

Final Project

Shamecca Marshall

2023-12-01

Load Libraries

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats 1.0.0      v readr 2.1.4
## v lubridate 1.9.3    v stringr 1.5.0
## v purrr 1.0.2       v tibble 3.2.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

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

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

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

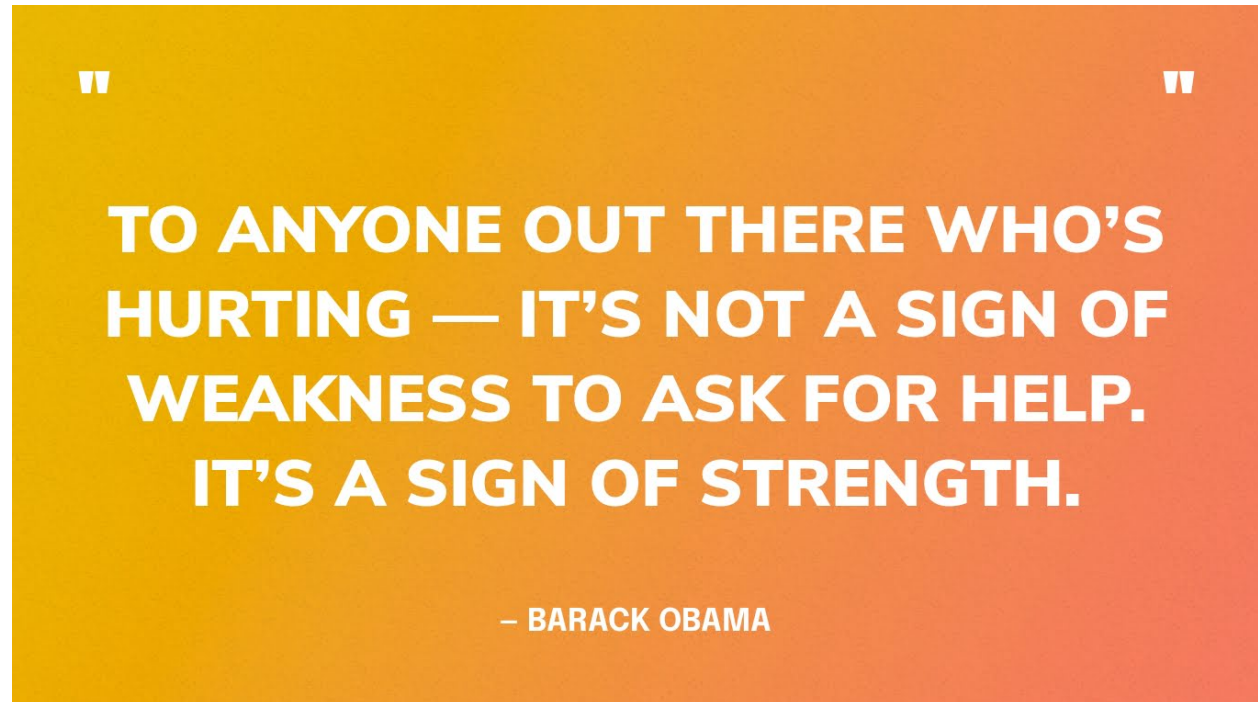
Introduction

Bringing awareness to suicide and suicidal thoughts is very dear to my heart. Growing up in Brooklyn, NY suicide was always forbidden fruit to. Granted, it could have been because the people I were around just did not take mental health seriously. It wasn't until after I became comfortable enough to admit that my depression left me in a suicidal state and I had thoughts of ending it all after the loss of a loved one, that I started to take mental health my seriously, seek help, and educate the people around me.

Suicide has not only affected me personally, but is one of the biggest problems in the world and has increased a significant amount around the world annually. There are a lot of factors that play a role in why people attempt and commit suicide such as mental health issues, suffering from a loss, and chronic illnesses just to name a few. Despite there being an ample amount of factors to why people contemplate or commit suicide, I will be taking few variables from the data set and I will check their relationship with suicide numbers throughout the world. I will be analyzing both men and women from mostly all age groups with I hope that this project will be able to shed light on suicide awareness and show people that it is okay to be vulnerable

and seek help. We should not have to take on all of our struggles alone, sometimes it is okay for the “strong friend” to ask for help when it is needed.

“Suicide is one of the leading causes of death among all American adults and rates are increasing in both men and women” (SSM Health).



