Senior Project

Jorge Jeronimo Fueyo Gonzalez

3/10/2022

S&DS Senior Project (S&DS 492)

Jorge Jeronimo Fueyo Gonzalez

Advisor: Professor Meyers

Spring 2022

Senior Project

Three major topics:

1. Factors that Contribute to Violence in Mexico
2. Effects of Violence on the Economic Development of Mexico
3. Economic Phenomenon of Collusion: Why do Cartels Collude and Cheat on Each Other?

```
# install packages
#install.packages("readxl")
#install.packages("tidyverse")
```

```
# load packages
library("readxl")
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("dplyr")
```

```
# load violence data

df_homicides_0 <- read.csv("INEGI_exporta_10_2_2022_13_0_2.csv")
```

Factors that Contribute to Violence in Mexico

1. How does level of education affects level of violence in a specific region (state)?

Relationship between level of education and:

a) homicides
b) feminicides
c) crime rate
d) kidnappings

- Homicides data is readily available
- Data on crime rate and kidnappings might be harder to find

2. What's the relationship between the political party in power in a region (state) with the level of violence in that region?

- Data set would have to be made by me based on information online
- Try to find time series data set

3. What's the relationship between the drug cartel in power in a region (state) with level of violence in that region?

- Data set would have to be made by me based on information online
- Try to find time series data set

4. Is there a pattern between both, the political party and drug cartel in power in a region (state) and their trends on violence?

```
# load economic data

df_GDP_PC_0 <- read.csv("/Users/jeronimofueyo/Documents/Jeronimo Fueyo/GitHub/Yale-SDS-492-Senior-Proje

df_GDP_0 <- read.csv("/Users/jeronimofueyo/Documents/Jeronimo Fueyo/GitHub/Yale-SDS-492-Senior-Project/

df_FDI_0 <- read.csv("/Users/jeronimofueyo/Documents/Jeronimo Fueyo/GitHub/Yale-SDS-492-Senior-Project/
```

```r
df_GDP_PC <- df_GDP_PC_0 %>%
  rename(date = DATE, GDP_PC = PCAGDPMXA646NWDB) %>%
  mutate(date = as.Date(date), year = as.numeric(format(date, format="%Y")))

df_GDP <- df_GDP_0 %>%
  rename(date = DATE, GDP = MKTGDPMXA646NWDB) %>%
  mutate(date = as.Date(date), year = as.numeric(format(date, format="%Y")))

df_FDI <- df_FDI_0 %>%
  rename(date = DATE, FDI = BPFADI03MXQ637N) %>%
  mutate(date = as.Date(date), year = as.numeric(format(date, format="%Y")))

df_FDI_per_year <- df_FDI %>%
  group_by(year) %>%
  summarize(FDI = mean(FDI))
```

Effects of Violence on the Economic Development of Mexico

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# head economic data

head(df_GDP_PC)
```

```
##         date   GDP_PC year
## 1 1960-01-01 345.2305 1960
## 2 1961-01-01 363.3933 1961
## 3 1962-01-01 378.1535 1962
## 4 1963-01-01 409.0456 1963
## 5 1964-01-01 469.4761 1964
## 6 1965-01-01 494.9703 1965
```

```r
head(df_GDP)
```

```
##         date       GDP year
## 1 1960-01-01 1.304e+10 1960
## 2 1961-01-01 1.416e+10 1961
## 3 1962-01-01 1.520e+10 1962
## 4 1963-01-01 1.696e+10 1963
## 5 1964-01-01 2.008e+10 1964
## 6 1965-01-01 2.184e+10 1965
```

```r
head(df_FDI)
```

```
##         date       FDI year
## 1 1980-01-01 400800000 1980
## 2 1980-04-01 509800000 1980
## 3 1980-07-01 680600000 1980
## 4 1980-10-01 498600000 1980
## 5 1981-01-01 530400000 1981
## 6 1981-04-01 970400000 1981
```

```r
head(df_FDI_per_year)
```

```
## # A tibble: 6 x 2
##    year       FDI
##   <dbl>     <dbl>
## 1  1980 522450000
## 2  1981 768950000
## 3  1982 475075000
## 4  1983 547925000
## 5  1984 385250000
## 6  1985 495875000
```

