

Correlation of Location, Year, Economic Status, and Demographic Information with Suicide Rate

STAT 420, Summer 2020, Kun Hu, Madeline Old, Yuwen Xiang, Patrick Willhalm

07/18/2020

```
library(tidyverse)
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.3.2    v purrr  0.3.4
## v tibble  3.0.3    v dplyr  1.0.0
## v tidyr   1.1.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(knitr)
```

```
options(scipen = 1, digits = 4, width = 80)
opts_chunk$set(cache = TRUE, autodep = TRUE)
```

```
'%!in%' <- function(x,y)!('%in%'(x,y))
```

```
suicide <- read.csv('master.csv') %>%
  as_tibble(.)
```

```
n <- nrow(suicide)
buffer <- rep(0, n)
```

```
sex <- buffer
sex[suicide$sex == 'male'] <- 1
```

```
age <- buffer
age[suicide$age == '5-14 years'] <- 10
age[suicide$age == '15-24 years'] <- 20
age[suicide$age == '25-34 years'] <- 30
age[suicide$age == '35-54 years'] <- 45
age[suicide$age == '55-74 years'] <- 65
age[suicide$age == '75+ years'] <- 85
```

```

suicide <- suicide %>%
  select(-one_of('age', 'sex')) %>%
  add_column(age, sex) %>%
  rename(gdp_year = gdp_for_year....,
         gdp_per_capita = gdp_per_capita....,
         country = i..country,
         hdi_year = HDI.for.year)

suicide$suicide_year <- buffer
suicide$suicide_year_female <- buffer
suicide$suicide_year_male <- buffer
suicide$population_year <- buffer

for (country_year in unique(suicide$country.year)) {
  data_curr <- suicide[suicide$country.year == country_year,]

  suicide_year_curr <- sum(data_curr$suicides_no)
  suicide_year_female_curr <- sum(data_curr[data_curr$sex == 0,]$suicides_no)
  suicide_year_male_curr <- sum(data_curr[data_curr$sex == 1,]$suicides_no)
  population_year_curr <- sum(data_curr[data_curr$sex == 1,]$population)

  suicide[suicide$country.year == country_year,]$suicide_year <- suicide_year_curr
  suicide[suicide$country.year == country_year,]$suicide_year_female <- suicide_year_female_curr
  suicide[suicide$country.year == country_year,]$suicide_year_male <- suicide_year_male_curr
  suicide[suicide$country.year == country_year,]$population_year <- population_year_curr
}

head(suicide)