Data

I found the data used for this project on Kaggle. This data set can be downloaded free of charge from <https://www.kaggle.com/data-sets/russellyates88/suicide-rates-overview-1985-to-2016>. This compiled data set pulled from four other data sets linked by time and place, and was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum.

Loading the data set

```
data<- read.csv('/Users/MECCA/Desktop/master suicide.csv')
```

```
summary(data)
```

```
##      country      year      sex      age
## Length:27820   Min.   :1985 Length:27820 Length:27820
## Class :character 1st Qu.:1995 Class :character Class :character
## Mode  :character Median :2002 Mode  :character Mode  :character
##                Mean   :2001
##                3rd Qu.:2008
##                Max.   :2016
##
## suicides_no      population      suicides.100k.population      Year
## Min.   :      0.0 Min.   :      278 Min.   :      0.00 Min.   :1985
```

```
## 1st Qu.: 3.0 1st Qu.: 97498 1st Qu.: 0.92 1st Qu.:1995
## Median : 25.0 Median : 430150 Median : 5.99 Median :2002
## Mean : 242.6 Mean : 1844794 Mean : 12.82 Mean :2001
## 3rd Qu.: 131.0 3rd Qu.: 1486143 3rd Qu.: 16.62 3rd Qu.:2008
## Max. :22338.0 Max. :43805214 Max. :224.97 Max. :2016
##
## HDI.for.year gdp_for_year.... gdp_per_capita.... generation
## Min. :0.483 Length:27820 Min. : 251 Length:27820
## 1st Qu.:0.713 Class :character 1st Qu.: 3447 Class :character
## Median :0.779 Mode :character Median : 9372 Mode :character
## Mean :0.777 Mean : 16866
## 3rd Qu.:0.855 3rd Qu.: 24874
## Max. :0.944 Max. :126352
## NA's :19456
```

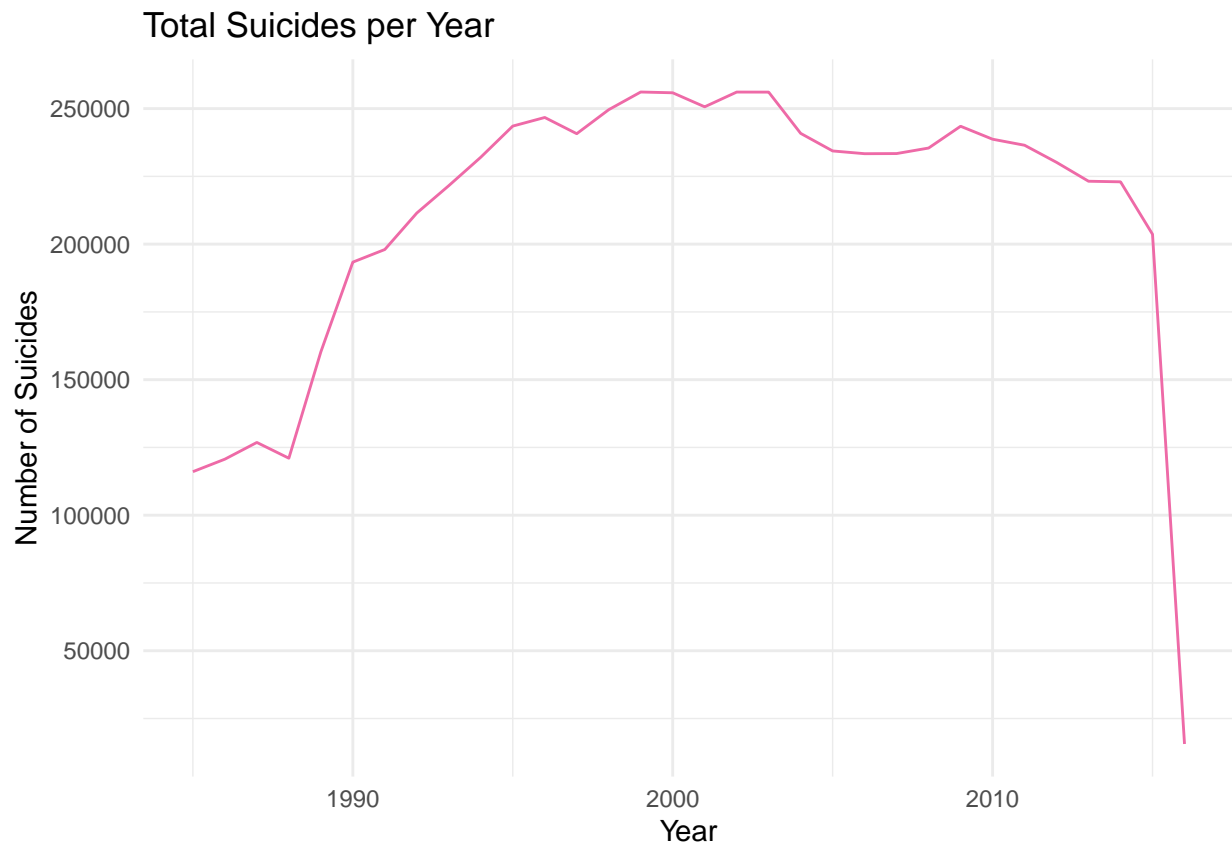
Grouping data by different categories to see the distribution of suicides

```
group_by_age <- data %>% group_by(age) %>% summarise(suicides_no = sum(suicides_no))
```