```r
# view economic data

View(df_GDP_PC)
View(df_GDP)
View(df_FDI)
View(df_FDI_per_year)
```

```r
df_homicides <- df_homicides_0[6:36,] %>%
  select(!X.35) %>%
  rename(year = Defunciones.por.homicidios, total_homicides = X,
         Aguascalientes_homicides = X.1, BajaCalifornia_homicides = X.2,
         BajaCaliforniaSur_homicides = X.3, Campeche = X.4,
         Coahuila = X.5, Colima = X.6,
         Chiapas = X.7, Chihuahua = X.8,
         CDMX = X.9, Durango = X.10,
         Guanajuato = X.11, Guerrero = X.12,
         Hidalgo = X.13, Jalisco = X.14,
         Mexico = X.15, Michoacan = X.16,
         Morelos = X.17, Nayarit = X.18,
         NuevoLeon = X.19, Oaxaca = X.20,
         Puebla = X.21, Queretaro = X.22,
         QuintanaRoo = X.23, SanLuisPotosi = X.24,
         Sinaloa = X.25, Sonora = X.26,
         Tabasco = X.27, Tamaulipas = X.28,
         Tlaxcala = X.29, Veracruz = X.30,
         Yucatan = X.31, Zacatecas = X.32,
         Extranjero = X.33, NotIdentified = X.34) %>%
  mutate(year = as.numeric(year),
         total_homicides = as.numeric(gsub(",", "", total_homicides)),
         Aguascalientes_homicides = as.numeric(gsub(",", "", Aguascalientes_homicides)),
         BajaCalifornia_homicides = as.numeric(gsub(",", "", BajaCalifornia_homicides)),
         BajaCaliforniaSur_homicides = as.numeric(gsub(",", "", BajaCaliforniaSur_homicides)),
         Campeche = as.numeric(gsub(",", "", Campeche)),
         Coahuila = as.numeric(gsub(",", "", Coahuila)),
         Colima = as.numeric(gsub(",", "", Colima)),
         Chiapas = as.numeric(gsub(",", "", Chiapas)),
         Chihuahua = as.numeric(gsub(",", "", Chihuahua)),
         CDMX = as.numeric(gsub(",", "", CDMX)),
         Durango = as.numeric(gsub(",", "", Durango)),
         Guanajuato = as.numeric(gsub(",", "", Guanajuato)),
         Guerrero = as.numeric(gsub(",", "", Guerrero)),
```

```
        Hidalgo = as.numeric(gsub(",", "", Hidalgo)),
        Jalisco = as.numeric(gsub(",", "", Jalisco)),
        Mexico = as.numeric(gsub(",", "", Mexico)),
        Michoacan = as.numeric(gsub(",", "", Michoacan)),
        Morelos = as.numeric(gsub(",", "", Morelos)),
        Nayarit = as.numeric(gsub(",", "", Nayarit)),
        NuevoLeon = as.numeric(gsub(",", "", NuevoLeon)),
        Oaxaca = as.numeric(gsub(",", "", Oaxaca)),
        Puebla = as.numeric(gsub(",", "", Puebla)),
        Queretaro = as.numeric(gsub(",", "", Queretaro)),
        QuintanaRoo = as.numeric(gsub(",", "", QuintanaRoo)),
        SanLuisPotosi = as.numeric(gsub(",", "", SanLuisPotosi)),
        Sinaloa = as.numeric(gsub(",", "", Sinaloa)),
        Sonora = as.numeric(gsub(",", "", Sonora)),
        Tabasco = as.numeric(gsub(",", "", Tabasco)),
        Tamaulipas = as.numeric(gsub(",", "", Tamaulipas)),
        Tlaxcala = as.numeric(gsub(",", "", Tlaxcala)),
        Veracruz = as.numeric(gsub(",", "", Veracruz)),
        Yucatan = as.numeric(gsub(",", "", Yucatan)),
        Zacatecas = as.numeric(gsub(",", "", Zacatecas)),
        Extranjero = as.numeric(gsub(",", "", Extranjero)),
        NotIdentified = as.numeric(gsub(",", "", NotIdentified))) %>%
  replace(is.na(.), 0)
```

```
# head homicides data

head(df_homicides)
```

```
##   year total_homicides Aguascalientes_homicides BajaCalifornia_homicides
## 1 1990           14493                       43                       260
## 2 1991           15128                       53                       283
## 3 1992           16594                       35                       313
## 4 1993           16040                       32                       290
## 5 1994           15839                       23                       352
## 6 1995           15612                       34                       398
##   BajaCaliforniaSur_homicides Campeche Coahuila Colima Chiapas Chihuahua CDMX
## 1                          12       82      182     78     272       306 1272
## 2                          16       65      235     99     470       454 1100
## 3                          18       87      229    121     437       462 1259
## 4                          21       77      220     70     504       456 1249
## 5                          28       74      214     78     648       573 1180
## 6                          32       98      176     62     708       747 1361
##   Durango Guanajuato Guerrero Hidalgo Jalisco Mexico Michoacan Morelos Nayarit
## 1     289        361      770     202     832   3449      1170     379     243
## 2     335        397     1005     150     782   3464      1163     342     232
## 3     601        352     1578     122     783   3354      1389     376     251
## 4     344        379     1296     131     781   3315      1345     493     253
## 5     347        324     1198     158     703   3185      1270     463     268
## 6     307        284     1260     156     781   2748      1124     353     193
##   NuevoLeon Oaxaca Puebla Queretaro QuintanaRoo SanLuisPotosi Sinaloa Sonora
## 1        72   1211    526        82          57           234     445    166
## 2        91   1282    422        77          53           210     542    193
## 3       120   1275    540        78          54           258     594    241
```

