

An Analysis of the Cases of Suicide from 1985 - 2016

Paolo G. Hilado

Abstract

This study works on the data set of suicide from 1985 until 2016 retrieved from Kaggle. It is guided by research questions pertaining to determining the reported cases of suicide and suicide per 100k population among males and females for each age group, visualizing the trend in the median recorded cases of suicide and median suicide per 100k population among Males and Females per age group and observing the correlation pattern and coefficient of the aforementioned variables when associated with GDP per capita (USD).

In light of the research questions of this study, the median is used to determine reported cases of suicide and suicide per 100k population together with the median absolute deviation as estimate of dispersion, Bar graphs are used to provide a comparison of these medians among males and females for each age group, the line graph was used in terms of visualizing the trend from 1985 to 2016 and a scatterplot has been utilized to check for correlational pattern which was reinforced by the Kendall's Tau coefficient.

Results show that Males have higher median reported cases of suicide and median suicide per 100k population among Females across all age groups. Starting 2005, an increasing trend in median suicide cases among males and females can be observed for age groups involving 35-54 and 55-74 years old; this, however, is not evident in terms of median per 100k population thus the increasing trend in reported case may probably due to a steady increase in human population. Lastly, both reported cases of suicide and suicide per 100k population have a weak positive relationship with GDP per capita (USD).

Motivation

Suicide is a tragedy that creates trauma among individuals left behind, communities and even at a macro-level. The World Health Organization (WHO, 2019) describes it as a serious public problem that exists as a global phenomenon which affects not only high income countries but more so among low and middle income countries.

Often, people who have suicidal ideation become withdrawn, show despair, engage in self-destructive activities, focus on death and exhibit mood swings (WebMD, 2019). At times, we tend to overlook at these tell-tale signs and eventually realize that everything is too late. During that time we loose someone dear to us, there is no coming back.

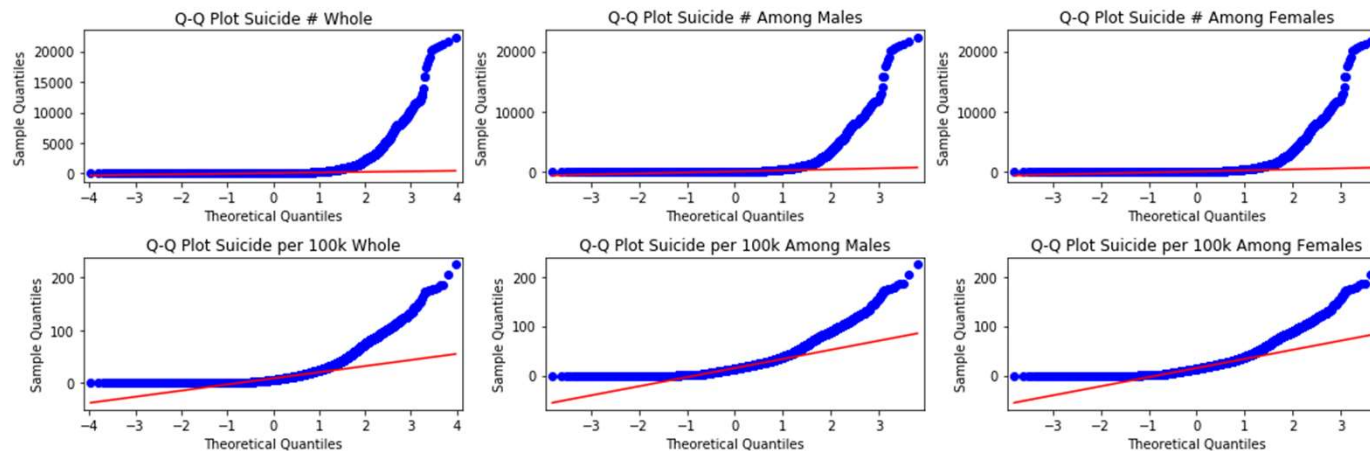
This study has been conducted in order to highlight this serious problem. Through its findings readers would be further informed about suicide with the hopes that it gets the attention it deserves thus curb the damages it brings knowing that it is preventable if provided with evidenced-based and timely interventions.

