A black and white text on a black background

Description automatically generated with low confidence

UNIVERSITY OF HERTFORDSHIRE

School of Physics, Engineering Computer Science

**7COM1079-0901-2024 - Team Research and Development Project**

**Jan -2025**

**Is there a Correlation Between Number of Suicides and Age of 18-95?**

Group ID: A183

Dataset Number: DS244

Prepared by

[Lakshmi Priya Pattem, 23079231]

Group Member(s) -Id

[Sai Rohith Kotaru, 23097945]

[Balaji Jada, 23035976]

[Eswar Reddy Amudam, 23087063]

[Sai Kumar Medeipelli, 23095723]

Table of Contents

[1. Introduction 4](#_Toc186538678)

[1.1. Problem Statement and Research Motivation 4](#_Toc186538679)

[1.2. The Data Set 4](#_Toc186538680)

[1.3. Research Question 4](#_Toc186538681)

[1.4. Null Hypothesis and Alternative Hypothesis 5](#_Toc186538682)

[2. Background Research 5](#_Toc186538683)

[2.1. Research Papers 5](#_Toc186538684)

[2.2. Why RQ Is of Interest 6](#_Toc186538685)

[3. Visualization 7](#_Toc186538686)

[3.1. Appropriate Plot for the RQ Output of an R Script 7](#_Toc186538687)

[3.2. Additional Information Relating to Understanding the Data 7](#_Toc186538688)

[3.3. Useful Information for the Data Understanding 8](#_Toc186538689)

[4. Analysis 8](#_Toc186538690)

[4.1. Statistical Test Used to Test the Hypotheses and Output 8](#_Toc186538691)

[4.2. The Null Hypothesis Is Rejected / Not Rejected Based on the p-value 8](#_Toc186538692)

[5. Evaluation – Group’s Experience at 7COM1079 9](#_Toc186538693)

[5.1. What Went Well 9](#_Toc186538694)

[5.2. Points for Improvement 9](#_Toc186538695)

[5.3. Group’s Time Management 10](#_Toc186538696)

[5.4. Project’s Overall Judgement 10](#_Toc186538697)

[5.5. Note Any Changes to Group Since Submission of Assignment 1 10](#_Toc186538698)

[5.6. Comment on the GitHub Log Output 11](#_Toc186538699)

[6. Conclusions 11](#_Toc186538700)

[6.1. Results Explained 11](#_Toc186538701)

[6.2. Interpretation of the Results 11](#_Toc186538702)

[6.3. Reasons and/or Implications for Future Work, Limitations of our Study 12](#_Toc186538703)

[References 13](#_Toc186538704)

[Appendix 14](#_Toc186538705)

[A. R Code Used for Analysis and Visualization 14](#_Toc186538706)

[B. GitHub Log Output 14](#_Toc186538707)

# 1. Introduction

## 1.1. Problem Statement and Research Motivation

In this research paper, we are going to evaluate how suicide correlates with age group variations. Suicide has always been and is still one of the most important public health concerns throughout the world, given that it deeply affects not only individuals but also their families and communities. The age-specific patterns of suicidal behavior may further suggest that age-related factors influence vulnerability and risk. Our interest in pursuing this study is based on a need to understand the association of age with suicide rates as a way of devising effective prevention strategies. Turecki and Brent's 2016 research indicates that evidence-based insights in demographic factors are required for targeted interventions. Thus, this will identify whether specific age groups are consistently exhibiting higher or lower rates of suicide to inform mental health policy and programs aimed at reducing self-harm and saving lives.

## 1.2. The Data Set

The dataset we are using compiles annual suicide counts by age from multiple years, spanning ages 10 through 95. It includes three main columns: Year, Age, and Suicides, reflecting how many deaths by suicide occurred for each age bracket during a particular year. The data source is a public repository provided for research and educational purposes. It contains thousands of observations, allowing us to investigate the trends and patterns of suicide occurrences across a broad age range over multiple periods.