```
## 4       121   1335    597        80          80           240     552    210
## 5       134   1175    496       101         156           234     596    249
## 6       123   1183    599        88          74           209     617    315
##   Tabasco Tamaulipas Tlaxcala Veracruz Yucatan Zacatecas Extranjero
## 1     132        266       47      776      65       145         67
## 2     119        338       31      843      63       138         81
## 3     136        407       46      753      48       165        112
## 4     154        369       45      696      59       144        102
## 5     239        355       42      703      47       121        105
## 6     234        321       71      617      49       164        126
##   NotIdentified
## 1             0
## 2             0
## 3             0
## 4             0
## 5             0
## 6             0
```

```
# view homicides data

View(df_homicides)
```

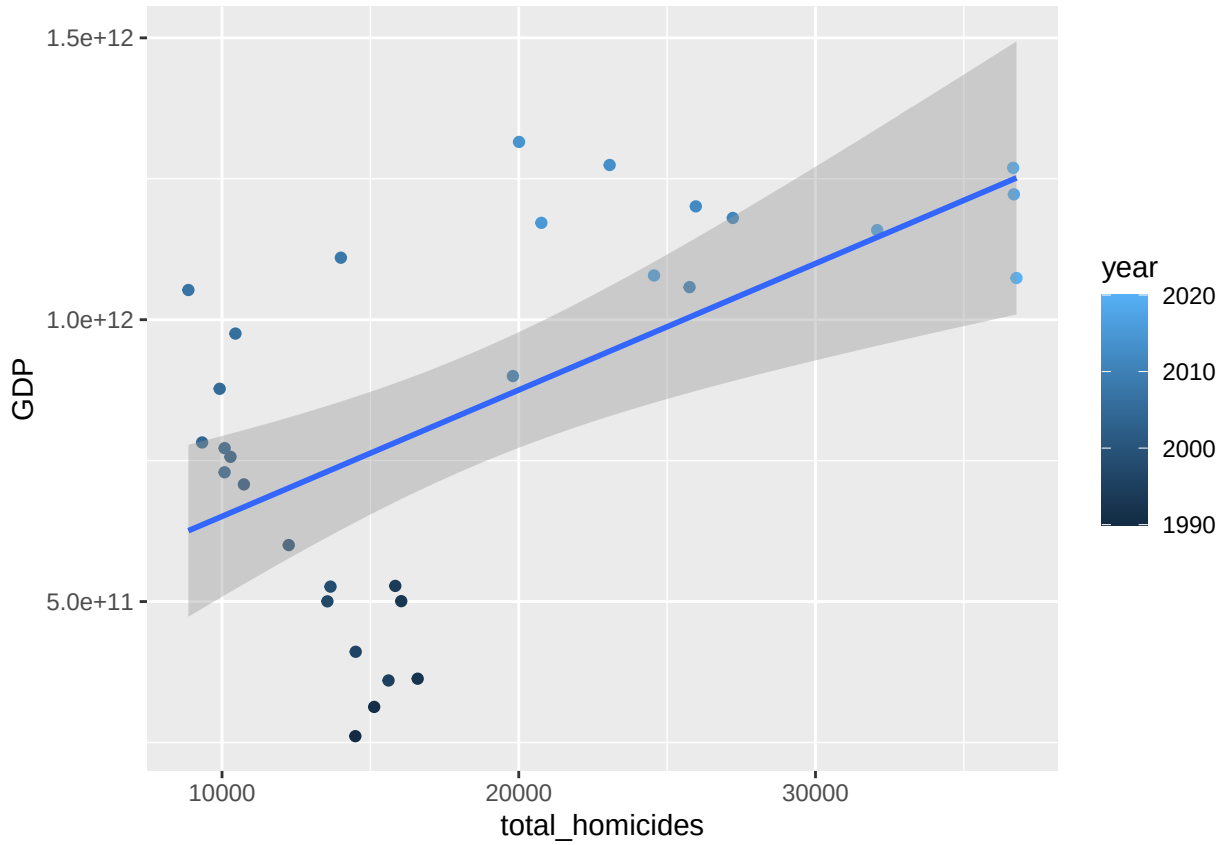
```
df_violence_econ <- df_homicides %>%
  left_join(df_GDP, by="year") %>%
  left_join(df_GDP_PC, by="year") %>%
  left_join(df_FDI_per_year, by="year") %>%
  select(year, total_homicides, GDP, GDP_PC, FDI) %>%
  mutate(years_since_1989 = year - 1989)

