.9	15.8	1.0	5.0	97.0		--
employee.num	165	0	4125.3	3204.5	58.0	3334.0	18922.0		--
poverty.ratio	17	0	25.8	13.5	11.4	22.7	73.3		-
food.poverty.index	16	0	5.2	5.3	0.3	3.6	25.0		-
poverty.gap.ratio	8	0	0.0	0.0	0.0	0.0	0.2		-
fem.literacy	111	0	87.8	12.2	27.4	91.5	98.7		
male.literacy	85	0	93.6	9.4	35.8	96.6	99.5		
total.literacy	102	0	90.4	10.7	31.7	93.7	99.1		
sex.ratio	123	0	92.2	9.4	74.8	91.5	174.2		-
conflicts.num	98	0	54.3	56.2	2.0	36.0	289.0		

MODELS AND ANALYSIS

Regression Model 1: regression on all variables

Linear Regression Model:

Now we run regression of detainees.per.1000 on all other variables, and analyze the results:

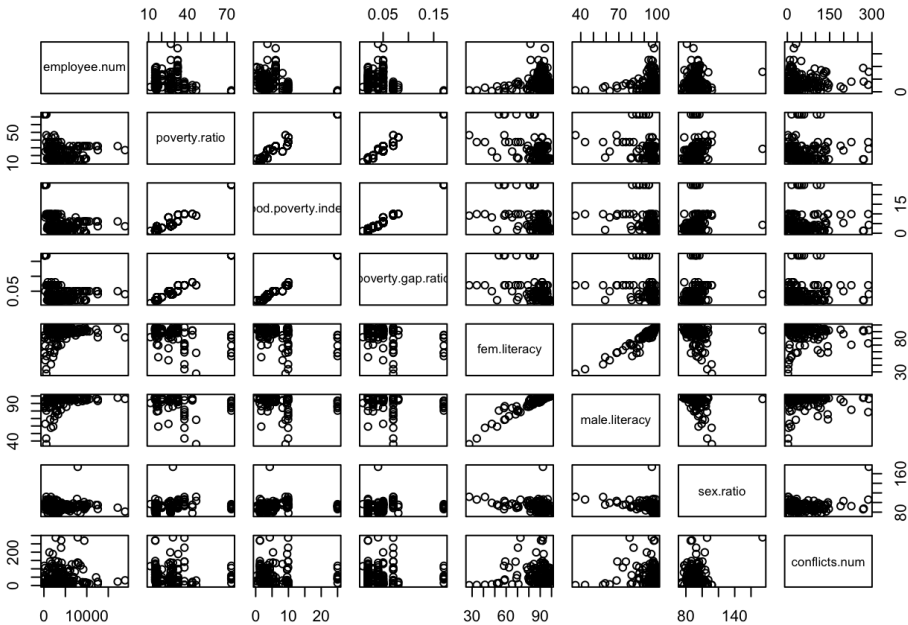
Code

```
##
## Call:
## lm(formula = detainees.per.1000 ~ employee.num + poverty.ratio +
##      food.poverty.index + poverty.gap.ratio + fem.literacy + male.literacy +
##      sex.ratio + conflicts.num, data = merged.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16139 -0.03934 -0.00950  0.02846  0.61527
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.291082057  0.125205440   2.325  0.021361 *
## employee.num   -0.000001942  0.000002195  -0.885  0.377658
## poverty.ratio  -0.000670352  0.002783850  -0.241  0.810024
## food.poverty.index 0.028733193  0.009188803   3.127  0.002105 **
## poverty.gap.ratio -3.016161561  2.103832360  -1.434  0.153660
## fem.literacy     0.008985560  0.001961178   4.582 0.00000935 ***
## male.literacy    -0.008778279  0.002477981  -3.543  0.000522 ***
## sex.ratio       -0.002336097  0.000740271  -3.156  0.001919 **
## conflicts.num     0.000520062  0.000112132   4.638 0.00000737 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07888 on 157 degrees of freedom
## Multiple R-squared:  0.3206, Adjusted R-squared:  0.286
## F-statistic: 9.261 on 8 and 157 DF,  p-value: 2.058e-10
```

Before we analyze the result, we first get into the correlation of the independent variables and take the assessment for collinearity.

Correlation analysis:

Code



From the correlation figures above, we can find that there are strong correlations between poverty.ratio , food.poverty.index and poverty.gap.ratio . In addition, fem.literacy and male.literacy have strong correlations. Therefore, we should delete some variables to fine tune the model.