Dataset(s)

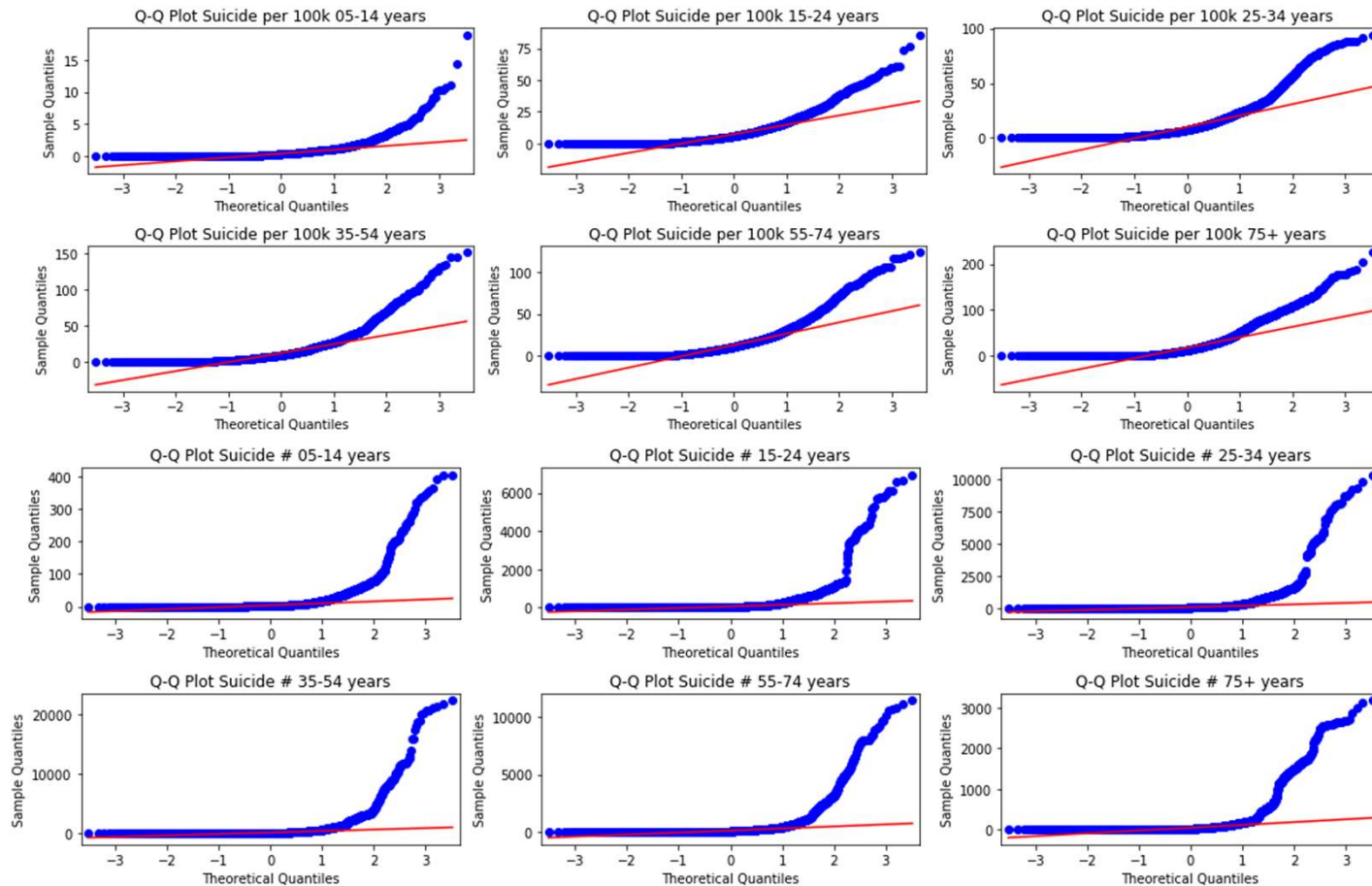
The dataset works on the number of suicides over the years of 1985 to 2016. This study works a dataset retrieved from Kaggle with the link (<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>). It combines data from varied sources such as the United Nations, Development Program, World Bank, World Health Organization and Suicide in the 21st Century Data set by Szamil. The raw data set has 12 Variables (columns) and 27,820 rows (2.58 mb). Given the research questions of this study, the raw data set was cleaned to only include 7 columns and same rows involving the necessary variables such as the number of suicides, suicides per 100k population, GDP per capita (in USD), age (05-14 years, 15-24 years, 25-34 years, 35-54 years, 55-74 years and 75+ years), sex (male and female), year (1985 – 2016) and country (101 unique entries). Furthermore, the type of the variables are a combination of integers, objects and float.

