

Suicide People in Japan and the USA

Descriptive Data Analysis

I worked on a dataset with the suicide population in over 100 countries through 1979-2015. It includes detailed information about suicide number and total population classified by country, age group, and sex group. The dataset is from Szamil (Szamil, 1), a user of Kaggle, and the World Health Organisation (WHO) collected the original data.

According to the limited awareness of demography and health in many countries, there is a massive amount of missing data of their population or the number of suicide people. In this report, I choose the data of Japan and USA as the objects to learn the trends and causes of the suicide population. However, there are still limitations in those two datasets. For example, evenly separated age gap, which in some groups is 10 years, while the age gap in other groups is 20 years make the group with long age gap has doomed large number. Also, there is not data of income or employment rate in the dataset, which limits my evaluation on the cause and effect of the suicide trends.

In this dataset, the age of people was classified into 6 groups: "5-14 years", "15-24 years", "25-34 years", "35-54 years", "55-74 years" and "75+ years". Since the age gap in the different group was varied, I used the ratio of suicide number over the total population to see the tendency of suicide (Fig.1&2). It was evident that the suicide number over population in both the USA and Japan increased as people grow older, while the children under 14-year-old almost didn't suicide.

The result is reasonable, because older people, especially those who are above 75-year-old, are more likely to face the plight of the health. Prolonging their life needs a lot of money, and suffering the extreme physical pain requires tremendous courage, many people will relief themselves by suicide.