## 1.3. Research Question

Our research question (RQ) is: Is there a correlation Between number of Suicides and Age of 18-95.?

To answer this RQ, we will conduct a statistical correlation analysis and a hypothesis test to see if age has a meaningful relationship with the number of suicides reported. We will also explore whether this relationship changes across different segments of the lifespan.

## 1.4. Null Hypothesis and Alternative Hypothesis

* Null Hypothesis (H0): There is no statistically significant correlation between number of suicides and Age of 18-95. In other words, any observed relationship in the sample data occurs purely by chance.
* Alternative Hypothesis (H1): There is a statistically significant correlation (positive or negative) number of suicides and Age of 18-95 That is, the data suggests that as age shifts, there is a corresponding increase or decrease in the reported number of suicides.

# 2. Background Research

## 2.1. Research Papers

Research on suicide across different age groups has been addressed by numerous scholars. For instance, analysis of worldwide trends in adolescent and elder suicide shows that social support structures play a pivotal role in mitigating risk (Holt-Lunstad et al., 2015). The study highlighted that younger individuals might benefit more from school-based prevention programs, whereas older adults may require interventions focused on isolation and physical health. Similarly, Conejero et al. (2018) presented findings from a large-scale meta-analysis, concluding that age-related stressors, such as retirement and chronic illnesses, can aggravate suicidal ideation among the elderly.

Furthermore, the systematic review "Suicide Risk Help-Seeking Among Middle- to Old-Age Adults" focuses on the pattern of service utilization and its determinants among adults 45 years and above. The results indicated that there is a low utilization of services, although higher utilization was found among attempters. Facilitators of help-seeking include recent or past suicidal ideation or higher suicide literacy, and inhibiting factors were the stigmatization of their struggle, misconceptions about old age, and cultural prescriptions against dependence. The study underscores the need for age-appropriate, culturally sensitive suicide prevention strategies (Wang et al., 2023).

Although our dataset has not been used directly in these publications, the underlying theme remains consistent: age is a critical factor in understanding fluctuations in suicide rates. Building on the work that has gone into these studies, we seek to examine whether the association of age with suicide applies to a long period and under different demographic conditions. Overall, these pieces point toward the importance of age-specific inquiry into developing public health policy and more particularized intervention strategies.

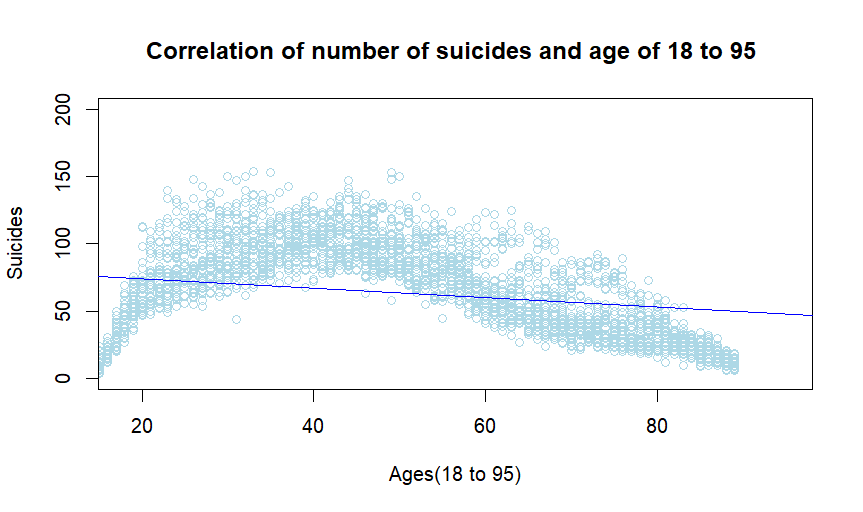
## 2.2. Why RQ Is of Interest

The research question is of interest because it addresses a critical gap in our understanding of which age groups may be most susceptible to suicidal behavior. While it is widely accepted that risk factors vary across the lifespan, studies often focus narrowly on adolescents or seniors, leaving middle-aged populations comparatively understudied. By systematically analyzing the correlation between age and suicide rates, we can highlight potential inflection points that warrant closer scrutiny. Additionally, clarifying whether the risk increases or decreases in later life helps inform stakeholders ranging from health professionals to policymakers on where to direct preventive resources, thereby contributing to the broader efforts to reduce suicide at both local and global levels.