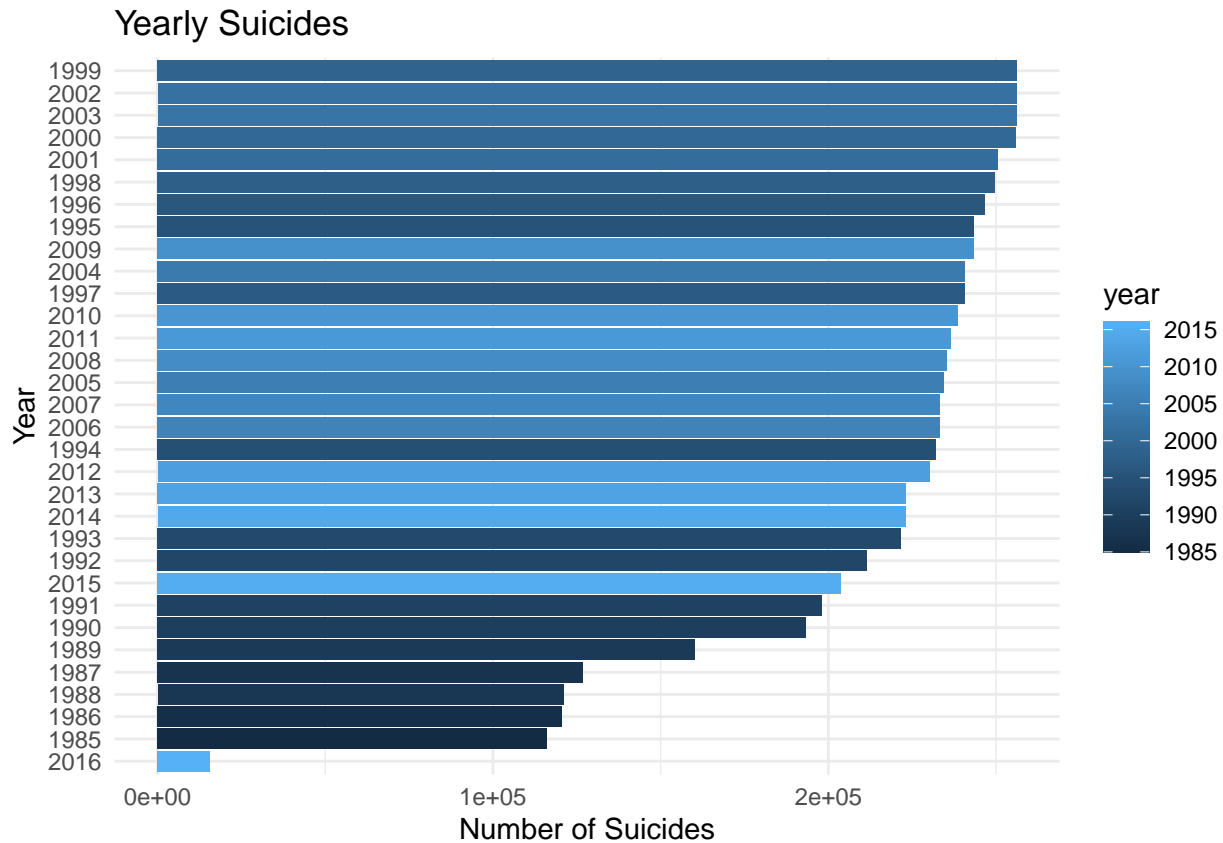
Research Question 1: Which year has the most suicides? Which year has the least suicides?

