Final Report of group\_179

Table of Contents

[Introduction 2](#_Toc61034353)

[Visualization: 3](#_Toc61034354)

[Analysis: 5](#_Toc61034355)

[Conclusion: 5](#_Toc61034356)

[List of Tables: 6](#_Toc61034357)

[List of Figures: 6](#_Toc61034358)

[Bibliography: 6](#_Toc61034359)

## 

## Introduction

The compiled dataset pulled from four other datasets 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.

Table 1 : Descriptive Statistics of our Dataset

|  |
| --- |
| country year sex age suicides no  Length:27820 Min. :1985 Length:27820 Length:27820 Min. : 0  Class :character 1st Qu.:1995 Class :character Class :character 1st Qu.: 3  Mode :character Median :2002 Mode :character Mode :character Median : 25  Mean :2001 Mean : 243  3rd Qu.:2008 3rd Qu.: 131  Max. :2016 Max. :22338    population suicides.100k.pop country.year HDI.for.year gdp\_for\_year....  Min. :2.8e+02 Min. : 0.00 Length:27820 Min. :0.48 Length:27820  1st Qu.:9.7e+04 1st Qu.: 0.92 Class :character 1st Qu.:0.71 Class: character  Median :4.3e+05 Median : 5.99 Mode :character Median :0.78 Mode :character  Mean :1.8e+06 Mean : 12.82 Mean :0.78  3rd Qu.:1.5e+06 3rd Qu.: 16.62 3rd Qu.:0.85  Max. :4.4e+07 Max. :224.97 Max. :0.94  NA's :19456  gdp\_per\_capita.... generation  Min. : 251 Length:27820  1st Qu.: 3447 Class :character  Median : 9372 Mode :character  Mean : 16866  3rd Qu.: 24874  Max. :126352 |

The dataset consists of the number of suicides each year in a given country, split by sex and age groups, plus the total population in according group, thus allowing us to calculate the **suicides\_rate per 100k people between male and female**.

We have formed our Research Question based on Suicides.100k.pop (Ordinal/Interval) and sex (Nominal).

Our Research Question, Null Hypothesis and Alternative Hypothesis are stated below:

**Research Question**: Is there a difference in the mean of the number of suicides per 100K Population between male and female?

**Null hypothesis**: There is no difference in the mean of the suicide numbers per 100K Population between male and female.

**Alternative hypothesis**: There is a difference in the mean of the suicide numbers per 100K Population between male and female.

## 

## Visualization:

Now we will look into the Distributions of our data by drawing the Box plot, Bar plot, Stacked Bar plot and Histogram.

Figure 1: Gender Boxplot

By Plotting Box plot it is clear that there is a significant difference in the median between male and female suicides rate.

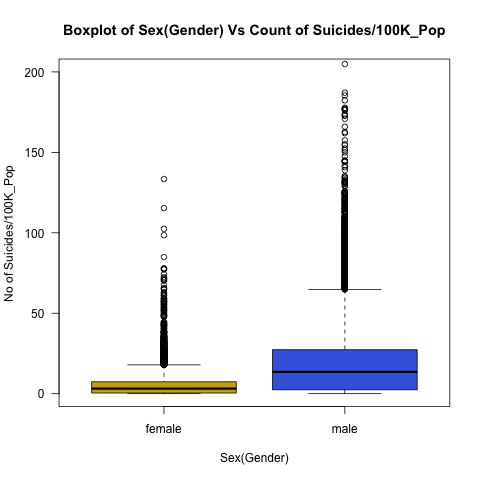


Figure 2: Gender Bar plot

The below Bar plot shows the summary statistics of male and female suicide rates.

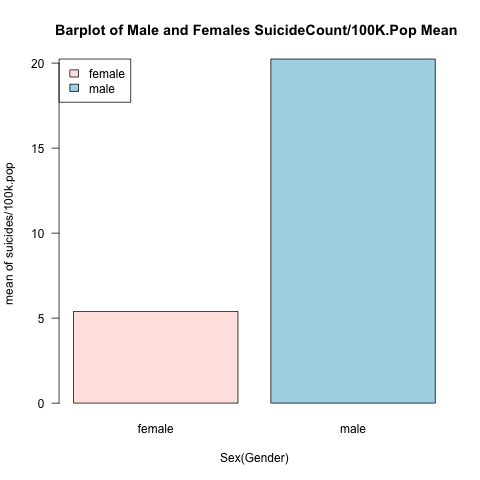


Figure 3: Gender Stacked Bar plot

Stacked bar plot shows based on Age, The frequency proportions of suicide rates of male and female.