Regression Model 2: regression on selected variables

Linear Regression Model:

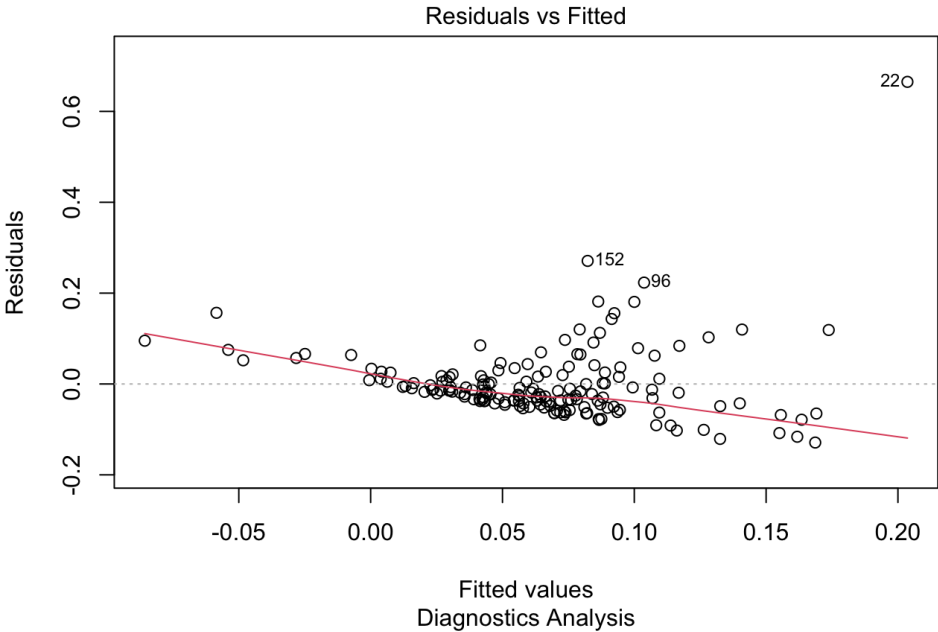
Code

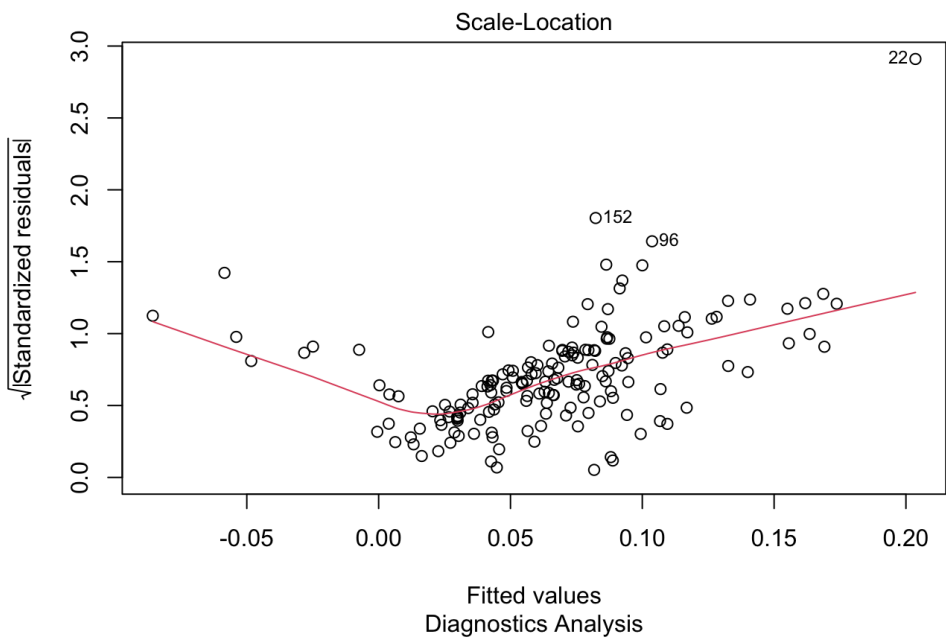
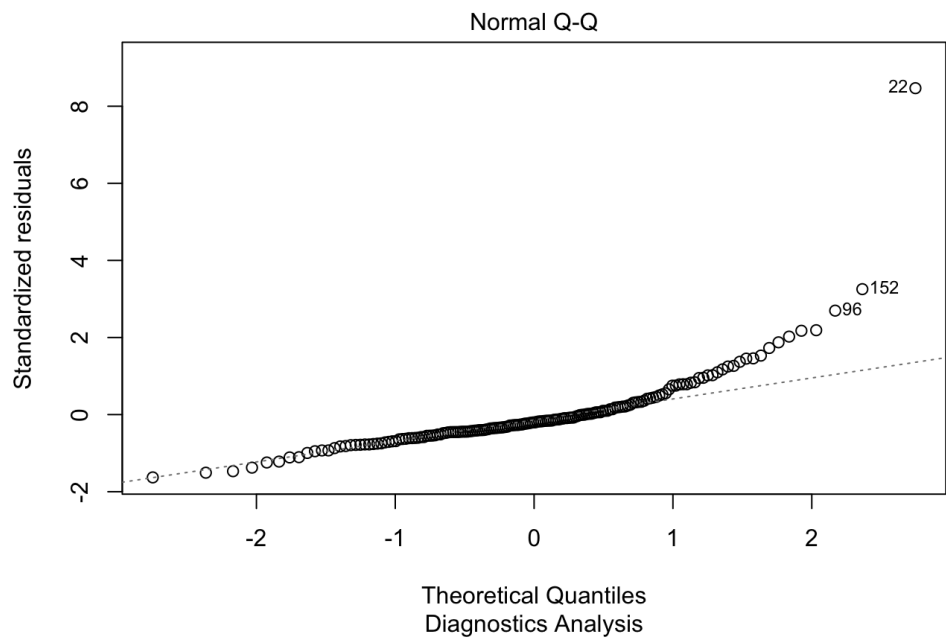
```
##
## Call:
## lm(formula = detainees.per.1000 ~ employee.num + food.poverty.index +
##     sex.ratio + fem.literacy + conflicts.num, data = merged.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.12893 -0.04202 -0.01660  0.01904  0.66506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.048733272  0.098331338  -0.496  0.620855
## employee.num  -0.000004970  0.000002139  -2.323  0.021439 *
## food.poverty.index  0.005003700  0.001448059   3.455  0.000704 ***
## sex.ratio     -0.001278334  0.000736266  -1.736  0.084446 .
## fem.literacy   0.002277046  0.000634194   3.590  0.000439 ***
## conflicts.num   0.000483128  0.000117541   4.110  0.000063 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08369 on 160 degrees of freedom
## Multiple R-squared:  0.2208, Adjusted R-squared:  0.1964
## F-statistic: 9.067 on 5 and 160 DF,  p-value: 0.0000001311
```