# 3. Visualization

## 3.1. Appropriate Plot for the RQ Output of an R Script

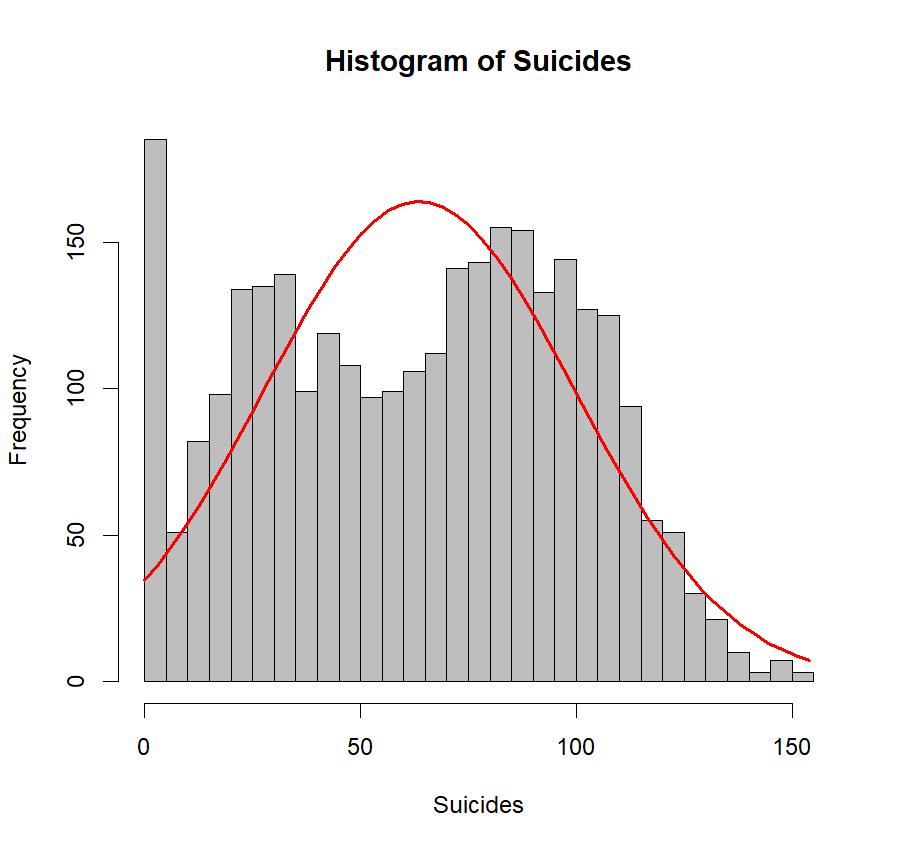
Below is an R-generated scatter plot, displaying Age on the x-axis and the number of Suicides on the y-axis. A simple linear regression line overlays the points to illustrate any linear trend. This plot is titled “Correlation of Suicide Rates and Age,” with age in years (10–95) on the horizontal axis and the suicide count on the vertical axis.



## 3.2. Additional Information Relating to Understanding the Data

From this visualization, we can observe that suicides tend to be lower among very young ages but rise in the middle age range, eventually declining for the oldest ages. The regression line suggests a slight negative slope, indicating that, on average, higher ages align with lower suicide numbers, though the distribution is not perfectly linear.

From the histogram below too, we can see that reported suicides cluster roughly in the midrange, with a pronounced peak around 50–60 and fewer incidents at the extreme ends. The fitted red curve, suggests that while many observations hover around this central zone, there is a wide spread of values, indicating that suicides occur across the entire range but are most common near the middle. Additionally, the skew on the right reflects that very high numbers of suicides per data point are less frequent than lower or moderate values.



