

Relationship Between Crime Rate , Gross Domestic Product, Unemployment ,Gender,and Education in Taiwan

Christine Hsieh, May Huang, Yi-Fang Lo

7/22/2019

Objective:

There are many factors that lead to the fluctuations of number of crime, and as many people assert that education and economic circumstances are some factors that we are interested in. In the report, we analysed the relationship between crime, GDP, Education, and Unemployment to see if obvious relationship can be seen.

Install library needed

```
rm(list=ls())
library(readr)
library(tidyverse)
library(dplyr)
```

Data about Crimes and Education in Taiwan

First import the data about GDP in Taiwan

```
GDP <- read_csv("https://raw.githubusercontent.com/Vladislav292/2019summer-datasciece/master/GDP.csv", )
```

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   GDP = col_double(),
##   ir = col_double(),
##   di = col_double(),
##   disir = col_double()
## )
```

```
head(GDP)
```

```
## # A tibble: 6 x 5
##   year    GDP    ir    di disir
##   <dbl>  <dbl> <dbl>  <dbl> <dbl>
## 1 1992 5614679 11.8 5151732 9.84
## 2 1993 6205338 10.5 5659516 9.86
## 3 1994 6784442 9.33 6164888 8.93
## 4 1995 7396650 9.02 6642268 7.74
## 5 1996 8036590 8.65 7231774 8.88
## 6 1997 8717241 8.47 7843998 8.47
```

Import the data about Crime and Education and clean it

source Steps 1. read.csv 2. assing proper data type 3. filter the years we need 4. rename the columns

```

edu <- read_csv("https://raw.githubusercontent.com/ChristineHsieh1231/mii/master/Educational%20Attainment%20Data.csv",
  col_types = cols(X10 = col_number(),
    X11 = col_number(), X12 = col_number(),
    X13 = col_number(), X14 = col_number(),
    X15 = col_number(), X16 = col_number(),
    X17 = col_number(), X18 = col_number(),
    X19 = col_number(), X2 = col_number(),
    X20 = col_number(), X21 = col_number(),
    X22 = col_number(), X23 = col_number(),
    X24 = col_number(), X25 = col_number(),
    X26 = col_number(), X27 = col_number(),
    X28 = col_number(), X3 = col_number(),
    X4 = col_number(), X5 = col_number(),
    X6 = col_number(), X7 = col_number(),
    X8 = col_number(), X9 = col_number()))

## Warning: Missing column names filled in: 'X2' [2], 'X3' [3], 'X4' [4],
## 'X5' [5], 'X6' [6], 'X7' [7], 'X8' [8], 'X9' [9], 'X10' [10], 'X11' [11],
## 'X12' [12], 'X13' [13], 'X14' [14], 'X15' [15], 'X16' [16], 'X17' [17],
## 'X18' [18], 'X19' [19], 'X20' [20], 'X21' [21], 'X22' [22], 'X23' [23],
## 'X24' [24], 'X25' [25], 'X26' [26], 'X27' [27], 'X28' [28]

## Warning: 27 parsing failures.
## row col expected          actual
##   1 X2  a number Grand Total      'https://raw.githubusercontent.com/ChristineHsieh1231/mii/master
##   1 X5  a number Illiterate        'https://raw.githubusercontent.com/ChristineHsieh1231/mii/master
##   1 X8  a number -                  'https://raw.githubusercontent.com/ChristineHsieh1231/mii/master
##   1 X11 a number Elementary School 'https://raw.githubusercontent.com/ChristineHsieh1231/mii/master
##   1 X14 a number Junior High School 'https://raw.githubusercontent.com/ChristineHsieh1231/mii/master
## ... ..
## See problems(...) for more details.

edu = filter(edu, Year<2017)
edu = filter(edu, Year>1996)
edu = edu %>%
  rename(
    year = Year,
    t = X2,
    m = X3,
    f = X4,
    t0 = X5,
    m0 = X6,
    f0 = X7,
    s = X8,
    ms = X9,
    fs = X10,
    t1 = X11,
    m1 = X12,
    f1 = X13,
    t2 = X14,
    m2 = X15,
    f2 = X16,
    t3 = X17,
    m3 = X18,
    f3 = X19,

```

```

t4 = X20,
m4 = X21,
f4 = X22,
t5 = X23,
m5 = X24,
f5 = X25,
t8 = X26,
m8 = X27,
f8 = X28

```

```
)
```

Merge two sets of data for further analysis

Merge by parameter “year”