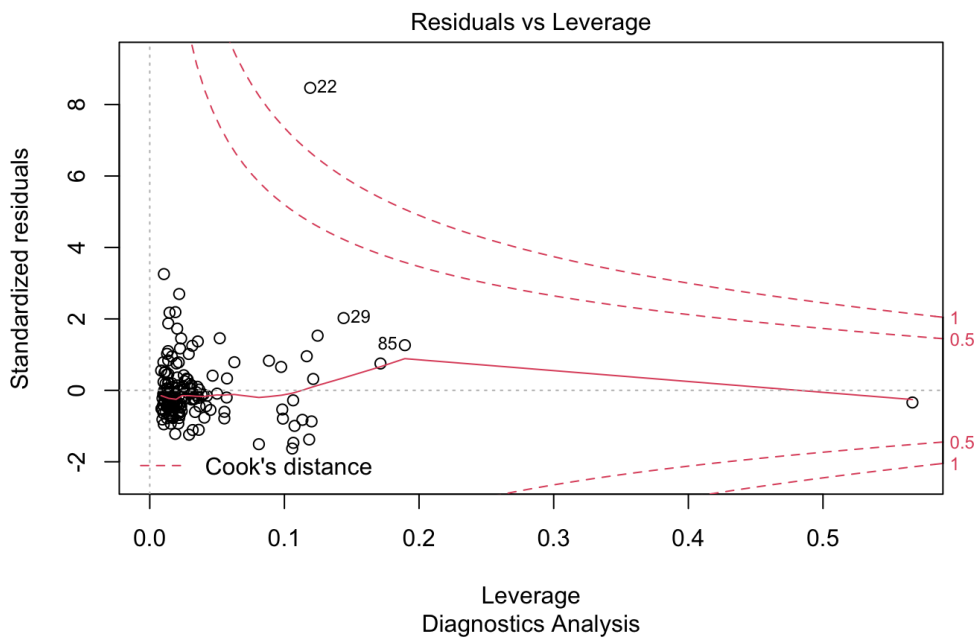
Regression diagnostics analysis

After getting the basic regression model, we would like to look at the diagnostic plots and access whether it is reasonable to use this model.

Code







Look at the four plots.

First of all, we expect **Residuals vs. Fitted** to have constant variance and consider residuals and fitted values uncorrelated. However, it seems like they are not uncorrelated. As for **Normal QQ plot**, it looks a little right-skewed, and the residuals from the regression are not normally distributed.

In **Scale-location plot** plot, there is no discernible trends.