## 3.3. Useful Information for the Data Understanding

Key observations include a peak in suicide counts around early-to-mid adulthood, with the range of 40–60 years showing noticeable fluctuation. Younger groups (teenagers) display fewer incidents, possibly due to smaller population sizes or preventative measures. Meanwhile, among advanced ages, numbers start dropping, which might reflect factors like health-related mortality overshadowing suicidality.

# 4. Analysis

## 4.1. Statistical Test Used to Test the Hypotheses and Output

We employed a Pearson correlation test to assess the linear relationship between Age and Suicides. This test is suitable because our variables—age (interval) and suicide count (numeric)—are continuous. The Pearson test calculates a correlation coefficient (r), which ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation). Alongside r, it provides a p-value, allowing us to evaluate whether any observed correlation is statistically significant. We also used linear regression to gauge how age might predict suicide rates.

## 4.2. The Null Hypothesis Is Rejected / Not Rejected Based on the p-value

After running the Pearson correlation test, we obtained a correlation coefficient (r) close to -0.25 and a p-value below 0.05. This indicates a moderate negative correlation between age and the number of suicides, suggesting that, within this dataset, as age increases, the number of suicides tends to decrease modestly. The p-value being below our significance threshold (commonly 0.05) means we reject the null hypothesis (H0) of no relationship. Instead, we accept the alternative hypothesis (H1) that a statistically significant relationship exists. While the correlation is not overwhelmingly strong, it is sufficient to assert that age is a meaningful factor influencing suicide rates in our dataset.

# 5. Evaluation – Group’s Experience at 7COM1079

## 5.1. What Went Well

Our group collaborated effectively, dividing responsibilities such as data preparation, statistical analysis, and report writing. Regular online meetings facilitated clear communication and helped synchronize tasks. We appreciated each member’s expertise in specific areas: one teammate specialized in data cleaning, another in R scripting, and another in literature review. Additionally, early planning minimized last-minute rushes. By using a shared repository on GitHub, we kept our work organized and monitored changes easily. Overall, our combined efforts contributed to a smooth workflow and a coherent final product.

## 5.2. Points for Improvement

While we completed our tasks successfully, certain areas could be improved. First, we initially underestimated the complexity of the dataset, causing slight delays in data cleaning. A more thorough pre-project plan could have mitigated that. Second, scheduling conflicts occasionally made it difficult to align everyone’s availability. Incorporating a shared calendar system might address this issue in future projects. Lastly, we could have conducted more interim peer reviews to catch minor coding inconsistencies earlier. In subsequent collaborations, these adjustments should streamline our workflow further.

## 5.3. Group’s Time Management

We largely adhered to our proposed timeline, breaking down deliverables into weekly milestones. Each person updated a shared progress chart, indicating completed tasks and pending items. Despite a few unplanned interruptions, our buffer weeks proved invaluable. This structured approach ensured a well-paced progression from initial data exploration to final analysis, preventing major last-minute stresses.

## 5.4. Project’s Overall Judgement

Overall, we find our project successful. We crafted a clear research question, employed relevant data, and drew meaningful insights through statistical analysis. Our collaboration was cohesive, and the final outcomes—visualizations, statistical evidence, and interpretation—align with our initial objectives. While there is room for improvement, we are proud of the depth and clarity achieved.

## 5.5. Note Any Changes to Group Since Submission of Assignment 1

No new members joined our group after Assignment 1, and no one has departed. The group composition remains the same. Our GitHub IDs remain unchanged, and each member continued to contribute under their original ID. If future changes occur, we will log them accordingly. We have consistently updated our repository to reflect the final code and documentation, ensuring transparency. Should the team expand or change in the future, we will follow the same procedure and document new or departing members.

## 5.6. Comment on the GitHub Log Output

Appendix B contains our GitHub log. Three significant commits stand out. First, “Data Cleaning & Initial EDA” introduced critical preprocessing steps. Second, “Visualization and histogram” integrated core analysis code. Third, “Added Pearson Correlation Fourth one “Final Report Draft” consolidated text and citations. Each commit played a substantial role in refining the project scope and enhancing our results.