head(df_violence_econ)
```

```
##   year total_homicides          GDP   GDP_PC        FDI years_since_1989
## 1 1990           14493 261253582806 3112.269  658325000                1
## 2 1991           15128 313142768453 3661.948 1190375000                2
## 3 1992           16594 363157598242 4170.623 1098200000                3
## 4 1993           16040 500736065605 5650.026 1044875000                4
## 5 1994           15839 527813238126 5854.418 2743125000                5
## 6 1995           15612 360073909244 3928.224 2381575000                6
```

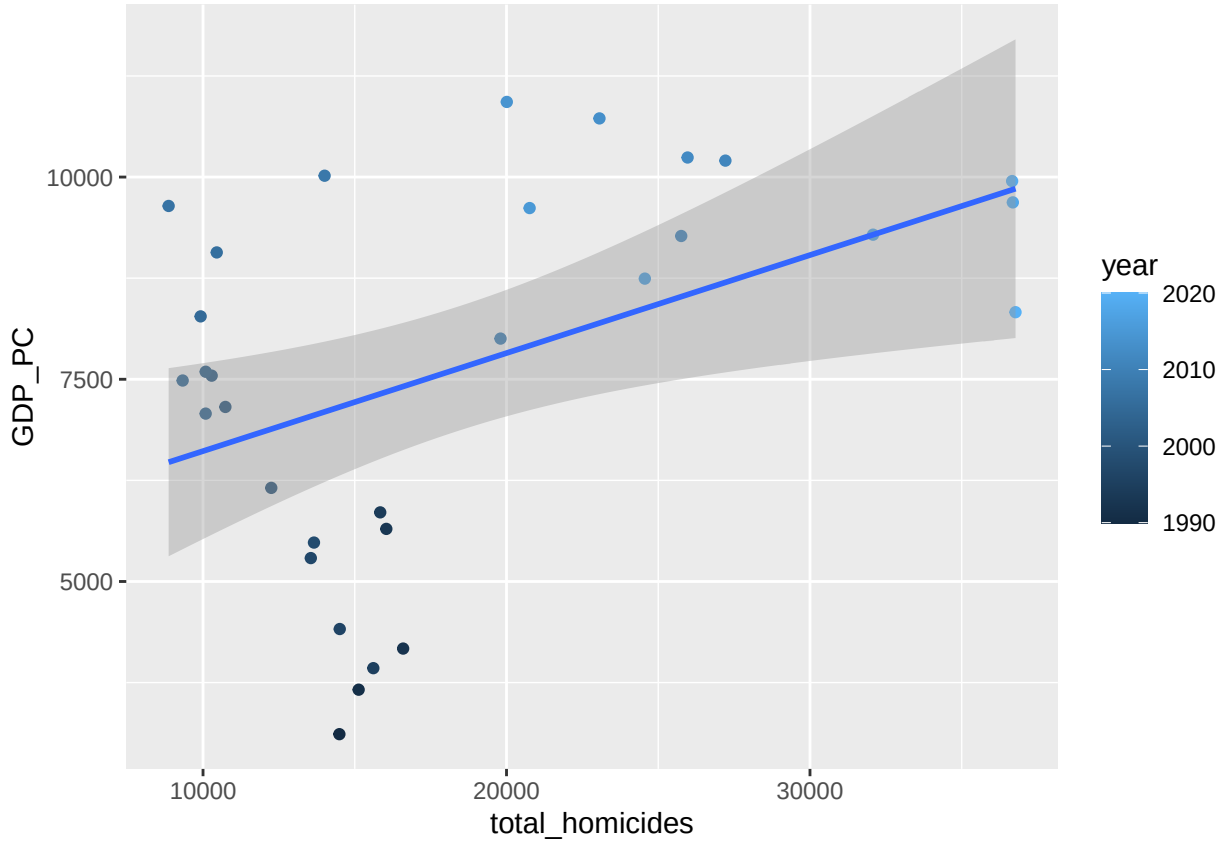
```
ggplot(df_violence_econ, aes(x=total_homicides, y=GDP, color=year)) +
  geom_point() +
  geom_smooth(method="lm")
```

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