Data Preparation and Cleaning

Columns were given with appropriate labels to facilitate convenience in data analysis at a later time. Missing Values were also removed. Upon exploration, it turns out that there were fewer observations on the Age Group involving 05-14 years. This Age Group does not have data for year 2016. It was ascertained that there are no Missing Values so this indicates that the encoding of the data for this age group simply stopped at 2015. More of exploratory data analysis, D'Agostino Normality testing was conducted on variables such as Suicide Number and Suicide per 100k when grouped by Sex and Age. Probability values for the aforementioned all turned out to be less than 0.05 thus these data fail to assume normal distribution. This is reinforced by the results of the Quantile-Quantile Plots generated (these are presented below and on the following slides). As such, the Median has been preferred as a measure for central tendency and the Median Absolute Deviation for Dispersion.



Data Preparation and Cleaning – QQ Plots Continuation



Research Question(s)

1. Which among Males and Females have higher median recorded cases of suicide and median suicide per 100k population when grouped by age, specifically, 05-14 year, 15-24 years, 25-34 years, 35-54 years, 55-74 years and 75+ years?
2. What is the trend in the median recorded cases of suicide and median suicide per 100k population among Males and Females per age group from 1985 until 2016?
3. Can a scatterplot present a reasonable correlation pattern of the number of recorded cases of suicide and suicide per 100k population when associated with GDP per capita (USD)?

Methods

The following are the methods used to analyze the data guided by the research questions.

In terms of presenting the median recorded cases of suicide and median suicide per 100k population among Males and Females for each age group, a table for Descriptives and the bar graph was used. The Descriptives tables includes the count, minimum value observed, quartiles (2nd quartile is the median), median absolute deviation and maximum value observed. The bar graph provided the visualization which compares the median reported cases of suicide and median suicide per 100k population between Males and Females for each age group.

As to visualizing the trend from 1985 to 2016, the line graph was used as it is most appropriate for comparison of parameters over time. Through it, the rise and fall in median reported cases of suicide and suicide per 100k population can easily be observed involving males and females per age group.

In terms of correlation, scatterplot was used to observe for patterns between variables. Considering that the data set is skewed, the Kendall's Tau Coefficient was used to check the direction and strength of the correlation and to assess whether observations of the scatterplot were reasonable.

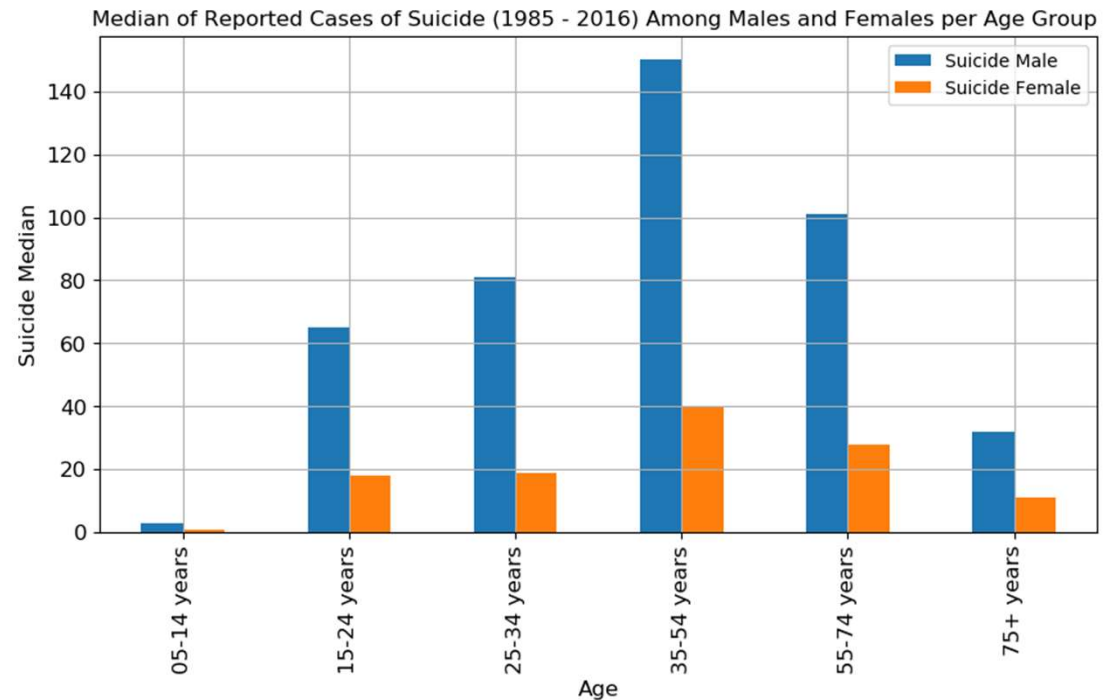
Findings – Median Reported Cases of Suicide among Males and Females per Age Group