```

data_merged <-merge(edu, GDP,by="year")
head(data_merged)

```

```

##   year      t      m      f  t0  m0  f0  s  ms fs   t1   m1  f1
## 1 1997 172540 143164 29376 3476 1852 1624 229 162 67 29243 22860 6383
## 2 1998 158923 132101 26822 2863 1602 1261 186 140 46 24720 19395 5325
## 3 1999 179597 152553 27044 2678 1546 1132 181 122 59 25663 20882 4781
## 4 2000 181614 155891 25723 2487 1463 1024 143  96 47 24639 20491 4148
## 5 2001 180527 155226 25301 2481 1472 1009 172 138 34 24842 20870 3972
## 6 2002 185751 155666 30085 3217 1785 1432 180 127 53 25744 21012 4732
##      t2   m2   f2   t3   m3   f3   t4   m4   f4   t5   m5   f5   t8
## 1 79299 68285 11014 49430 40853  8577  8659  7343 1316 169 152 17 2035
## 2 73966 63746 10220 46321 38185  8136  8555  7089 1466 192 168 24 2120
## 3 81220 70989 10231 56365 47625  8740 10664  9019 1645 241 197 44 2585
## 4 80153 70832  9321 59447 50567  8880 12353 10431 1922 369 316 53 2023
## 5 75061 66369  8692 61793 52583  9210 14197 12077 2120 460 394 66 1521
## 6 70780 61337  9443 67179 55876 11303 16342 13653 2689 537 468 69 1772
##      m8  f8      GDP      ir      di disir
## 1 1657 378  8717241  8.47 7843998  8.47
## 2 1776 344  9381141  7.62 8345435  6.39
## 3 2173 412  9815595  4.63 8694931  4.19
## 4 1695 328 10351260  5.46 9188051  5.67
## 5 1323 198 10158209 -1.87 8923433 -2.88
## 6 1408 364 10680883  5.15 9417956  5.54

```

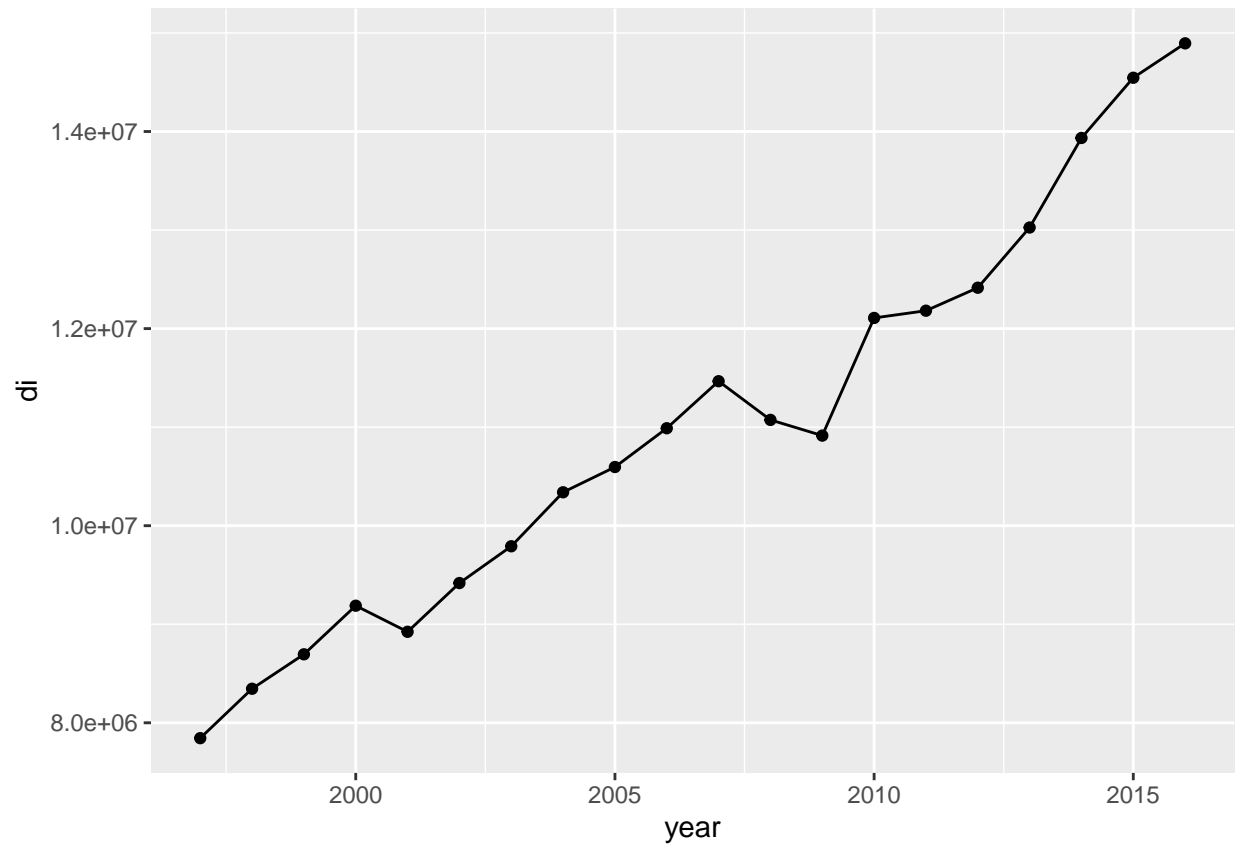
Plotting

Disposable income(a portion of GDP that can actually be used) - year unit di Million NTD