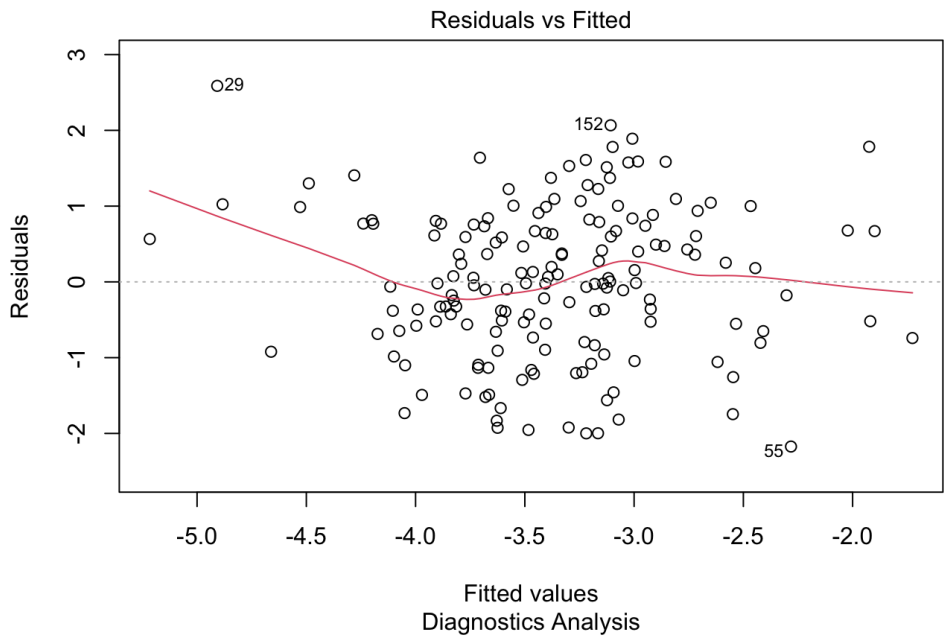
But in **Residuals vs Leverage** plot, it seems that there are some obvious outliers.

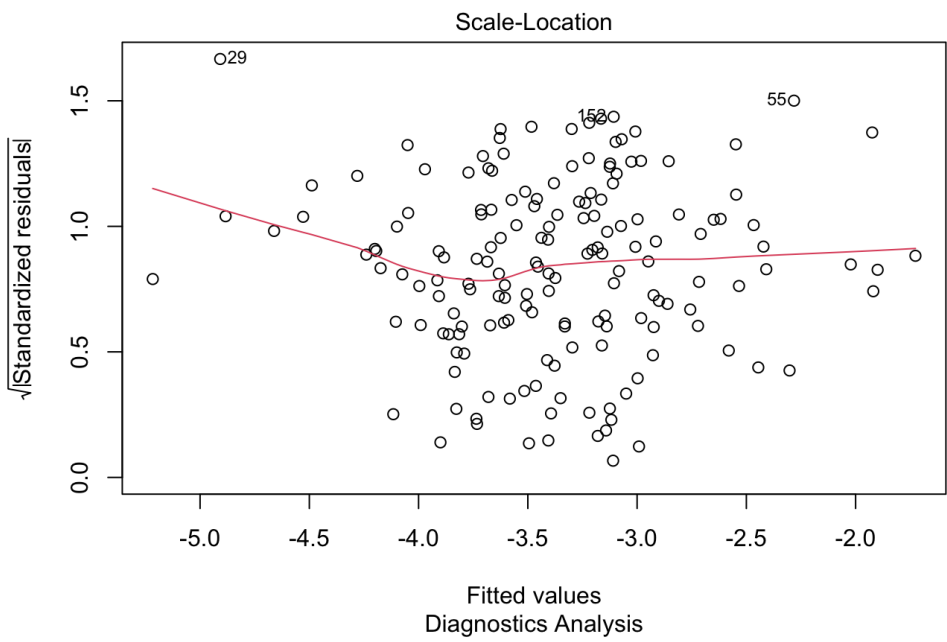
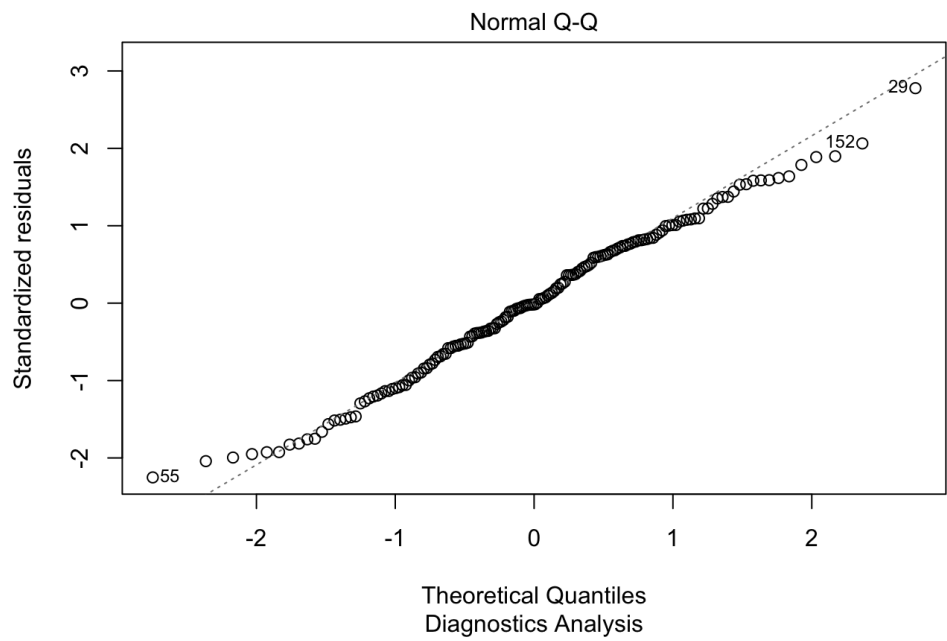
Based on the diagnostics analysis, we assume there is a better version for this regression model. Considering the QQ plot, we expect to do log calculation on the dependent variable and do linear regression on `log(detainees.per. 1000)` .