Males

	count	min	25%	50%	MAD	75%	max
Age Group							
05-14 years	2305.0	0.0	0.0	3.0	61.0	12.0	404.0
15-24 years	2321.0	0.0	15.0	65.0	77.0	229.0	6945.0
25-34 years	2321.0	0.0	20.0	81.0	147.0	310.0	10310.0
35-54 years	2321.0	0.0	31.0	150.0	3.0	585.0	22338.0
55-74 years	2321.0	0.0	13.0	101.0	101.0	370.0	11431.0
75+ years	2321.0	0.0	3.0	32.0	32.0	151.0	3171.0

Females

	count	min	25%	50%	MAD	75%	max
Age Group							
05-14 years	2305.0	0.0	0.0	1.0	17.0	6.0	158.0
15-24 years	2321.0	0.0	4.0	18.0	18.0	72.0	1132.0
25-34 years	2321.0	0.0	4.0	19.0	40.0	74.0	1444.0
35-54 years	2321.0	0.0	6.0	40.0	1.0	177.0	4053.0
55-74 years	2321.0	0.0	3.0	28.0	28.0	143.0	3499.0
75+ years	2321.0	0.0	1.0	11.0	11.0	58.0	2013.0

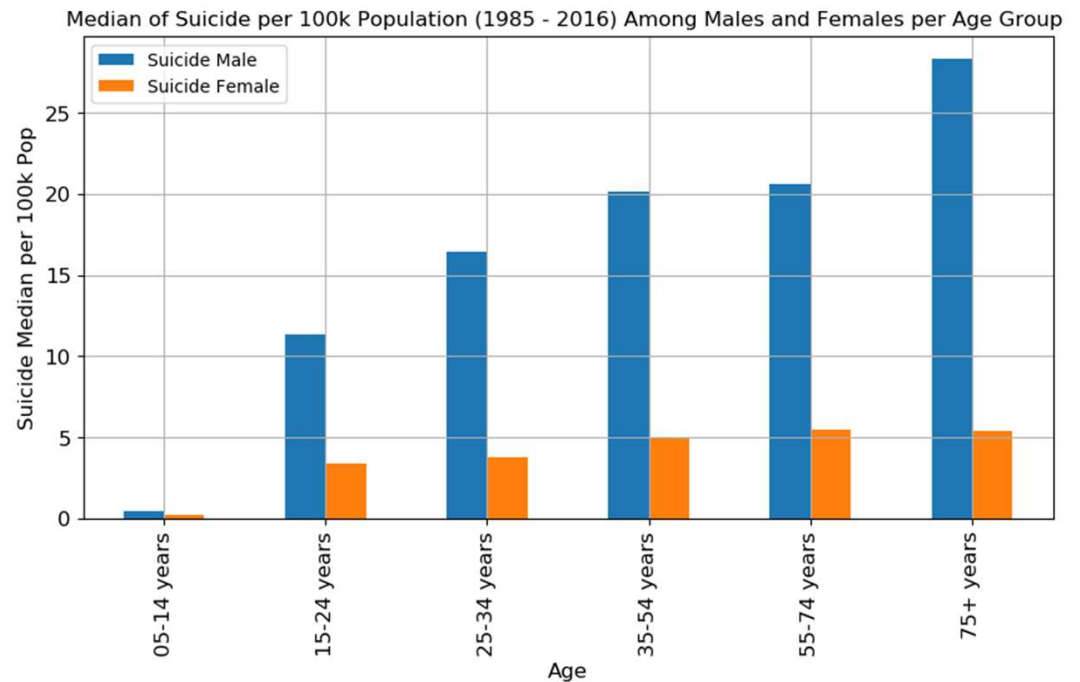


It can be observed that Males have higher median reported cases of suicide compared to females across all age groups. Both males and females have higher median reported cases of suicide at age 35 – 54 years with 150 and 40 respectively. It is also interesting to note that it is the same age range that has consistency in values given its low median absolute deviation.