```
ggplot(df_violence_econ, aes(x=total_homicides, y=GDP_PC, color=year)) +
  geom_point() +
  geom_smooth(method="lm")
```

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

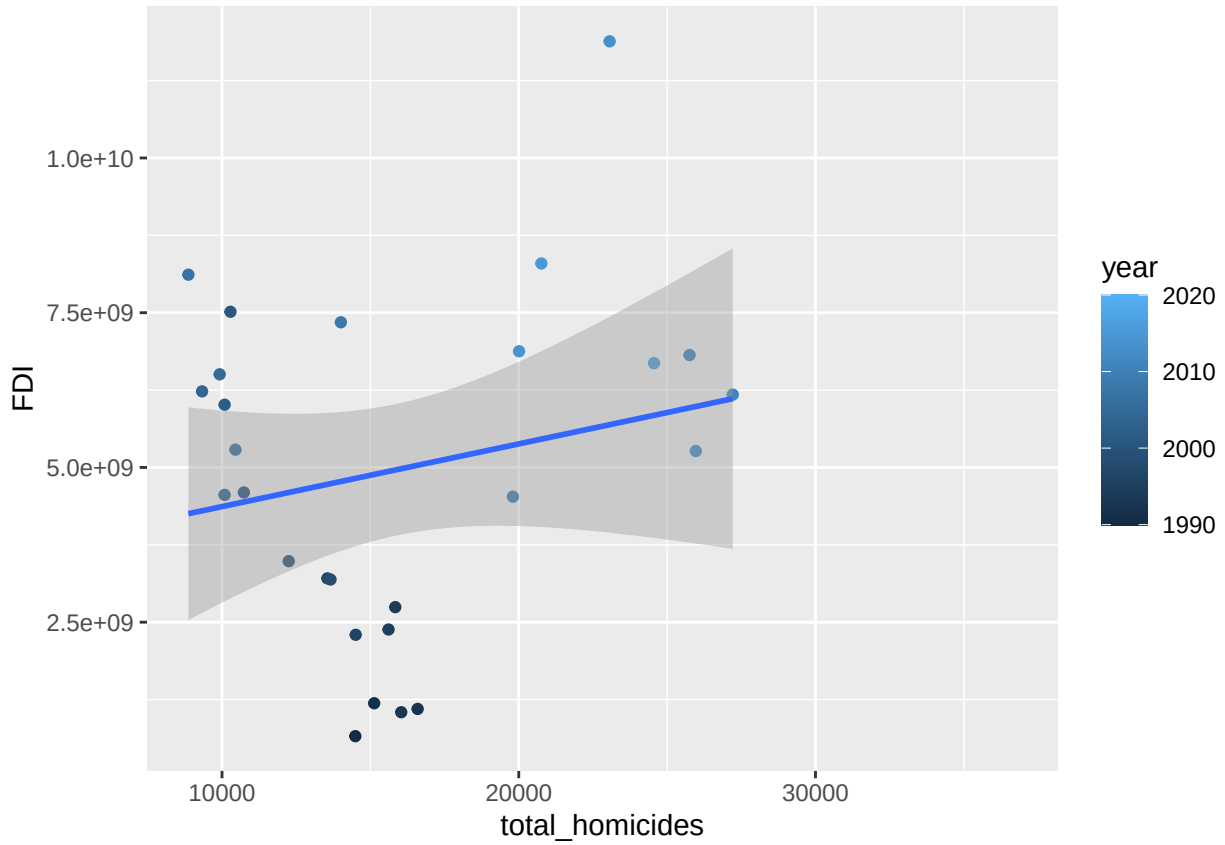



```
ggplot(df_violence_econ, aes(x=total_homicides, y=FDI, color=year)) +
  geom_point() +
  geom_smooth(method="lm")
```

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

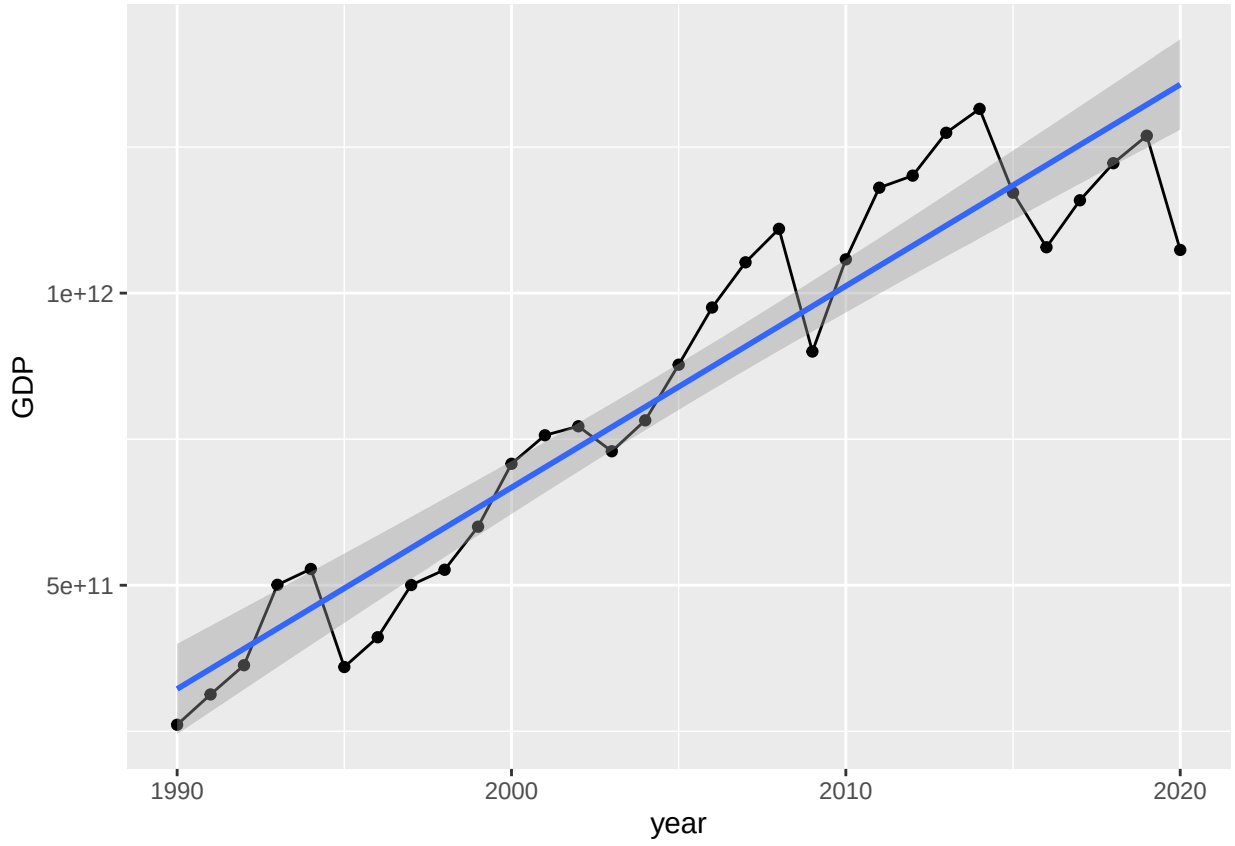
```
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