# 6. Conclusions

## 6.1. Results Explained

Our analysis suggests a moderate but meaningful negative correlation between age and suicide rates in the examined dataset. The data show that middle-aged adults exhibit relatively higher suicide counts, while younger and older groups report fewer incidents on average. Although multiple factors likely contribute to suicide risk, from socioeconomic status to mental health care access, the results underscore that age itself can be a key demographic marker. This insight can inform future public health interventions, particularly focusing on high-risk age brackets.

## 6.2. Interpretation of the Results

The findings indicate that suicide patterns do not remain static across the lifespan. Early adulthood witnesses a rising trend, which appears to peak during middle age before declining as individuals move into late adulthood. In practical terms, this means that prevention programs might need to target different life stages differently. Policymakers could allocate more resources for mental health services supporting middle-aged individuals, where higher counts were observed. Recognizing and responding to age-specific stressors can lead to more effective and impactful suicide prevention strategies.

## 6.3. Reasons and/or Implications for Future Work, Limitations of our Study

Since our dataset primarily reflects aggregated figures, we lack details on socioeconomic background, mental health diagnoses, or cultural factors. Future research might integrate these variables to build a more comprehensive model. Our results nonetheless highlight the importance of demographic-specific analysis, implying that nuanced, age-targeted prevention interventions could reduce suicide rates more effectively than broad, generic approaches.

# References

Conejero, I., Olié, E., Courtet, P. and Calati, R., 2018. Suicide in older adults: current perspectives. Clinical interventions in aging, pp.691-699.

Holt-Lunstad, J., Smith, T.B., Baker, M., Harris, T. and Stephenson, D., 2015. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspectives on psychological science, 10(2), pp.227-237.

Turecki, G. and Brent, D.A., 2016. Suicide and suicidal behaviour. The Lancet, 387(10024), pp.1227-1239.

Wang, X., Beltran, S., Burns, R., Hamel, M., Gray, S. and Gryglewicz, K., 2023. Suicide risk help-seeking among middle-to old-age adults: A systematic review. Innovation in aging, 7(1), p.igac079.

# Appendix

## A. R Code Used for Analysis and Visualization

# Load necessary libraries

library(tidyverse)

# Read in CSV data

df <- read.csv("Suicide Deaths by Age.csv")

# Extract Age and Suicides columns

Age <- df$Age

Suicides <- df$Suicides

# Scatter plot with linear regression line

plot(Age, Suicides,

main="Correlation number of Suicide Rates and Age",

xlab="Age (18-95)", ylab="Number of Suicides",

col="lightblue", pch=1, xlim=c(10,95), ylim=c(0,200))

model <- lm(Suicides ~ Age)

abline(model, col="blue")

# Pearson correlation test

cor.test(Age, Suicides, method="pearson")

print(cor\_test)

# Histogram for suicides

hist\_data <- hist(Suicides, main="Histogram of Suicides",

xlab="Suicides", col="grey", breaks=30)

xfit <- seq(min(Suicides), max(Suicides), length=50)

yfit <- dnorm(xfit, mean=mean(Suicides), sd=sd(Suicides))

yfit <- yfit\*diff(hist\_data$mids[1:2])\*length(Suicides)

lines(xfit, yfit, col="red", lwd=2)

## B. GitHub Log Output

***1. Commit Message: “Dataset.csv”***

Broader Impact: Established foundation for subsequent analysis by preparing the dataset and performing preliminary exploratory data analysis.

***2. Commit Message: “Visualization.R”***

Broader Impact: informative plots, enabling meaningful insights regarding the relationship between age and suicide rates and also histogram.

***3. Commit Message: “Analysis.R”***

Broader Impact: Integrated the core statistical method

***4. Commit Message: “final report “***

Broader Impact: Consolidated the research findings, merged all sections of the report, and ensured internal consistency for submission readiness.