Findings – Median Suicide per 100k Population among Males and Females per Age Group

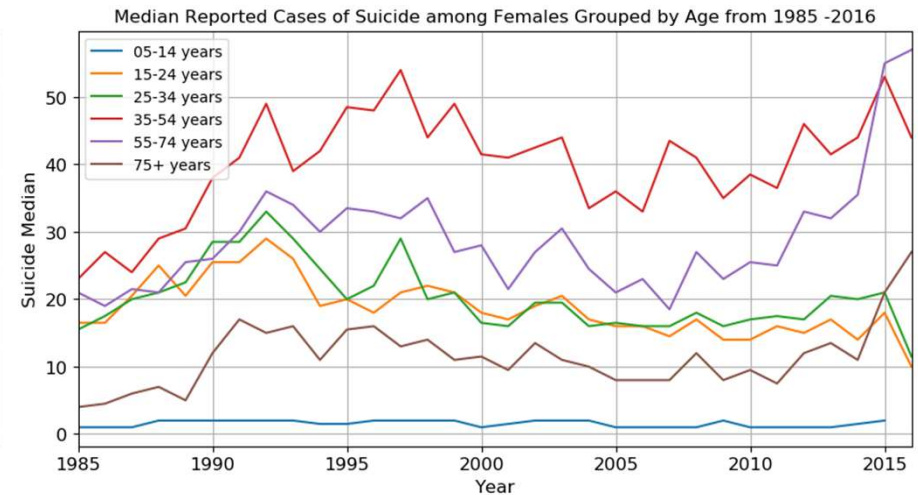
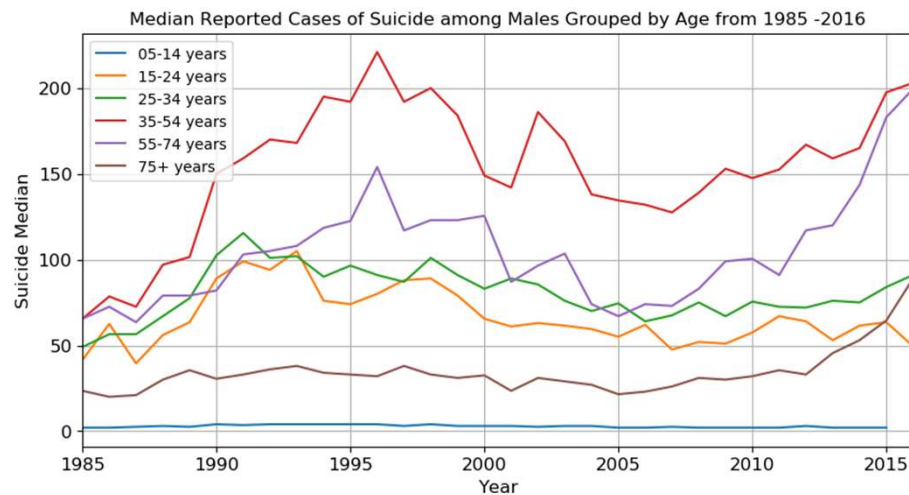
Males	count	min	25%	50%	MAD	75%	max
Age Group							
05-14 years	2305.0	0.0	0.00	0.46	6.04	1.02	18.92
15-24 years	2321.0	0.0	6.14	11.39	8.38	18.50	85.59
25-34 years	2321.0	0.0	9.08	16.45	10.59	26.04	94.18
35-54 years	2321.0	0.0	9.57	20.16	0.46	30.68	151.33
55-74 years	2321.0	0.0	9.68	20.61	12.15	35.01	123.83
75+ years	2321.0	0.0	10.39	28.32	21.86	59.00	224.97

Females	count	min	25%	50%	MAD	75%	max
Age Group							
05-14 years	2305.0	0.0	0.00	0.23	2.06	0.58	14.33
15-24 years	2321.0	0.0	1.53	3.40	2.29	5.76	57.08
25-34 years	2321.0	0.0	1.79	3.81	3.41	6.42	33.60
35-54 years	2321.0	0.0	1.76	4.91	0.23	8.80	30.97
55-74 years	2321.0	0.0	1.40	5.48	4.42	10.88	55.25
75+ years	2321.0	0.0	0.54	5.37	5.37	14.83	133.42