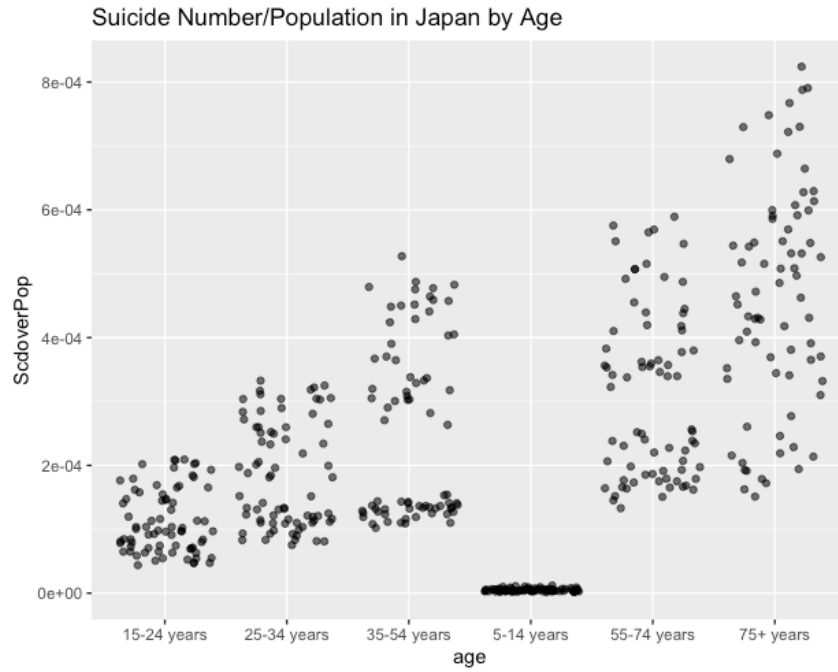


Fig.1

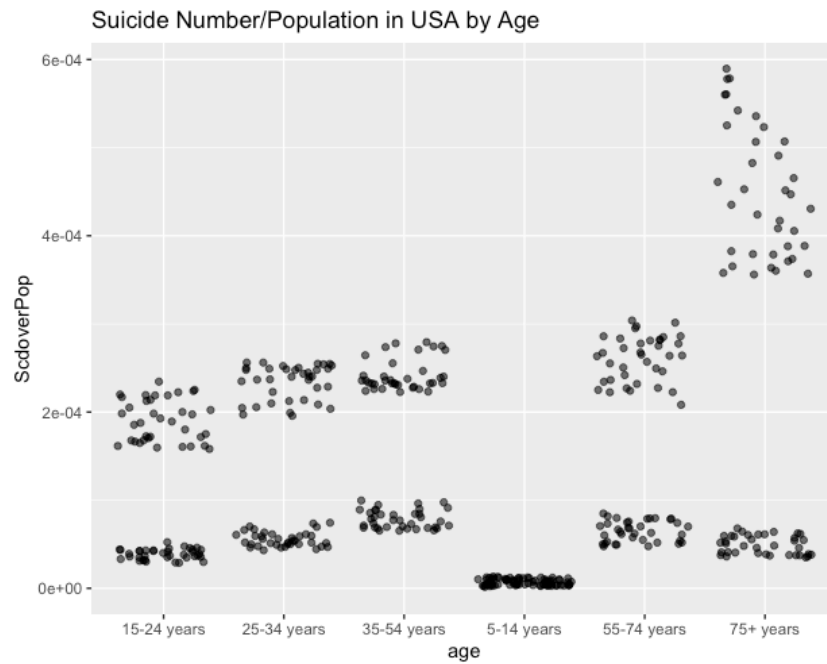


Fig.2

The range of the suicide number in Japan was 12-8949, while the range of the suicide number in USA was 26-11767. Although the total population of the USA was higher,

the mean of the suicide number in Japan was 3578 which was much more than 2706, the mean of the suicide number in USA. This drove to the fact that people in Japan were more likely to suicide than people in the USA. Based on the suicide number in Japan and USA (Fig.3&4), it was evident that in both countries, much more males suicided than females, especially the age group of 35-54 years. There was also an increasing suicide number of Japanese females as their age growing, while the number of suicide female in the USA and the number of suicide male in both counties were the highest among the age group of 35-54 years.

The result is expected since the female in Japan hardly needs to work compared with the other groups, and the age of 35-54 for most people is the period taking huge social pressure which leads to suicide.