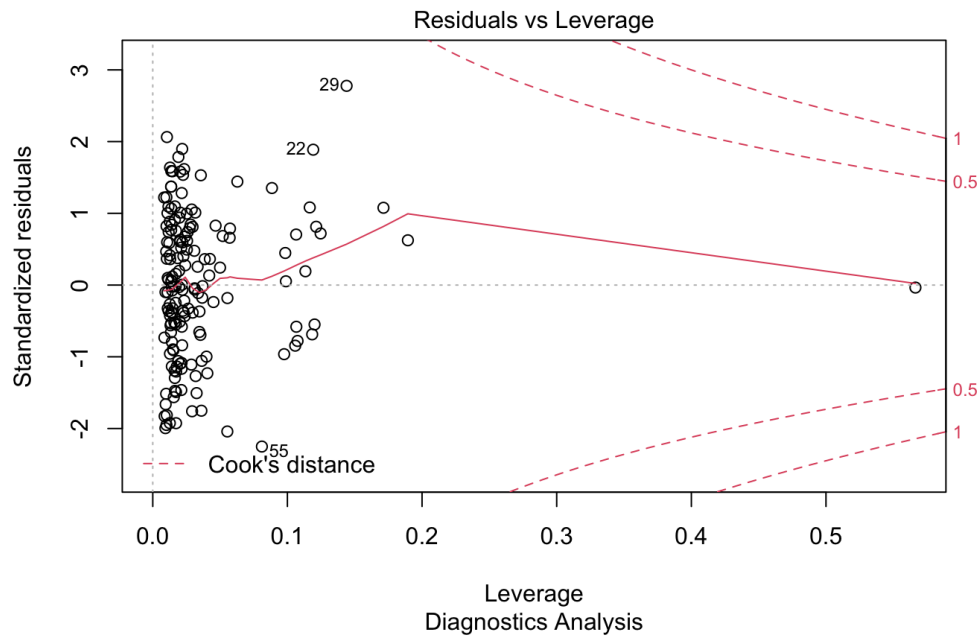
Regression Model 3: log calculation on dependent variable

Linear Regression Model:

Code







Code

```
##
## Call:
## lm(formula = log(detainees.per.1000) ~ employee.num + food.poverty.index +
##     fem.literacy + conflicts.num + sex.ratio, data = merged.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1730 -0.6564 -0.0167  0.7525  2.5871
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4.36376676  1.18293744  -3.689   0.000308 ***
## employee.num    -0.00006342  0.00002574  -2.464   0.014791 *
## food.poverty.index  0.04331169  0.01742032   2.486   0.013936 *
## fem.literacy     0.02491904  0.00762943   3.266   0.001334 **
## conflicts.num    0.00763076  0.00141403   5.396  0.000000241 ***
## sex.ratio      -0.01706271  0.00885737  -1.926   0.055829 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.007 on 160 degrees of freedom
## Multiple R-squared:  0.2471, Adjusted R-squared:  0.2236
## F-statistic: 10.5 on 5 and 160 DF, p-value: 9.81e-09
```

Looking at the p-value of sex.ratio, its p-value is 0.05583 and it is greater than 0.05. So it would be better if we dropped this column.