```

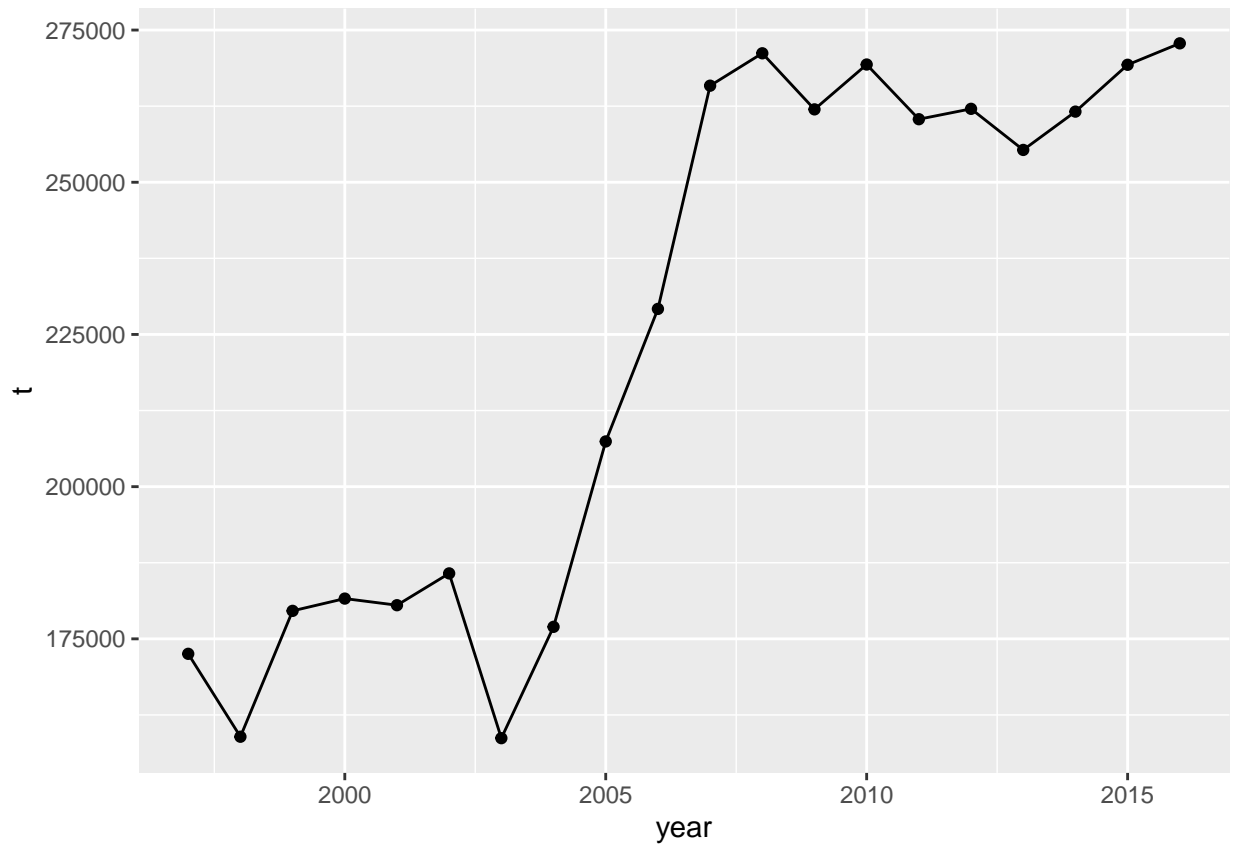
ggplot(data = data_merged)+
geom_line(mapping=aes(x=year,y=di,group=1))+geom_point(mapping = aes(x=year,y=di))

```



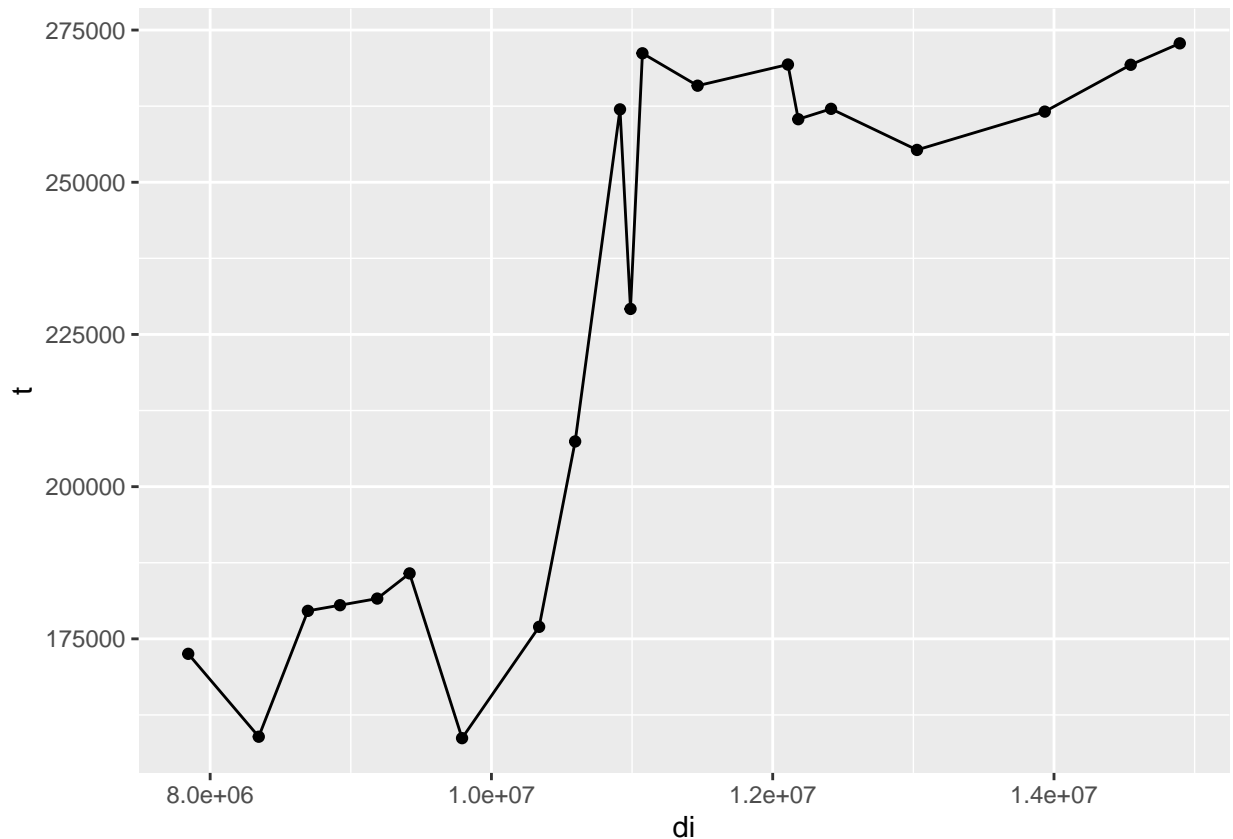
Crime-year