Plotting suicides over years

```
group_by_year <- data %>% group_by(year) %>% summarise(suicides_no = sum(suicides_no))
ggplot(group_by_year, aes(x = year, y = suicides_no)) +
  geom_line(color="hotpink2") +
  labs(title = 'Total Suicides per Year', x = 'Year', y = 'Number of Suicides') +
  theme_minimal()
```



```
ggplot(group_by_year, aes(x = reorder(year, suicides_no), y = suicides_no, fill = year)) +  
  geom_bar(stat = "identity") +  
  labs(title = 'Yearly Suicides', x = 'Year', y = 'Number of Suicides') +  
  theme_minimal() +  
  coord_flip()
```



```
year_most_suicides <- group_by_year[which.max(group_by_year$suicides_no), ]
year_least_suicides <- group_by_year[which.min(group_by_year$suicides_no), ]

list(most = year_most_suicides, least = year_least_suicides)
```

```
## $most
## # A tibble: 1 x 2
##   year suicides_no
##   <int>         <int>
## 1  1999         256119
##
## $least
## # A tibble: 1 x 2
##   year suicides_no
##   <int>         <int>
## 1  2016          15603
```

```
gender_suicides <- data %>%
  group_by(year, sex) %>%
  summarise(suicides_no = sum(suicides_no, na.rm = TRUE)) %>%
  spread(key = sex, value = suicides_no)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

From observing my line plot and bar graph, I noticed that suicide rates before 1990s were less than 150K annually. This low rate could be due to awareness of mental health and suicide in the 80s. After noticing this I decided to do more research and found out that this is accurate, as the research, “Suicide in the elderly” supports this claim”

“Female suicide rates have shown a similar overall decrease, falling by between 45 and 60% during the years 1983–1995 in the 45–84 age group.”

Creating a pivot wider

```
## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.
```

year	female	male
1985	32479	83584
1986	33852	86818
1987	35006	91836
1988	33015	88011
1989	41361	118883
1990	50118	143243
1991	49622	148398
1992	51567	159906
1993	51331	170234
1994	51532	180531
1995	54504	189040
1996	54583	192142
1997	54126	186619
1998	55631	193960
1999	56215	199904
2000	55254	200578
2001	52999	197653
2002	55549	200546
2003	55627	200452
2004	53232	187629
2005	52035	182340
2006	52039	181322
2007	53324	180084
2008	53973	181474
2009	54920	188567
2010	54222	184480
2011	54616	181868
2012	53011	177149
2013	51459	171740
2014	51556	171428
2015	47248	156392
2016	3504	12099

I created this pivot table to go more into detail. This allows us to see the total amount of women vs men that committed suicide per year.

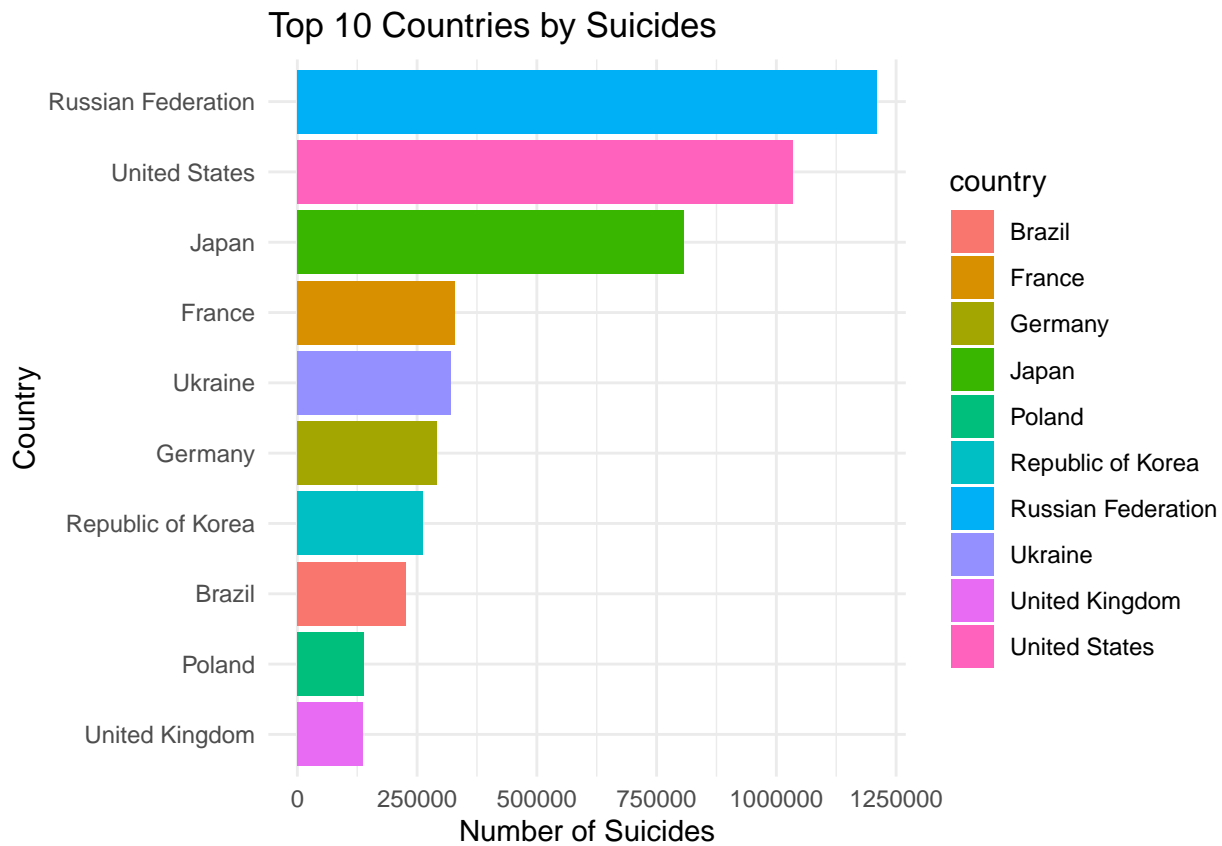
Research Question 2: Which country has the most Suicides? Which country has the least Suicides?