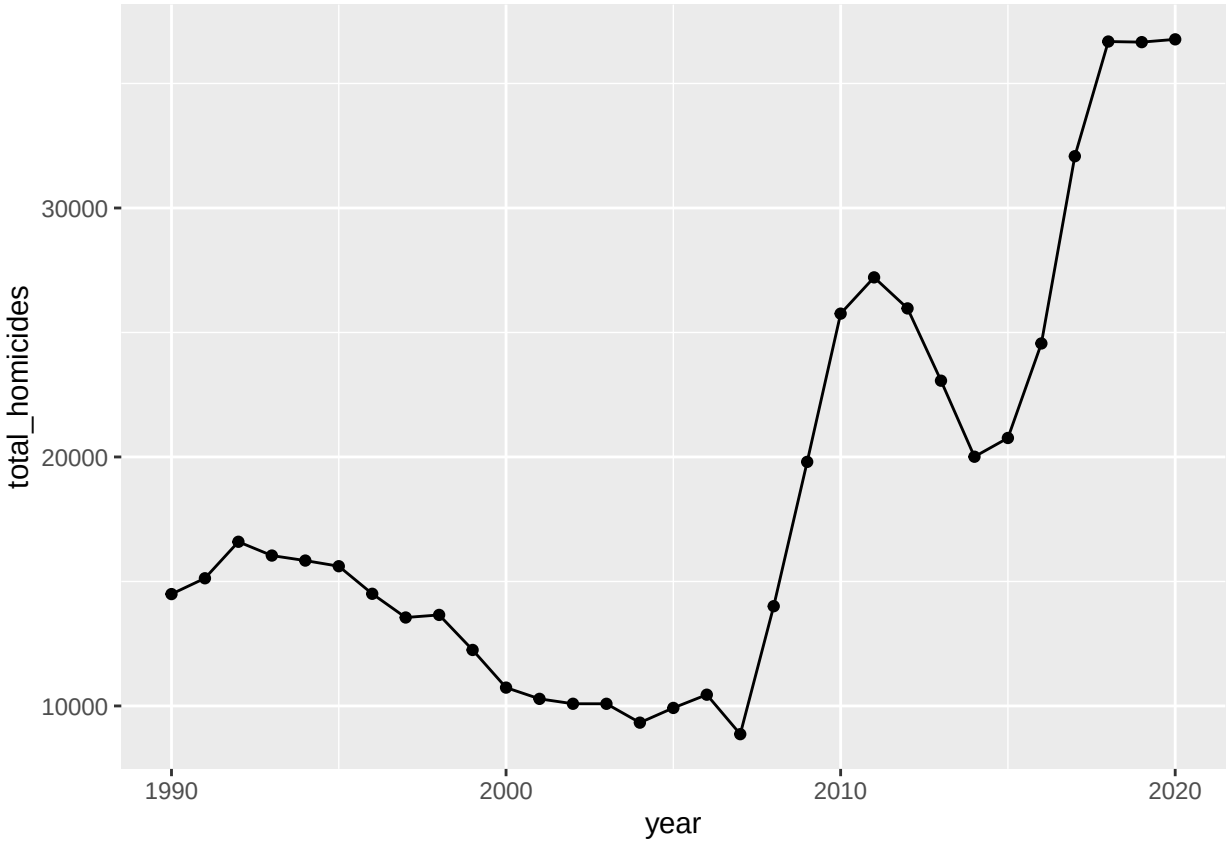



```
ggplot(df_violence_econ, aes(x=year, y=GDP)) +
  geom_point() +
  geom_line() +
  geom_smooth(method="lm")
```