```
ggplot(data = data_merged)+  
geom_line(mapping=aes(x=year,y=t,group=1))+geom_point(mapping = aes(x=year,y=t))
```



Disposable income(a portion of GDP that can actually be used) - Crime unit di Million NTD Since no obvious relationship between the two factors can be seen, we moved to the other one-education.

```
ggplot(data = data_merged)+  
geom_line(mapping=aes(x=di,y=t,group=1))+geom_point(mapping = aes(x=di,y=t))
```



Make another dataframe to plot the following graph work summery: create a new column to enable ggplot to group by “class”

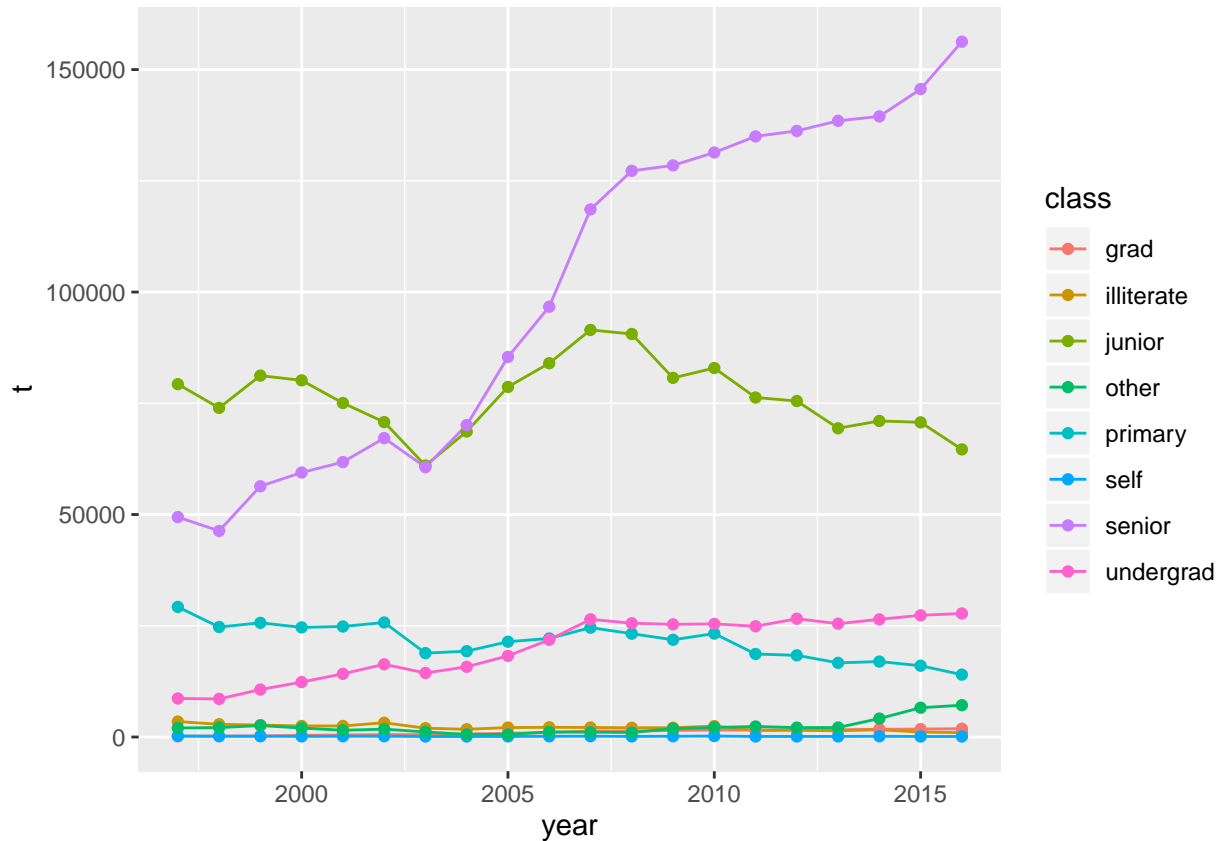
```
edu0 = edu %>% select(year,t0,m0,f0) %>%
  mutate(class = "illiterate") %>% rename(t=t0, m=m0,f=f0)
edu1 = edu %>% select(year,t1,m1,f1) %>%
  mutate(class = "primary")%>% rename(t=t1, m=m1,f=f1)
edu2 = edu %>% select(year,t2,m2,f2) %>%
  mutate(class = "junior")%>% rename(t=t2, m=m2,f=f2)
edu3 = edu %>% select(year,t3,m3,f3) %>%
  mutate(class = "senior")%>% rename(t=t3, m=m3,f=f3)
edu4 = edu %>% select(year,t4,m4,f4) %>%
  mutate(class = "undergrad")%>% rename(t=t4, m=m4,f=f4)