```
highest_suicide_country <- data %>% group_by(country) %>%  
  summarise(suicides_no = sum(suicides_no)) %>%
```

```
arrange(desc(suicides_no)) %>%
top_n(10, suicides_no)
```

Plotting by country (top 10)

```
ggplot(highest_suicide_country, aes(x = reorder(country, suicides_no), y = suicides_no, fill = country)) +
  geom_bar(stat = "identity") +
  labs(title = 'Top 10 Countries by Suicides', x = 'Country', y = 'Number of Suicides') +
  theme_minimal() +
  coord_flip()
```



```
groupby_country <- data %>% group_by(country) %>% summarise(suicides_no = sum(suicides_no))
```

Finding the grand total population per country.

```
group_by_population <- data %>% group_by(country) %>% summarise(population = sum(population))
print(group_by_population)
```

```
## # A tibble: 101 x 2
##   country      population
##   <chr>         <dbl>
## 1 Albania      62325467
## 2 Antigua and Barbuda 1990228
## 3 Argentina    1035985431
## 4 Armenia      77348173
```

```
## 5 Aruba 1259677
## 6 Australia 542377786
## 7 Austria 243853094
## 8 Azerbaijan 111790300
## 9 Bahamas 6557048
## 10 Bahrain 16753926
## # i 91 more rows
```

Grouping Country, Year, Sex, Age, and Population together to find which countries have the largest and lowest suicides, which year have the largest amount of suicides, the gender that committed the most amount of suicides, and the age group with the most suicides.

```
group_by_CYSAP <- data %>% group_by(country, sex, age, population, year) %>%
  summarise(suicides_no = sum(suicides_no)) %>%
  arrange(desc(suicides_no)) %>%
  top_n(101, suicides_no)
```

```
## `summarise()` has grouped output by 'country', 'sex', 'age', 'population'. You
## can override using the `.groups` argument.
```