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




```
ggplot(df_violence_econ, aes(x=year, y=total_homicides)) +
  geom_point() +
  geom_line()
```




```
fit_1 <- lm(total_homicides ~ years_since_1989 + GDP, df_violence_econ)
summary(fit_1)
```

```
##
## Call:
## lm(formula = total_homicides ~ years_since_1989 + GDP, data = df_violence_econ)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8974  -4381   -763   5068   8044
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.471e+04  3.456e+03   4.257 0.000210 ***
## years_since_1989  1.511e+03  3.510e+02   4.306 0.000184 ***
## GDP              -2.439e-08  9.641e-09  -2.530 0.017319 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5593 on 28 degrees of freedom
## Multiple R-squared:    0.6,  Adjusted R-squared:  0.5715
## F-statistic:    21 on 2 and 28 DF,  p-value: 2.682e-06
```

```
fit_2 <- lm(total_homicides ~ years_since_1989 + GDP_PC, df_violence_econ)
summary(fit_2)
```

```
##
## Call:
## lm(formula = total_homicides ~ years_since_1989 + GDP_PC, data = df_violence_econ)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8169  -3630  -1038   4342   8631
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      20384.1827  4018.5454   5.073 2.27e-05 ***
## years_since_1989  1374.9902   222.2731   6.186 1.11e-06 ***
## GDP_PC              -3.1413     0.8776  -3.580  0.00128 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5135 on 28 degrees of freedom
## Multiple R-squared:  0.6629, Adjusted R-squared:  0.6388
## F-statistic: 27.53 on 2 and 28 DF,  p-value: 2.45e-07
```

```
fit_3 <- lm(total_homicides ~ years_since_1989 + FDI, df_violence_econ)
summary(fit_3)
```

```
##
## Call:
## lm(formula = total_homicides ~ years_since_1989 + FDI, data = df_violence_econ)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7333.6 -2993.6    -6.2  2729.4  9898.0
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.231e+04  1.859e+03   6.623 7.51e-07 ***
## years_since_1989  8.058e+02  2.040e+02   3.949 0.000599 ***
## FDI              -1.555e-06  6.013e-07  -2.586 0.016193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4461 on 24 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.4211, Adjusted R-squared:  0.3728
## F-statistic: 8.728 on 2 and 24 DF,  p-value: 0.001417
```

```
fit_3 <- lm(total_homicides ~ years_since_1989 + GDP_PC + FDI, df_violence_econ)
summary(fit_3)
```

```
##
## Call:
## lm(formula = total_homicides ~ years_since_1989 + GDP_PC + FDI,
##     data = df_violence_econ)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -6795.7 -2873.1  -333.7  2868.4  9620.7
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.406e+04  4.236e+03   3.319  0.00299 **
## years_since_1989  9.137e+02  3.126e+02   2.923  0.00766 **
## GDP_PC           -5.388e-01  1.167e+00  -0.462  0.64874
## FDI              -1.411e-06  6.868e-07  -2.054  0.05147 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4536 on 23 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.4264, Adjusted R-squared:  0.3516
## F-statistic: 5.699 on 3 and 23 DF,  p-value: 0.004544
```

```
fit_4 <- lm(FDI ~ years_since_1989 + total_homicides, df_violence_econ)
summary(fit_4)
```

```
##
## Call:
## lm(formula = FDI ~ years_since_1989 + total_homicides, data = df_violence_econ)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.901e+09 -5.858e+08 -6.275e+07  2.994e+08  4.567e+09
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2479456648  789968321   3.139  0.00445 **
## years_since_1989   336242406   38453014   8.744 6.31e-09 ***
## total_homicides      -140147      54185  -2.586  0.01619 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.339e+09 on 24 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.7718, Adjusted R-squared:  0.7528
## F-statistic: 40.59 on 2 and 24 DF,  p-value: 1.991e-08
```