edu5 = edu %>% select(year,t5,m5,f5) %>%
  mutate(class = "grad")%>% rename(t=t5, m=m5,f=f5)
edus = edu %>% select(year,s,ms,fs) %>%
  mutate(class = "self")%>% rename(t=s, m=ms,f=fs)
edu8 = edu %>% select(year,t8,m8,f8) %>%
  mutate(class = "other")%>% rename(t=t8, m=m8,f=f8)
edue = rbind(edu0,edu1,edu2,edu3,edu4,edu5,edus,edu8)
head(edue)
```

```
## # A tibble: 6 x 5
##   year      t      m      f class
##   <dbl> <dbl> <dbl> <dbl> <chr>
## 1  1997  3476  1852  1624 illiterate
## 2  1998  2863  1602  1261 illiterate
```

```
## 3 1999 2678 1546 1132 illiterate
## 4 2000 2487 1463 1024 illiterate
## 5 2001 2481 1472 1009 illiterate
## 6 2002 3217 1785 1432 illiterate
```

Plotting year-education-Crime relationship It can easily be seen that people with only high school offend the law more frequently.

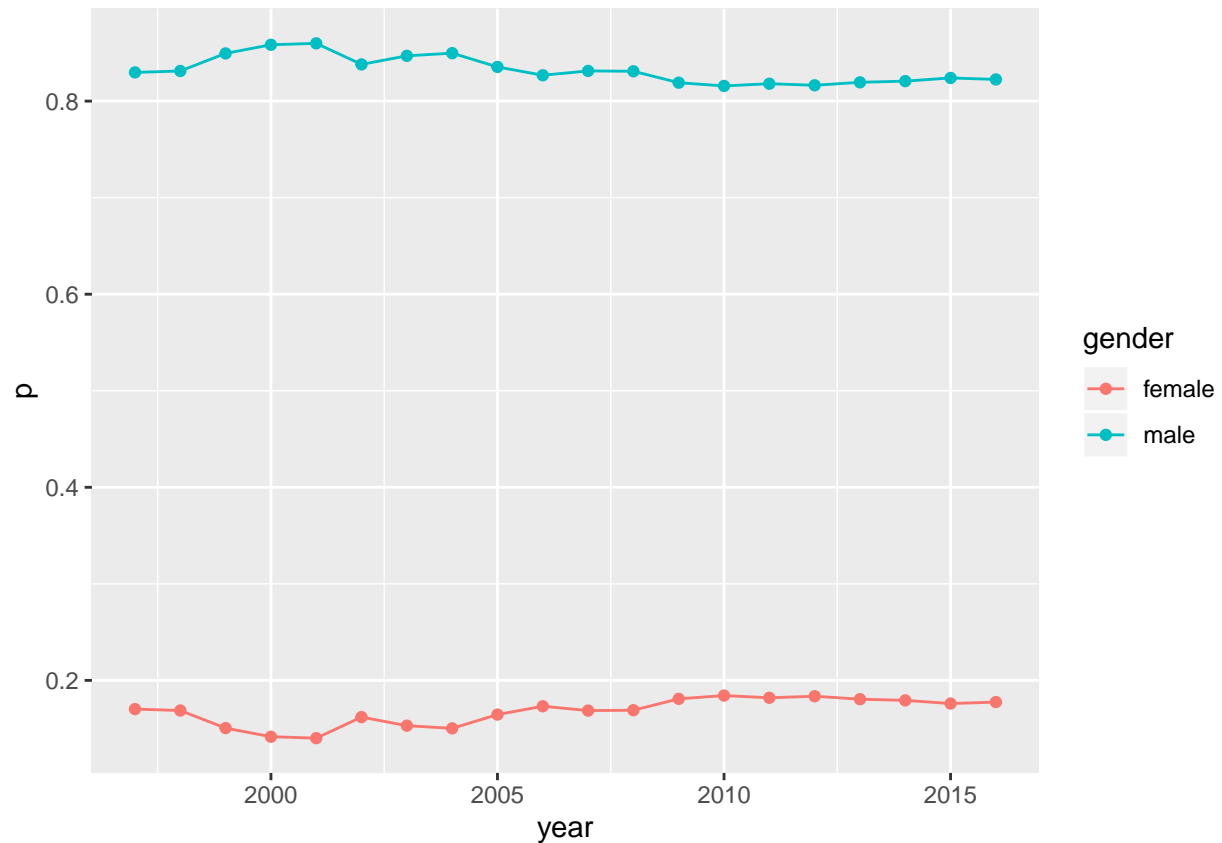
```
ggplot(data = edue,aes(x = year, y = t ,color=class ,group = class)) +
  geom_line()+
  geom_point()
```



Plotting Gender - Crime relationship conclusion: Men violate the Law much more frequently than women do