```
country_most_suicides <- group_by_CYSAP[which.max(group_by_CYSAP$suicides_no), ]
country_least_suicides <- group_by_CYSAP[which.min(group_by_CYSAP$suicides_no), ]

list(most = country_most_suicides, least = country_least_suicides)
```

```
## $most
## # A tibble: 1 x 6
## # Groups:   country, sex, age, population [1]
##   country      sex   age      population  year suicides_no
##   <chr>      <chr> <chr>      <int> <int>      <int>
## 1 Russian Federation male 35-54 years 19044200 1994      22338
##
## $least
## # A tibble: 1 x 6
## # Groups:   country, sex, age, population [1]
##   country sex   age      population  year suicides_no
##   <chr>  <chr> <chr>      <int> <int>      <int>
## 1 Albania female 15-24 years 270003 2009          0
```

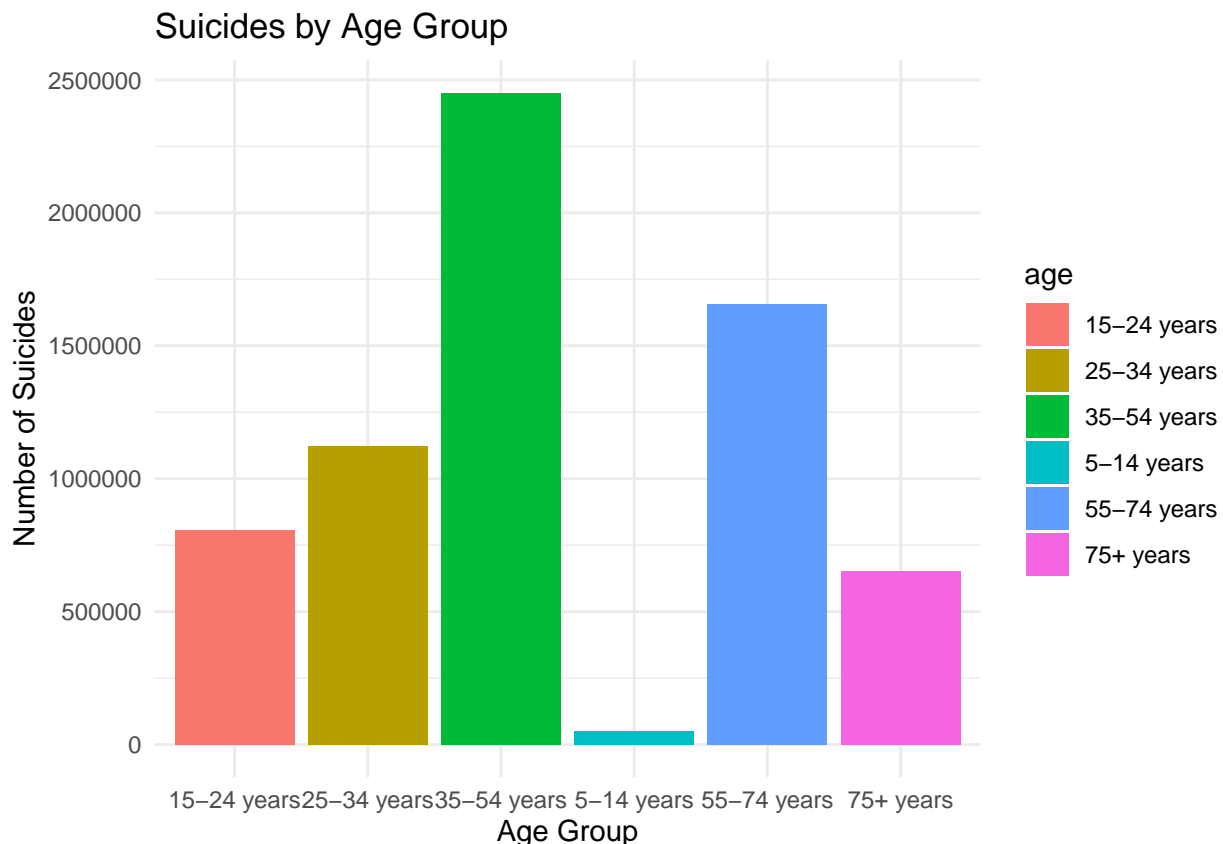
Both the graph & the min/max function above, confirms that Albania had the lowest suicide count, while Russian Federation, had the largest suicide count. A reason the Russian Federations may have a large suicide count may be because they have a very large population (Albania have a population of 2.8 million, while Russian Federation has of population of 144.3 million). It has been reported that Russian levels of alcohol consumption plays an immense role in it's large suicide count, but there is a lack of data to support this due to Soviet secrecy.

“Russian levels of alcohol consumption and suicide are among the highest in the world.”

Research Question 3: Are certain age groups more inclined to suicide?

Plot by age

```
ggplot(group_by_age, aes(x = age, y = suicides_no, fill = age)) +  
  geom_bar(stat = "identity") +  
  labs(title = 'Suicides by Age Group', x = 'Age Group', y = 'Number of Suicides') +  
  theme_minimal()
```