Regression Model 4: use selected variables to regress on log(dependent variable)

Linear Regression Model:

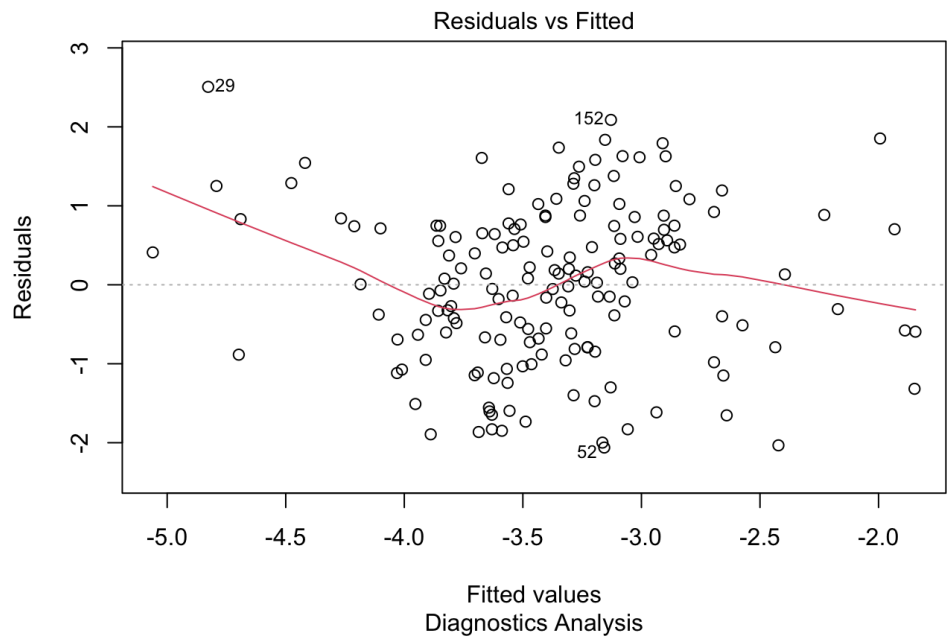
Now drop the variable and rerun the regression.

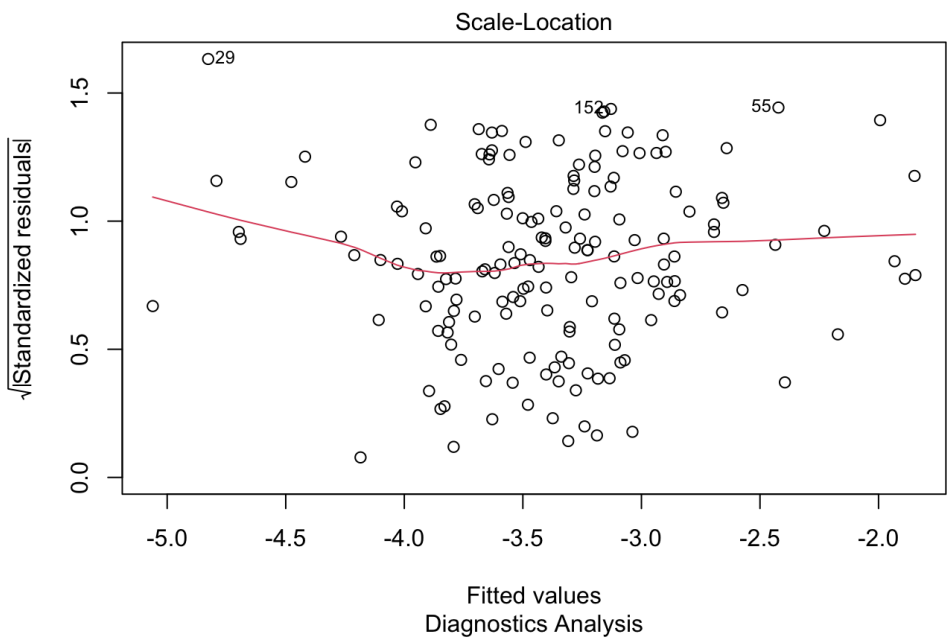
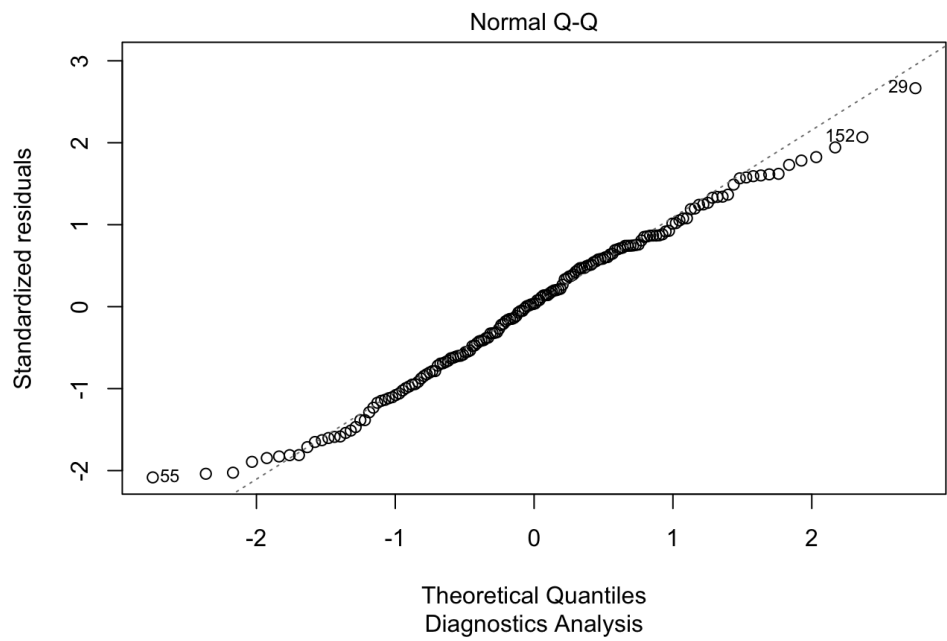
Code