```
gender<-edu %>%
  select(year,t,m,f) %>%
  mutate(fp = f / t, mp = m / t)
gendera= gender %>% select(year,fp)%>% mutate(gender="female")%>% rename(p=fp)
genderb= gender %>% select(year,mp)%>% mutate(gender="male")%>% rename(p=mp)
gender= rbind(gendera,genderb)

ggplot(data = gender,aes(x=year,y=p,group=gender,color=gender))+
  geom_line()+
  geom_point()
```



Unemployment and Crime

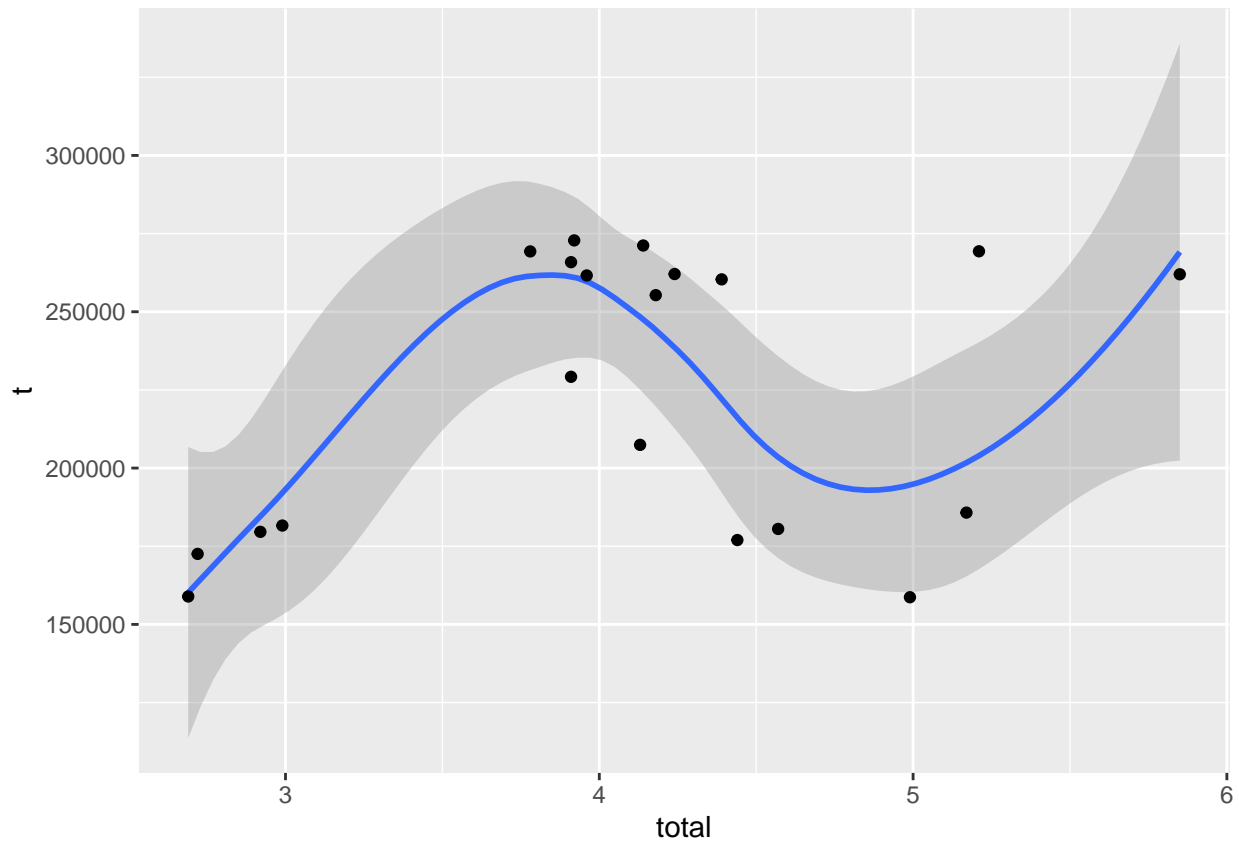
Data importing , cleaning and plotting Graph :total unemployment and crime