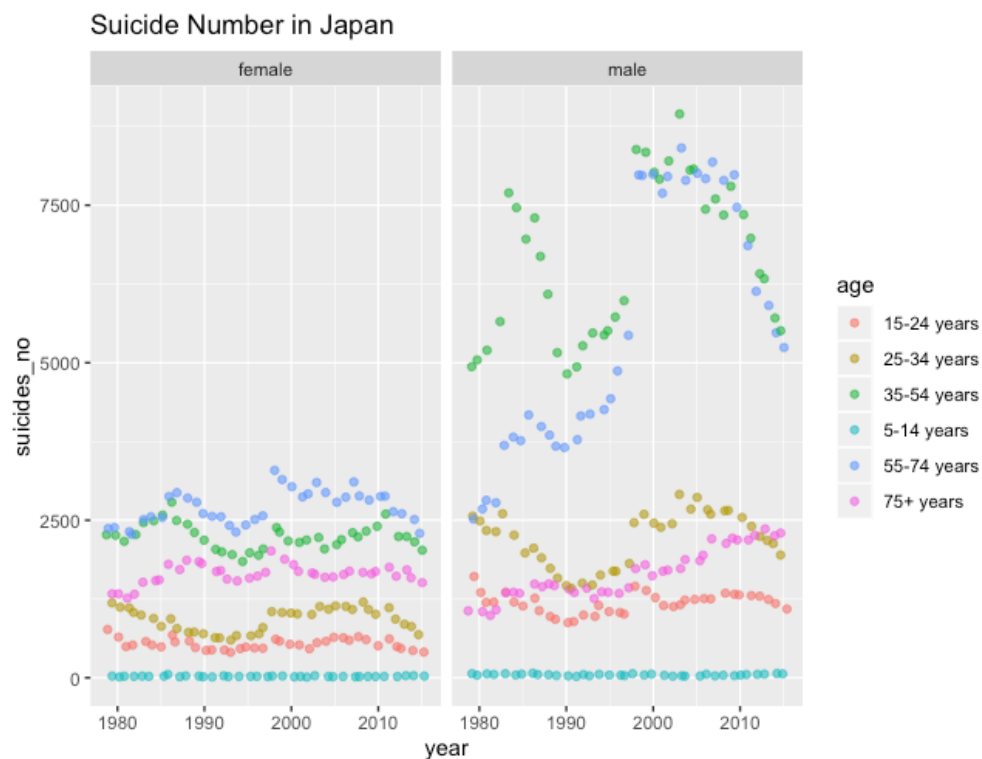


Fig.3

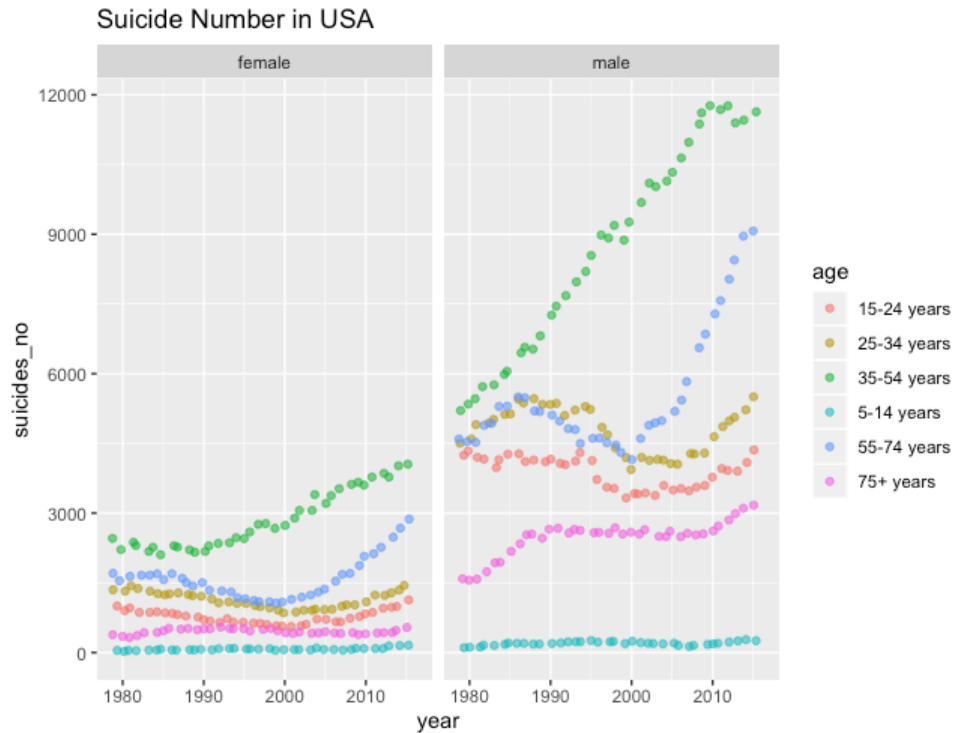


Fig.4

Focus on the age group of 35-54 years (Fig.5&6), more males were likely to suicide than female. The mean of the suicide female was 2239, but the mean of the suicide male was 6642. Similar situation happens in the USA that the mean of the suicide female was 2848, while the mean of suicide male was 8725. Looking at the graph of Japan, there were two periods that the number of suicide male boosting and tripled the number of suicide female: one was in the mid-1980s, and another was around 2000. In the USA, the number of suicide people was increasing every year, and the number of suicide male was double to triple the number of suicide female.

On the population charts of Japan (Fig.7), the total populations in "35-54 years" group of female and male were similar in shapes. The reason leads to the number of suicide male doubling number of suicide female should be the high financial pressure, so when there was an economic crisis (around 1985 or 2000), the number of suicide male tripled the number of suicide female. Compared to the curve of the USA population in 35-54

years (Fig.8) with the curve of USA suicide number in 35-54 years, I found the increasing patterns of the male were almost overlapping. However, the suicide number of females didn't increase much no matter the total population grew a lot, which was unexpected.