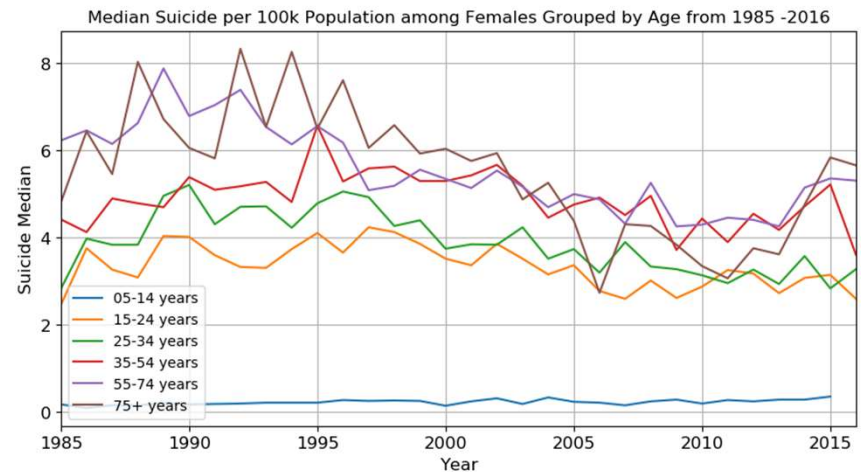
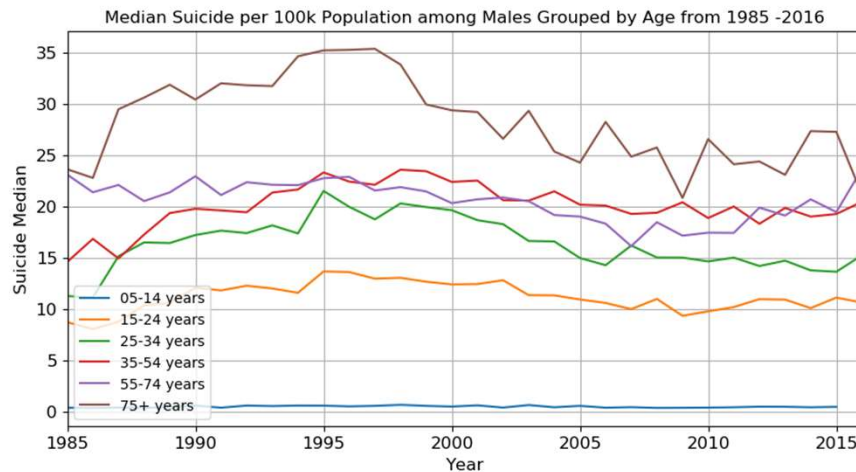
In terms of median suicide per 100k population, this time it can be observed that among males it is higher for age 75+ years and, among females, higher for 55-74 years as reflected in the Descriptives Table. Looking at the Bar Graph, we can still say that males tend to have higher median suicide per 100k population compared to females across all the age groups.

Findings – Trend in Median Reported Cases of Suicide among Males and Females per Age Group from 1985 - 2016