```

uneplay=read.csv("https://raw.githubusercontent.com/Vladislav292/finalproj/master/uneplay.csv")
plot2 = merge(edu,uneplay,by="year")
ggplot(data=plot2)+
  geom_smooth(mapping=aes(x=total,y=t))+
  geom_point(mapping =aes(x=total,y=t))

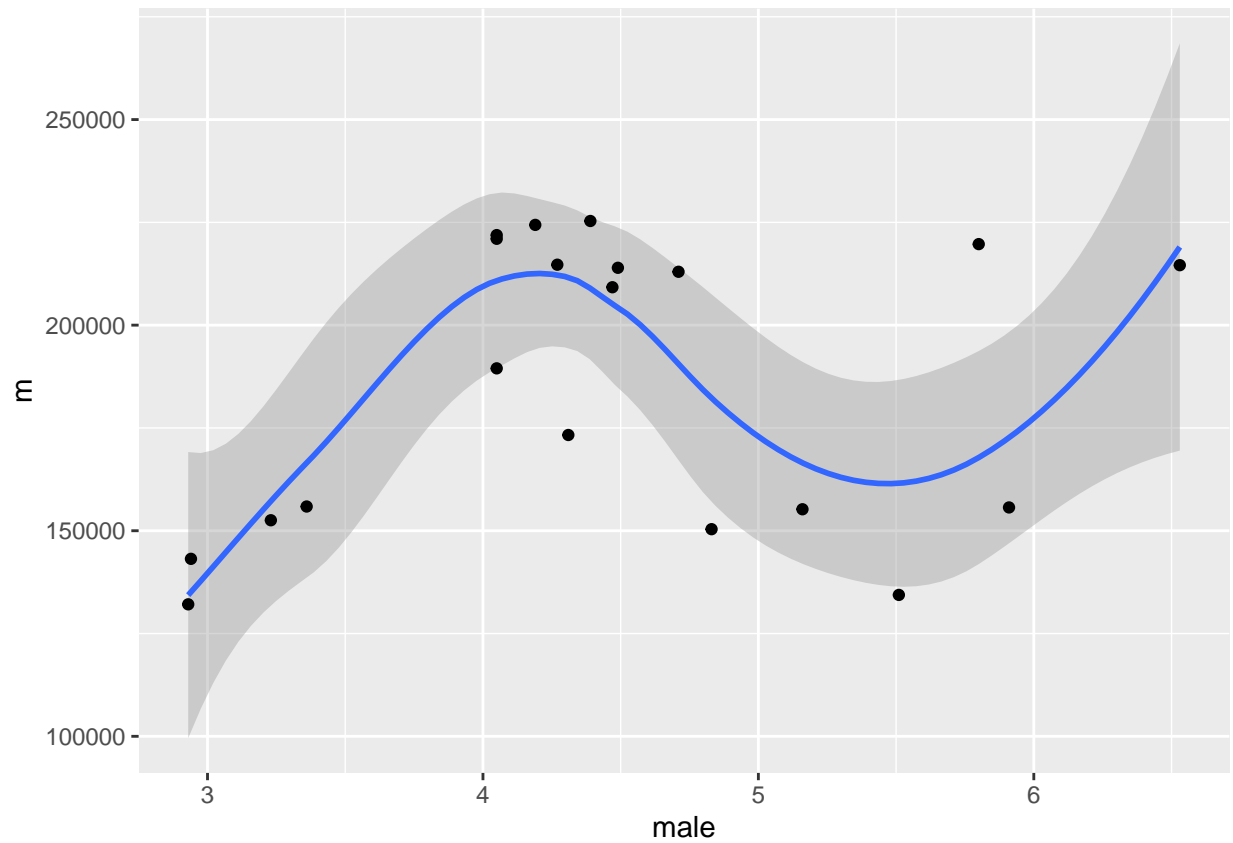
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Graph: female unemployment and crime

```
ggplot(data=plot2)+  
  geom_smooth(mapping=aes(x=total,y=t))+  
  geom_point(mapping =aes(x=total,y=t))  
  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

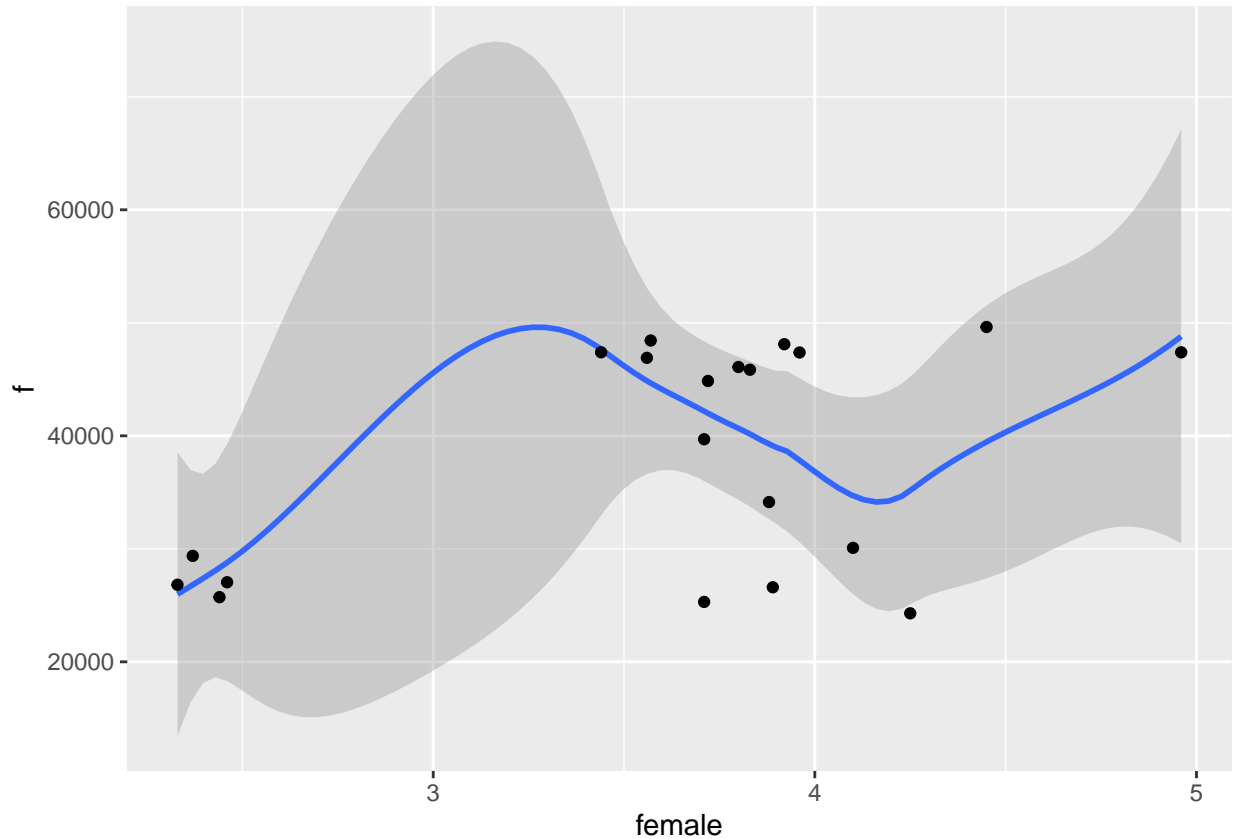


Graph: female unemployment and crime