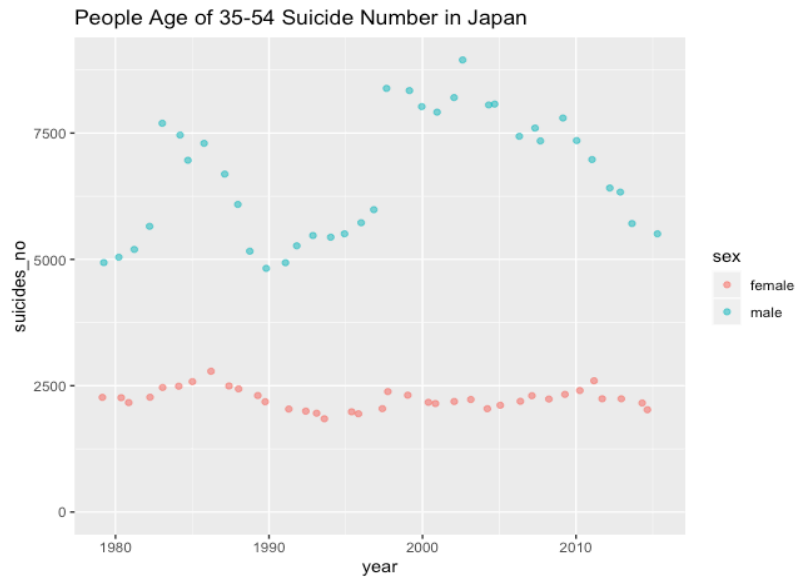


Fig.5

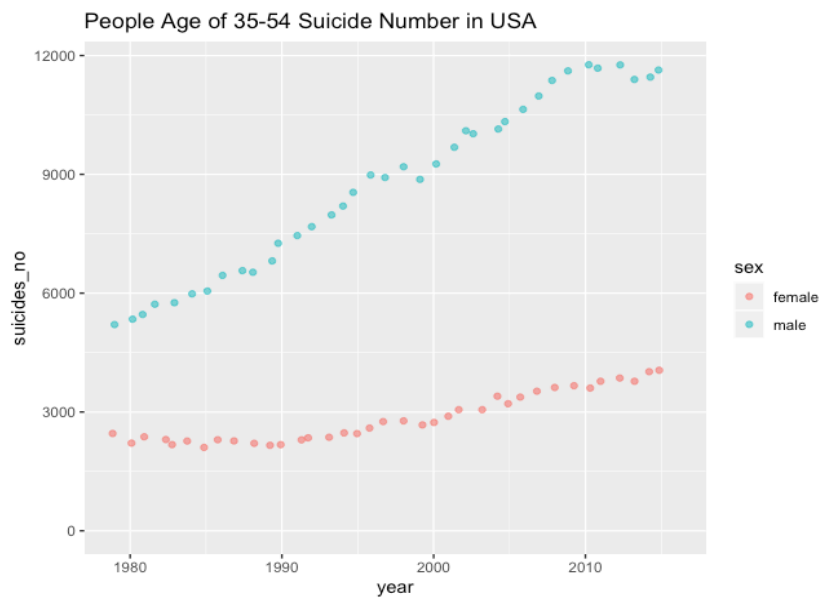


Fig.6

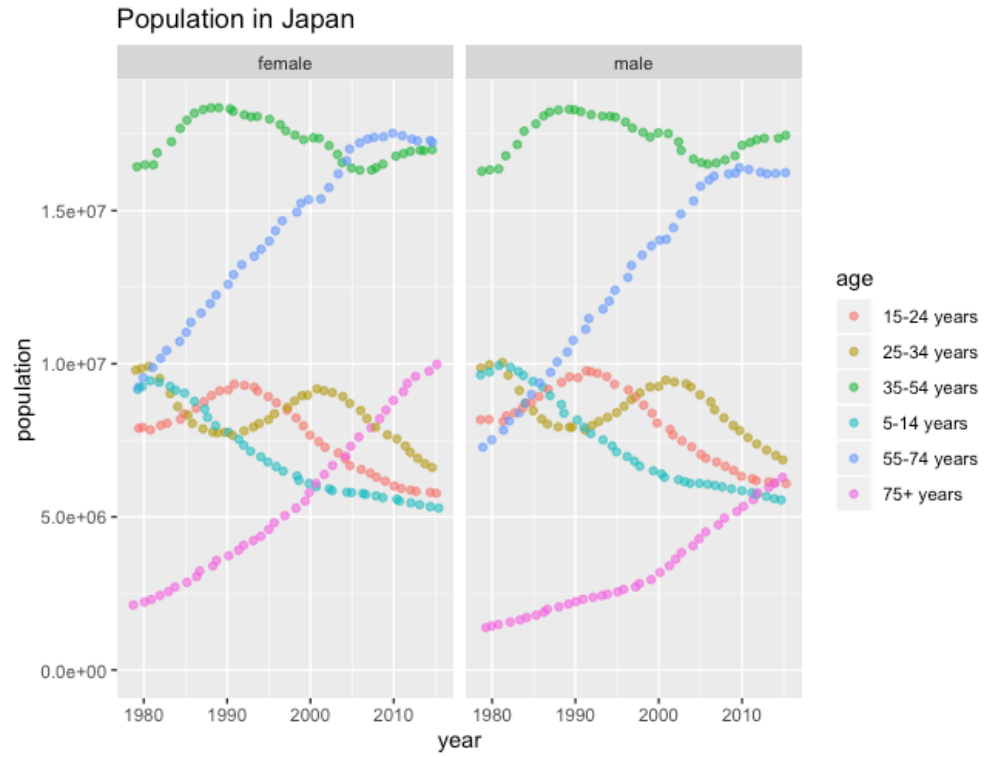


Fig. 7

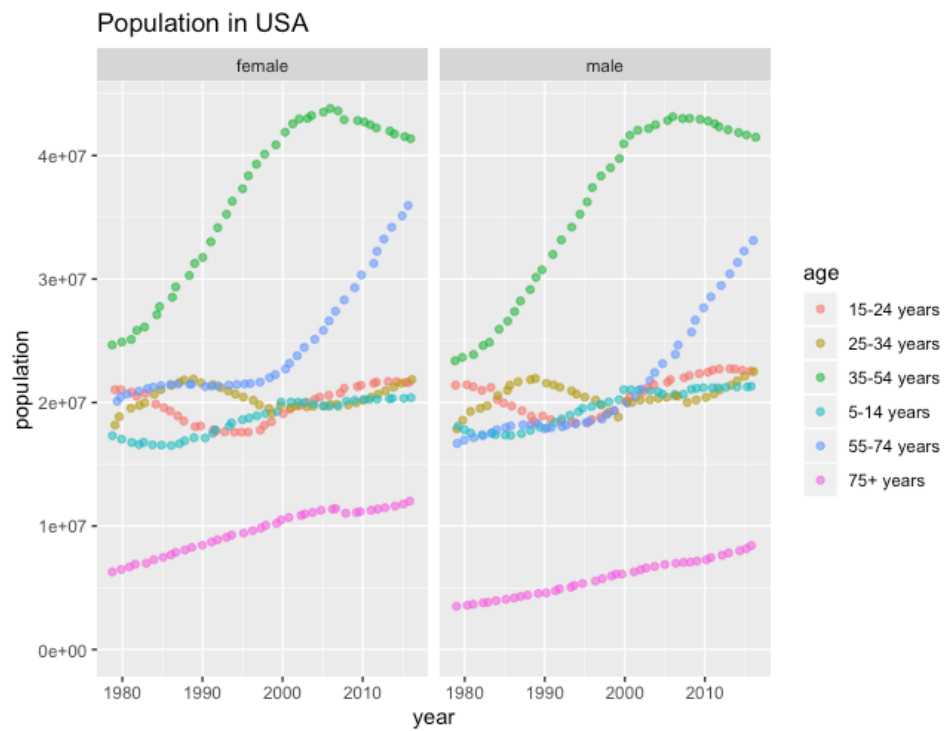


Fig. 8

There was no outlier in the dataset, but there were 12 missing values in the population of the USA.

Overall, there was a general increase in the suicide number every year, and people were more likely to suicide as ages grew. I compared to the suicide number of male and female, more males suicided than females. And compared two countries, suicide people in Japan were more than suicide people in the USA.