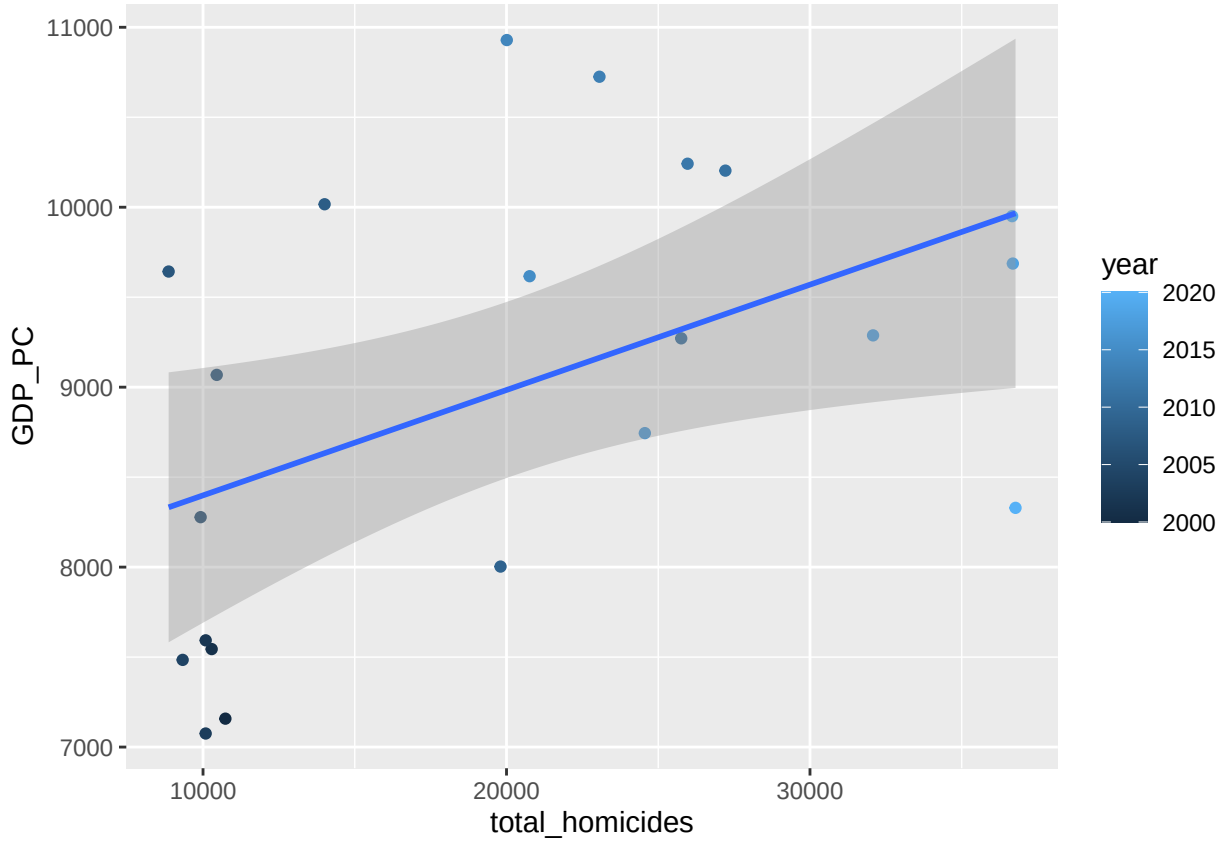
```
df_violence_econ_not_90s <- df_violence_econ %>%
  mutate(years_since_1999 = year - 1999) %>%
  filter(year >= 2000)
head(df_violence_econ_not_90s)
```

```
##   year total_homicides          GDP   GDP_PC        FDI years_since_1989
## 1 2000           10737 707906744575 7157.814 4595600000               11
## 2 2001           10285 756706300590 7544.569 7515000000               12
## 3 2002           10088 772106378935 7593.137 6013775000               13
## 4 2003           10087 729336319677 7075.370 4555900000               14
## 5 2004            9329 782240601985 7484.486 6229000000               15
## 6 2005            9921 877476221382 8277.672 6504575000               16
```

```
##   years_since_1999
## 1                1
## 2                2
## 3                3
## 4                4
## 5                5
## 6                6
```

```
ggplot(df_violence_econ_not_90s, aes(x=total_homicides, y=GDP_PC, color=year)) +
  geom_point() +
  geom_smooth(method="lm")
```

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




```
fit_5 <- lm(total_homicides ~ years_since_1999 + GDP_PC, df_violence_econ_not_90s)
summary(fit_5)
```

```
##
## Call:
## lm(formula = total_homicides ~ years_since_1999 + GDP_PC, data = df_violence_econ_not_90s)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6651.3 -3067.3   105.2  3243.8  6871.1
```

```
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      12999.628   7983.842   1.628    0.121
## years_since_1999  1597.749    194.945   8.196 1.73e-07 ***
## GDP_PC              -1.159      1.012  -1.146    0.267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4208 on 18 degrees of freedom
## Multiple R-squared:  0.8388, Adjusted R-squared:  0.8209
## F-statistic: 46.83 on 2 and 18 DF,  p-value: 7.347e-08
```

```
fit_6 <- lm(total_homicides ~ years_since_1989 + GDP_PC + FDI + GDP_PC*FDI, df_violence_econ)
summary(fit_6)
```

```
##
## Call:
## lm(formula = total_homicides ~ years_since_1989 + GDP_PC + FDI +
##     GDP_PC * FDI, data = df_violence_econ)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5776.4 -2578.5  -264.6  1296.9  7389.0
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.591e+04  4.764e+03   5.438 1.84e-05 ***
## years_since_1989  9.788e+02  2.544e+02   3.847 0.000875 ***
## GDP_PC           -2.138e+00  1.047e+00  -2.042 0.053312 .
## FDI              -5.655e-06  1.306e-06  -4.329 0.000270 ***
## GDP_PC:FDI        4.790e-10  1.333e-10   3.593 0.001619 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3682 on 22 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.6385, Adjusted R-squared:  0.5728
## F-statistic: 9.714 on 4 and 22 DF,  p-value: 0.0001105
```

1. What's the effect of violence (homicides/crime rates) on economic development in the country as a whole?

Effect of violence (homicides/crime rates) on:

a) GDP/GDP per capita? i) Already have the data
b) foreign investment i) Already have the data

2. What's the effect of violence (homicides/crime rates) on employment and business proliferation?

Effect of violence on:

a) level of unemployment per region i) not sure if I can get data on unemployment per region each year
b) number of business created per region i) might be hard to get data on this

Economic Phenomenon of Collusion: Why do Cartels Collude and Cheat on Each Other? In Microeconomics, we see that firms have an incentive to collude and sometimes cheat on each other. This seldom happens in regulated markets. The drug market through which these criminal groups operate is illegal, which makes it unregulated. Drug cartels are incentivized to collude and form aliances, but sometimes also cheat on each other and create war, which results in more violence. In this part of the project, I aim to discuss this phenomenon and how it has affected levels of violence in the country.

- I see it hard to find a way to use data to support this discussion so I don't think I'll be using data here