```
ggplot(data=plot2)+
  geom_smooth(mapping=aes(x=female,y=f),model=lm)+
  geom_point(mapping =aes(x=female,y=f))
```

```
## Warning: Ignoring unknown parameters: model
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Correlation Coefficients

1.GDP-Crime 2.Unemployment-Crime 3.Male unemployment- Male crime 4.Female unemployment-Female crime 5.year- crime

```
cor(data_merged$di,data_merged$t)
```

```
## [1] 0.8476391
```

```
cor(plot2$total,data_merged$t)
```

```
## [1] 0.3167359
```

```
cor(plot2$male,plot2$m)
```

```
## [1] 0.2387764
```

```
cor(plot2$female,plot2$f)
```

```
## [1] 0.4785071
```

```
cor(data_merged$year,data_merged$t)
```

```
## [1] 0.8934133
```

Conclusions

1.As GDP increases, Crime increases. This might not be a good conclusion since GDP increase with year and crime also increases with year and the latter has a greater correlation coefficient. 2.High school graduates violate the law more often than other groups do . 3.Men violates the law more often than women do. 4.The

unemployment-crime correlation is more obvious in women than in men. 5. There is only little correlation in unemployment and crime.

To the authority

1. Getting a lot of money (aka. FADATSAI) cannot stop people from violating our law. 2. Education is relatively important compared to other factors.

Possible further research

1. Why Crime increases with year, are people today more disobedient than before?
2. Why is the unemployment-crime correlation more obvious in women than in men?
3. What caused a soar in crime during 2003-2007?