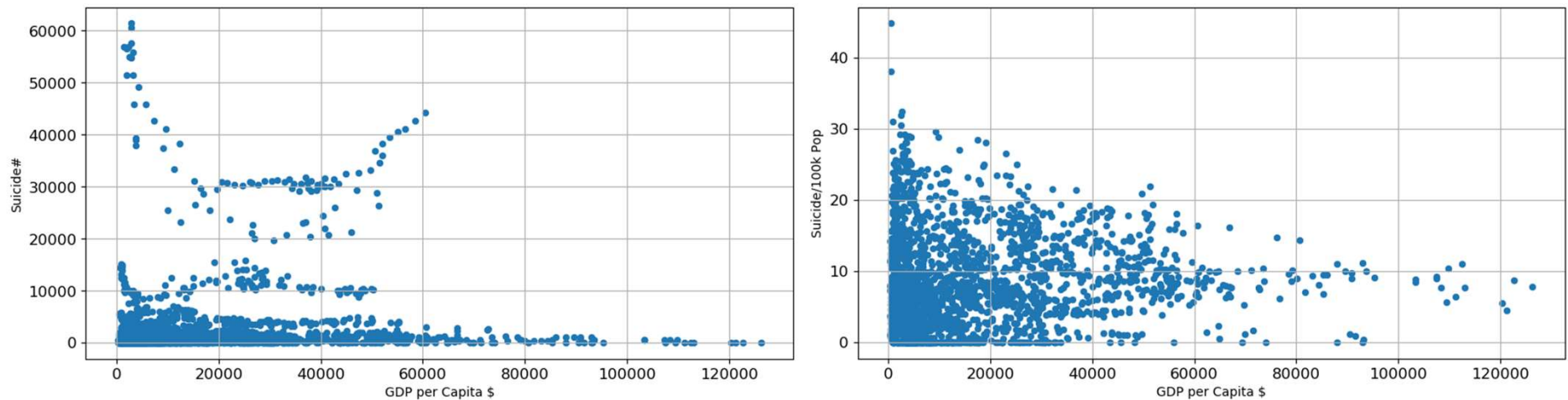
The trend presented in the line graphs shown above showcase a commonality among the Median Reported Cases of Suicide Among Males and Females when grouped by Age. It can be observed, that both males and females at age groups 35 – 54 years and 55 – 74 years had a decline in the aforementioned at around 1995 to 2005, however, an uptrend is evident at after 2005.

Findings – Trend in Median Suicide per 100k Population among Males and Females per Age Group from 1985 - 2016



Contrary to the observation on the median reported cases for suicide, here we can observe that in terms of median suicide per 100k population it is usually the 75+ years age group that are high. For males, the aforesaid age group continues to be at the top through the years except for 2016 where it was superseded by the 55 – 74 years age group. For the females, a rise and fall can be observed for this age group but it is ascertained that in the later years around 2015 – 2016, it is the one that garnered the top spot. An interesting thing to point out is why the trend of median reported cases and median per 100k population for suicide is different. One probable reason is that, in time, the population among age groups 35 – 54 and 55 – 74 years () may have continued to increase which led to more people and more cases but is curbed when considered for every 100k population.

Findings – Recorded cases of Suicide and Suicide per 100k population when associated with GDP per capita (USD)



Observing the scatterplot on cases of suicide and GDP per Capita (USD), countries with GDP over 80,000 tend to have low reported cases of suicide and vice versa. In terms of suicide per 100k population and GDP per capita, those with more than 60,000 GDP tend to have less than 20 suicide per 100k population. Kendall's Tau Coefficient informs us that Suicide Cases and GDP per Capita have a very weak correlation (0.091) which is same for Suicide per 100k Population and GDP per Capita (0.086). It would be unwise to make sense of its direction given the very weak correlation.

Limitations

One limitation of the data is that the entries for age groups 05 – 14 years ended at 2015. As such, upon a closer look at the line graph, it can be observed that the its line does not continue to 2016 unlike other age groups. However, considering that the median reported case of suicide and suicide per 100k for this age group is low compared to others, it may not be a big of an issue.

Conclusions

Given the research findings, it can be concluded that Males tend to have higher tendencies for suicide compared to females across all age groups. Starting from 1995 up to until 2005, there seems to be a downtrend in the median suicide reported cases for both males and females in the age groups 35 – 54 and 55 – 74 years. However, after 2005, an uptrend can be observed in the aforementioned involving the same age groups. The 75+ years age group tend to stand out in terms of suicide per 100k population for both males and females. Countries with higher GDP per capita seem to have lower suicide reported cases but with reservation considering that the correlation is very weak.

Acknowledgements

Acknowledgement is given to the following organizations and individuals who have made the data available for use:

Kaggle. Suicide Rates Overview from 1985 to 2016. Retrieved from <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>

United Nations Development Program. (2018). Human development index (HDI). Retrieved from <http://hdr.undp.org/en/indicators/137506>

World Bank. (2018). World development indicators: GDP (current US\$) by country:1985 to 2016. Retrieved from <http://databank.worldbank.org/data/source/world-development-indicators#>

Szamil. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from <https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook>

World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

References

World Health Organization (2019). Key Facts on Suicide, Retrieved from <https://www.who.int/news-room/fact-sheets/detail/suicide>

WebMD (2019). Suicide Warning Signs: What to Watch for and Do, Retrieved from <https://www.webmd.com/depression/guide/depression-recognizing-signs-of-suicide#1>