```
group_by_sex <- data %>% group_by(sex) %>% summarise(suicides_no = sum(suicides_no))
```

The bar graph shows that ages 35 through 54, have the highest suicide count. While ages 55 through 74 have the second highest suicide count. This high suicide rate in adults 35 and older can be due to the “U-Shape Happiness Curve. When people reach middle age they may review their earlier goals in the context of their achievements. For some, the realization of unmet aspirations or the perceived failure to have accomplished goals set as young adults could lead to a midlife low.

The U-shape of Happiness Across the Life Course: Expanding the Discussion

Research Question 3: Which gender is more likely to commit suicide?

```
group_by_sex <- data %>% group_by(sex) %>% summarise(suicides_no = sum(suicides_no))
```

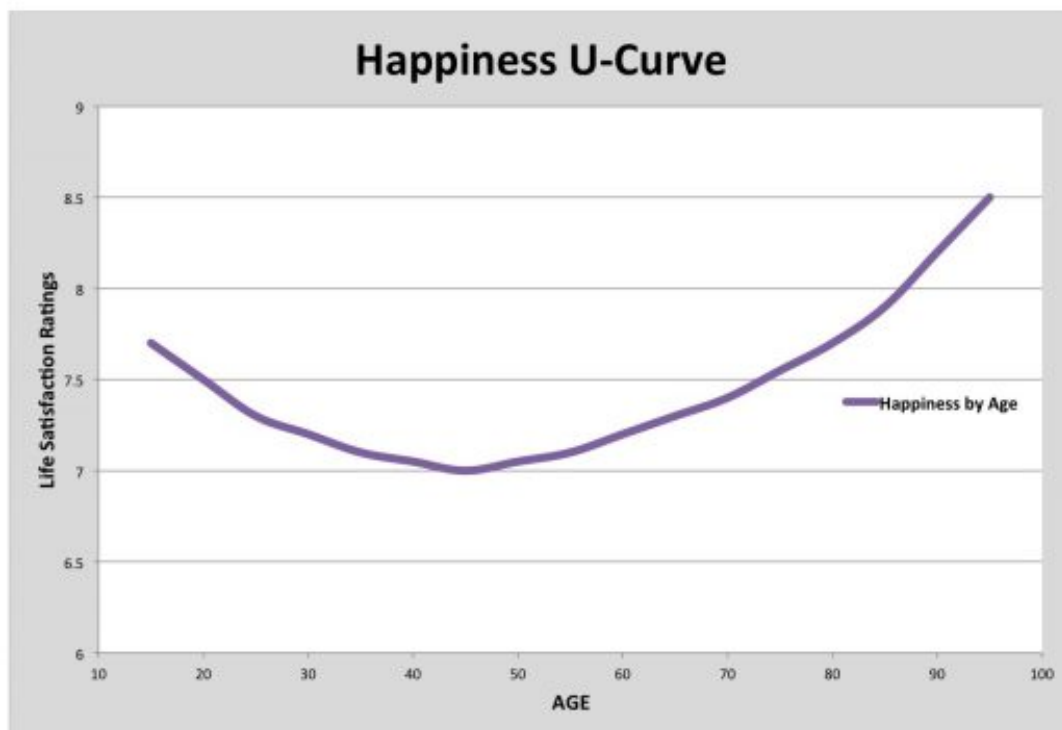
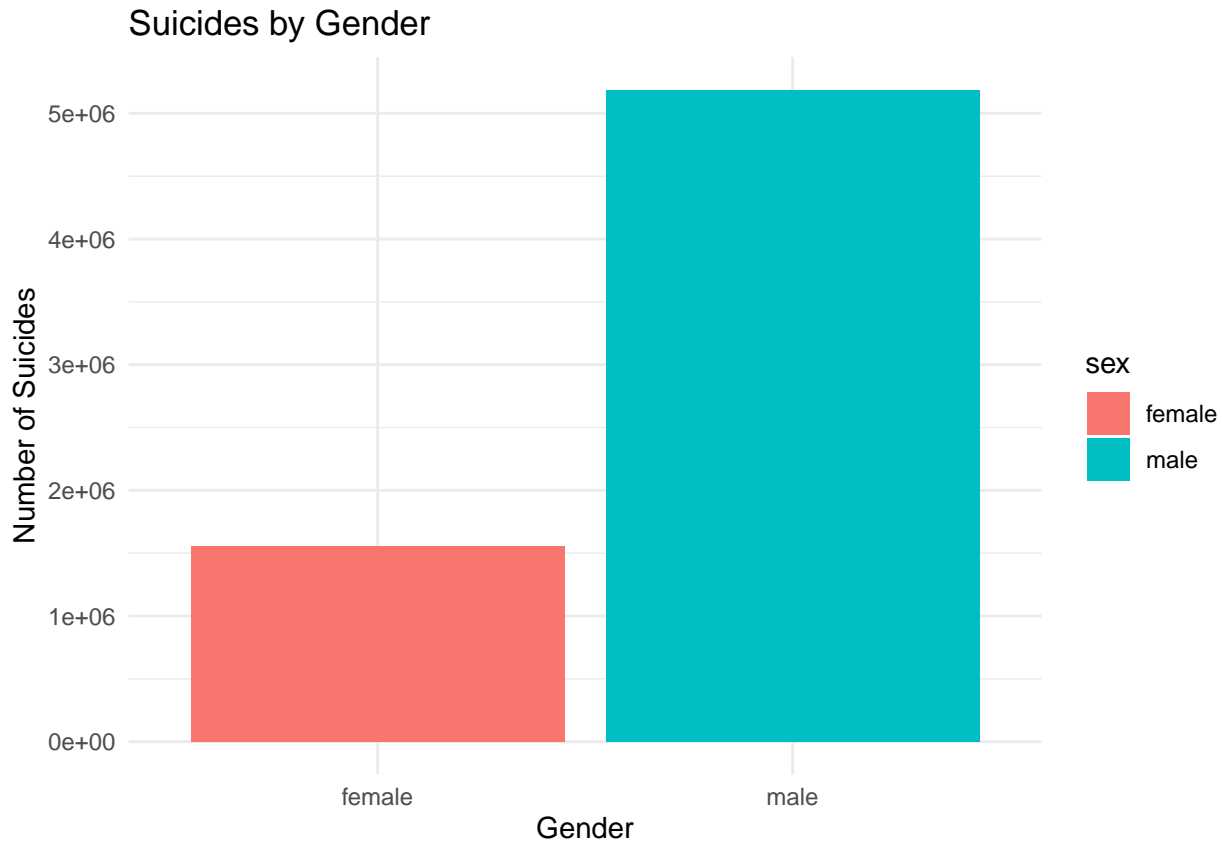


Figure 1: U-Shaped Happiness Curve

Plot by sex