Appendix

```
library(lubridate)
```

```
library(dplyr)
```

```
library(tidyverse)
```

```
scd<-read.csv("Desktop/suicide.csv",header=TRUE,check.names =FALSE)
```

```
summary(scd)
```

```
#SCD and POP in USA
```

```
scd_USA <-scd %>% filter(country=="United States of America")
```

```
summary(scd_USA)
```

```
ggplot(scd_USA,aes(x=year,y=suicides_no,color=age))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+facet_grid(.~sex)+labs(title="Suicide Number in USA")
```

```
ggplot(scd_USA,aes(x=year,y=population,color=age))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+facet_grid(.~sex)+labs(title="Population in USA")
```

```
sum(is.na(scd_USA))
```

```
#Suicide Number in USA by Age
```

```
ggplot(scd_USA,aes(x=age,y=suicides_no))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+labs(title="Suicide Number in USA by Age")
```

```
#SCD/POP in USA
```

```
scd_USA_SpP<-scd_USA%>%mutate(ScdoverPop=suicides_no/population)
```

```
head(scd_USA_SpP)
```



```

ggplot(scd_USA_SpP,aes(x=age,y=ScdoverPop))+geom_jitter(alpha=.
6)+expand_limits(y=0)+labs(title="Suicide Number/Population in USA by Age")

#Mid-Age in USA

scd_USA_mid<-scd %>% filter(country=="United States of America") %>%
filter(age=="35-54 years")

head(scd_USA_mid)

sum(is.na(scd_USA_mid))

glimpse(scd_USA_mid)

tail(scd_USA_mid)

scd_USA_mid_f<-scd %>% filter(country=="United States of America") %>%
filter(age=="35-54 years") %>% filter(sex=="female")

summary(scd_USA_mid_f)

scd_USA_mid_m<-scd %>% filter(country=="United States of America") %>%
filter(age=="35-54 years") %>% filter(sex=="male")

summary(scd_USA_mid_m)

ggplot(scd_USA_mid,aes(x=year,y=suicides_no,color=sex))+geom_jitter(alpha=.
6)+expand_limits(y=0)+labs(title="People Age of 35-54 Suicide Number in USA")

ggplot(scd_USA_mid,aes(x=year,y=population,color=sex))+geom_jitter(alpha=.
6)+expand_limits(y=0)+labs(title="Population in USA")

#SCD and POP in Japan

scd_JP <- scd %>% filter(country=="Japan")

summary(scd_JP)

```

```
head(scd_JP)%>%arrange(desc(suicides_no))
```

```
ggplot(scd_JP,aes(x=year,y=suicides_no,color=age))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+facet_grid(.~sex)+labs(title="Suicide Number in Japan")
```

```
ggplot(scd_JP,aes(x=year,y=population,color=age))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+facet_grid(.~sex)+labs(title="Population in Japan")
```

#Suicide Number in USA by Age

```
head(scd_JP)
```

```
ggplot(scd_JP,aes(x=age,y=suicides_no))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+labs(title="Suicide Number in Japan by Age")
```

#SCD/POP in Japan

```
scd_JP_SpP<-scd_JP%>%mutate(ScdoverPop=suicides_no/population)
```

```
head(scd_JP_SpP)
```

```
ggplot(scd_JP_SpP,aes(x=age,y=ScdoverPop))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+labs(title="Suicide Number/Population in Japan by Age")
```

#Mid-Age in Japan

```
scd_JP_mid<-scd %>% filter(country=="Japan") %>% filter(age=="35-54 years")
```

```
head(scd_JP_mid)
```

```
ggplot(scd_JP_mid,aes(x=year,y=suicides_no,color=sex))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+labs(title="People Age of 35-54 Suicide Number in Japan")
```

```
ggplot(scd_JP_mid,aes(x=year,y=population,color=sex))+geom_jitter(alpha=.  
6)+expand_limits(y=0)+labs(title="Population in Japan")
```

```
scd_JP_mid_f<-scd %>% filter(country=="Japan") %>% filter(age=="35-54 years")  
%>% filter(sex=="female")
```

```
summary(scd_JP_mid_f)
```

```
scd_JP_mid_m<-scd %>% filter(country=="Japan") %>% filter(age=="35-54  
years") %>% filter(sex=="male")
```

```
summary(scd_JP_mid_m)
```

Reference

1. Szmail, WHO Suicide Statics, <https://www.kaggle.com/szamil/who-suicide-statistics>