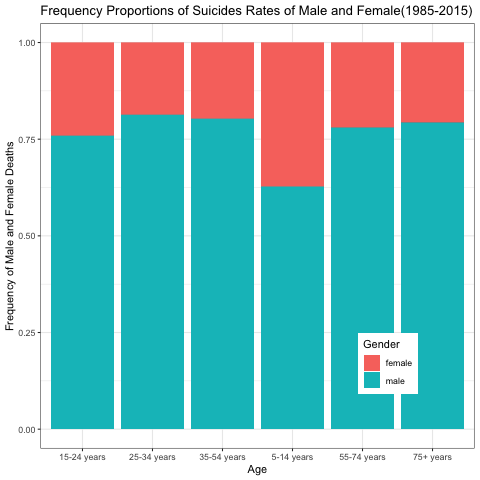
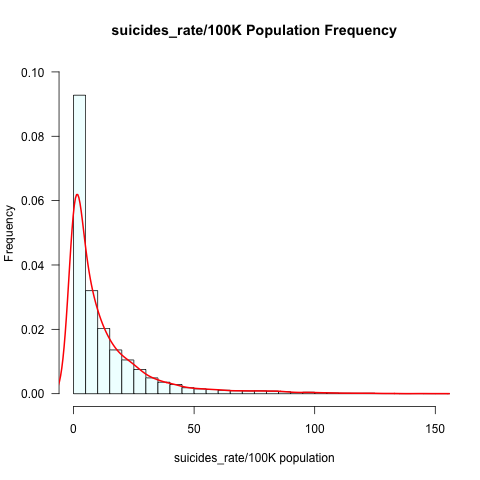


Figure 4:Histogram of Suicides\_rate

The Histogram is the most commonly used graph to show the frequency distributions. A frequency distribution shows how often each different value in a set of data occurs. The below histogram shows the graph is left Skewed.



## 

## Analysis:

When you perform a hypothesis test in statistics, a p value helps you determine the significance of your results. Hypothesis tests are used to test the validity of a claim that is made about a population. This claim that's on trial, in essence, is called the null hypothesis.

The alternative hypothesis is the one you would believe if the null hypothesis is concluded to be untrue. The evidence in the trial is your data and the statistics that go along with it. All hypothesis tests ultimately use a p-value to weigh the strength of the evidence.

The p-value is a number between 0 and 1 and interpreted in the following way:

• A small p-value (typically **≤** 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis.

• A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis. p-values very close to the cutoff (0.05) are considered to be marginal (could go either way).

The mean value of suicides/100K is 12.92. We have performed the Student’s T-test to find out if we could reject the Null Hypothesis or not based on the p-value. After performing T-test we got the p-value as 2.2e-16 i.e., (0.000000000000000022) which is less than 0.05.

## Conclusion:

With this we can conclude that the Alternative Hypothesis is **True** at 95% confidence level (or whatever level we have chosen) and It is clear that we can confidently **reject our Null Hypothesis** based on our analysis. It appears that there is a significant difference in the mean of the suicide numbers per 100K Population between male and female.

## List of Tables:

[Table 1 : Descriptive Statistics of our Dataset 2](#_Toc60943850)

## List of Figures:

[Figure 1: Gender Boxplot 3](#_Toc60949205)

[Figure 2: Gender Barplot 3](#_Toc60949206)

[Figure 3: Gender Stacked Barplot 4](#_Toc60949207)

[Figure 4:Histogram of Suicides\_rate 4](#_Toc60949208)

## Bibliography:

We used the Dataset suicide rates from Kaggle Dataset (Suicide Rates Overview 1985 to 2016)

Suicide Rates Overview 1985 to 2016. “Suicide Rates Overview 1985 to 2016,” December 1, 2018.

<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>.

Miron, Oren, Kun-Hsing Yu, Rachel Wilf-Miron, and Isaac S. Kohane. “Suicide Rates Among Adolescents and Young Adults in the United States, 2000-2017.” *JAMA* 321, no. 23 (June 18, 2019): 2362. https://doi.org/10.1001/jama.2019.5054.

Bould, Helen, Becky Mars, Paul Moran, Lucy Biddle, and David Gunnell. “Rising Suicide Rates among Adolescents in England and Wales.” *The Lancet* 394, no. 10193 (July 2019): 116–17. <https://doi.org/10.1016/S0140-6736(19)31102-X>.

Sher, Leo. “The Impact of the COVID-19 Pandemic on Suicide Rates.” *QJM: An International Journal of Medicine* 113, no. 10 (October 1, 2020): 707–12. <https://doi.org/10.1093/qjmed/hcaa202>.

Leske, Stuart, Kairi Kõlves, David Crompton, Ella Arensman, and Diego de Leo. “Real-Time Suicide Mortality Data from Police Reports in Queensland, Australia, during the COVID-19 Pandemic: An Interrupted Time-Series Analysis.” *The Lancet Psychiatry* 8, no. 1 (January 2021): 58–63. <https://doi.org/10.1016/S2215-0366(20)30435-1>.

Stack, S. “Media Coverage as a Risk Factor in Suicide.” *Journal of Epidemiology & Community Health* 57, no. 4 (April 1, 2003): 238–40. <https://doi.org/10.1136/jech.57.4.238>.