```
ggplot(group_by_sex, aes(x = sex, y = suicides_no, fill = sex)) +  
  geom_bar(stat = "identity") +  
  labs(title = 'Suicides by Gender', x = 'Gender', y = 'Number of Suicides') +  
  theme_minimal()
```



From the above chart, it shows that men are more likely to commit suicide than women. Why is that? For years boys and men have been told that it is not okay to cry and showing emotions make them less manly. Despite both women and men dealing with depression, women are more likely than men to seek help for it. Men take strong value in independence and purposefulness, and they sometimes believe that admitting that they need help as a sign of weakness and avoid it. Meanwhile, despite women valuing their independence that are willing to consult friends and are more likely to accept help.

“Men seek help for mental health less often,” Harkavy-Friedman says. “It’s not that men don’t have the same issues as women – but they’re a little less likely to know they have whatever stresses or mental health conditions that are putting them at greater risk for suicide.”

Conclusion:

Despite suicide having a decrease before the 1990’s, it is now at an all time high. Suicide is something that should be talked about more often because if more people are aware that they are not alone, they would possibly be more comfortable reaching out for help before it is too late. It is evident that middle aged men are more likely to commit suicide and the difference between men and women suicide rates are pretty alarming. Mental health is something that should not be brushed off because it is a major predictor for suicide. If you know someone who is suffering please use the resources below below.

Where to get help:

988 Suicide and Crisis Lifeline

References:

<https://www.ssmhealth.com/blogs/ssm-health-matters/october-2019/middle-aged-men-more-likely-to-die-by-suicide> <https://www.cambridge.org/core/journals/advances-in-psychiatric-treatment/article/suicide-in-the-elderly/A4A9F7695DCA8D9B2796453FF166B8F3> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1642767/> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7529452/> <https://www.bbc.com/future/article/20190313-why-more-men-kill-themselves-than-women>