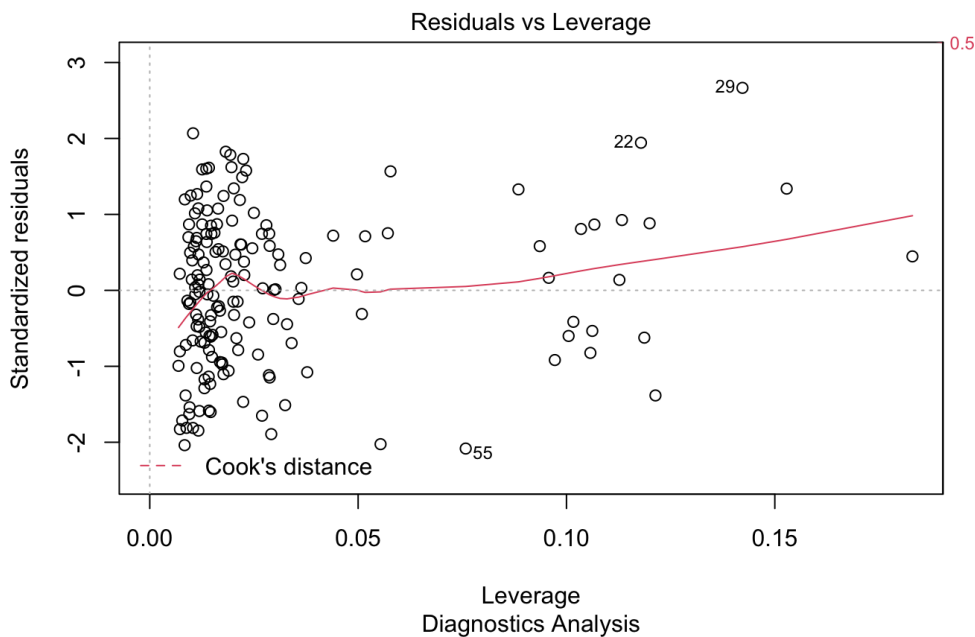
```
##
## Call:
## lm(formula = log(detainees.per.1000) ~ employee.num + food.poverty.index +
##     fem.literacy + conflicts.num, data = merged.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06098 -0.69550  0.03575  0.74387  2.50645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.19601109  0.70924030  -8.736 2.98e-15 ***
## employee.num   -0.00006797  0.00002584  -2.630 0.009365 **
## food.poverty.index  0.04163700  0.01754448   2.373 0.018813 *
## fem.literacy    0.02847186  0.00746523   3.814 0.000195 ***
## conflicts.num    0.00717594  0.00140587   5.104 9.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.015 on 161 degrees of freedom
## Multiple R-squared:  0.2297, Adjusted R-squared:  0.2105
## F-statistic: 12 on 4 and 161 DF,  p-value: 0.00000001468
```

Regression diagnostics analysis

Code







Now look at the four plots again.

Residuals vs. Fitted: It is obvious that residuals and fitted are unrelated with each other.

Normal QQ plot, The residuals from the regression are almost normally distributed. It is more normal distributed than the previous model.

Scale-location plot, there is no discernible trends.

Residuals vs Leverage , it seems that there are no apparent outliers.

Therefore, based on the diagnostics analysis, we consider it is reasonable to use this model. Our final model is to run linear regression of `log(detainees.per.1000)` on `employee.num`, `food.poverty.index`, `fem.literacy` and `conflicts.num`.