```

```

## # A tibble: 6 x 16
##   country year suicides_no population suicides.100k.p~ country.year hdi_year
##   <chr>   <int>      <int>      <int>          <dbl> <chr>          <dbl>
## 1 Albania 1987         21    312900          6.71 Albania1987      NA
## 2 Albania 1987         16    308000          5.19 Albania1987      NA
## 3 Albania 1987         14    289700          4.83 Albania1987      NA
## 4 Albania 1987          1    21800          4.59 Albania1987      NA
## 5 Albania 1987          9   274300          3.28 Albania1987      NA
## 6 Albania 1987          1    35600          2.81 Albania1987      NA
## # ... with 9 more variables: gdp_year <chr>, gdp_per_capita <int>,
## #   generation <chr>, age <dbl>, sex <dbl>, suicide_year <dbl>,
## #   suicide_year_female <dbl>, suicide_year_male <dbl>, population_year <dbl>

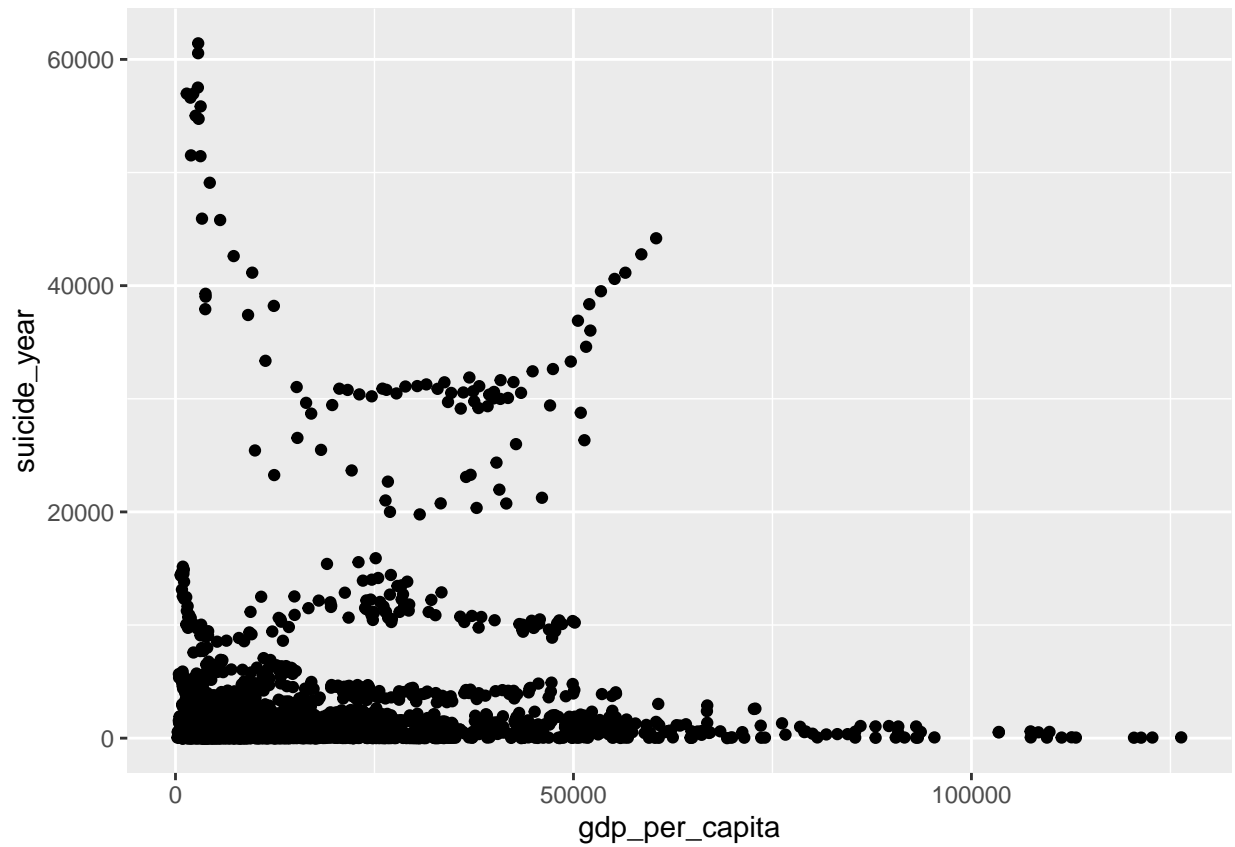
```

```

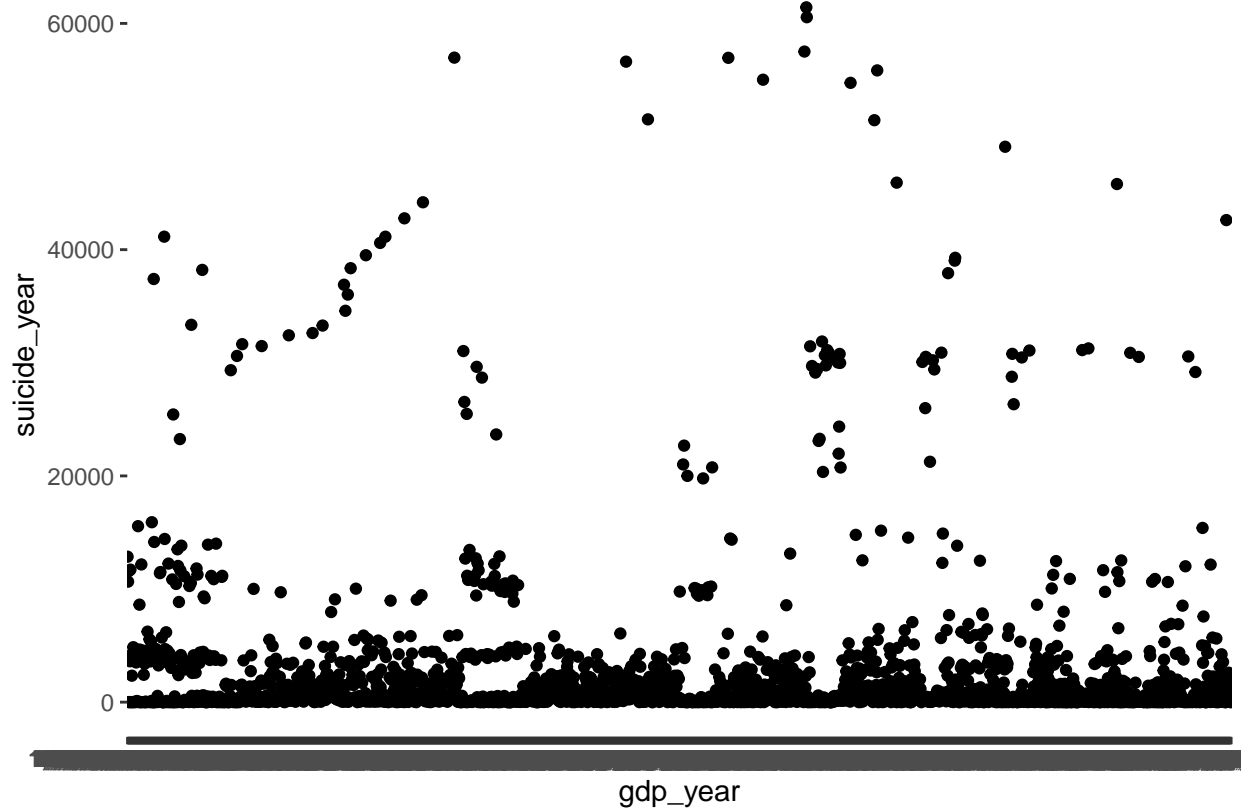
df_suicide_gdp <- suicide %>%
  distinct(country.year, .keep_all = TRUE)

ggplot(df_suicide_gdp) +
  geom_point(aes(x = gdp_per_capita, y = suicide_year))

```



```
ggplot(df_suicide_gdp) +  
  geom_point(aes(x = gdp_year, y = suicide_year))
```



```
list_year = sort(unique(df_suicide_gdp$year))
suicide_by_year <- c()
for (year in list_year) {
  suicide_by_year[length(suicide_by_year) + 1] <- sum(df_suicide_gdp[df_suicide_gdp$year == year,]$suicide)
}

ggplot() +
  geom_line(aes(x = list_year, y = suicide_by_year))
```