RESULTS

[Code](#)

```
##
## Call:
## lm(formula = log(detainees.per.1000) ~ employee.num + food.poverty.index +
##     fem.literacy + conflicts.num, data = merged.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06098 -0.69550  0.03575  0.74387  2.50645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.19601109   0.70924030  -8.736 2.98e-15 ***
## employee.num   -0.00006797   0.00002584  -2.630 0.009365 **
## food.poverty.index  0.04163700   0.01754448   2.373 0.018813 *
## fem.literacy    0.02847186   0.00746523   3.814 0.000195 ***
## conflicts.num    0.00717594   0.00140587   5.104 9.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.015 on 161 degrees of freedom
## Multiple R-squared:  0.2297, Adjusted R-squared:  0.2105
## F-statistic: 12 on 4 and 161 DF, p-value: 0.0000001468
```

[Code](#)

From the summary report, we can see that based on the confidence level of 95%, `employee.num`, `food.poverty.index`, `fem.literacy` and `conflicts.num` are all statistically significant predictors of `log(detainees.per.1000)`. The overall p-value is significantly less than 0.05. The p-value of `employee.num` is 0.00936, the p-value of `food.poverty.index` is 0.01881, the p-value of `fem.literacy` is 0.00019, the p-value of `conflicts.num` is 0. They are all smaller than 0.05.

When all else being equal between two townships, a 1 employment increase in average employee number appears to be associated with a -0.00007 decrease in `log(detainment rates per thousand)`. When all else being equal between two townships, a 1 food poverty increase in average appears to be associated with a 0.04164 increase in `log(detainment rates per thousand)`. When all else being equal between two townships, a 1

fem.literacy increase in average appears to be associated with a 0.02847 increase in log(detainment rates per thousand). When all else being equal between two townships, a 1 conflict number increase in average appears to be associated with a 0.00718 increase in log(detainment rates per thousand).

We can write the model as:

$$\log(\text{detaineesPer1000}) = -6.196 - 0.00007 \times \text{employeeNum} + 0.042 \times \text{povertyIndex} + 0.028 \times \text{fem.literacy} + 0.007 \times \text{conflictNum}$$

DISCUSSIONS

Limitations of the Analysis

Although we have 5 data sets, several columns have too many missing values. For example, almost detainees and imprisoned don't have a valid age number. Therefore, we have to give up some columns that may have useful information.

In addition, the ACLED data set has only valid data that are based on 2021. However, the data from MIMU data set don't have data on 2021. They only have data before 2020. What's worse, it has so many missing columns that we have to use data from different years for analysis. For example, when selecting variable, we used the Employee Number, Female Literacy that are from 2014, Poverty data that are from 2010, but conflicts number data that are from 2021. This too large time span may lead to some problems.

Beside, when preprocessing data, we simply drop the missing value. It may have some other ways to make use of them like filling the missing value with average data. And we didn't pay too much attention on outliers, but it is quite reasonable because we will do log calculation on dependent variables. The outliers would disappear in that case.

About the dependent variables, there are also some limitations. The value of Female literacy as a percentage of total female population is almost all greater than 90% and the variance is quite small. Although it is related to the detainment rates per thousand, it is possible that they don't have cause and effect relationship.

Inferences from Results

From the model, we can see that the detainment rates per thousand is positively influenced by female literacy percentage, poverty index and conflict number. It is negatively influenced by employee number.

Therefore, the higher the employment number is, the less the detainees percentage will be. The less the female literacy percentage is, the less the detainees percentage will be. The less the poverty index is, the less the detainees percentage will be. The less the conflict numbers, the less the detainees percentage will be.

Potential Stakeholders

We can connect some background about the detainee event with our analysis. As it is said in the introduction of final project, the background is that the Burmese military staged a coup, toppling the quasi-democratic government and removing Aung San Suu Kyi, the civilian leader supported by the National League of Democracy.

The potential stakeholders can be the Burmese military, Burmese government, or some other international organizations who want to stabilize the situation and make detainee number reduce.

Contextualization and Policy

To calm down such an event, it is important that reduce the local conflicts number. It would be wiser if some policies about bans on conflicts are announced. What's more, government can try to develop the local economy, to relief the poverty.

Conclusions

In summary, the detainment rates per thousand is positively related to poverty index, female literacy, conflict number and negatively related to employee number. The linear regression model can be shown like below.

$$\log(\text{detaineesPer1000}) = -6.196 - 0.00007 \times \text{employeeNum} + 0.042 \times \text{povertyIndex} + 0.028 \times \text{fem.literacy} + 0.007 \times \text{conflictNum}$$

Future Directions

For further analysis, we can gain more variables from these data sets instead of just 8 variables. When dealing with missing values, if there is only a few missing, we can use some method like average value to fill them. Also, we can pay more attention to the outliers in case of bad effects on the model.