MEMORANDUM

From: Geneway Tang
To: Dr. Craig
Date: 12/05/2022
Subj: Data Analysis on Global Annual Suicide Deaths

SECTION 1: INTRODUCTION/OVERVIEW

Mental health issues have recently been on the rise due to various reasons and this is having an effect on death numbers in suicides. Death by suicide is a very complex issue that affects thousands of people worldwide yearly and this is a topic that I personally care about and am curious to dive deeper into learning trends associated with suicide rates. This dataset illustrates the annual number of deaths globally by cause and I believe grasping an understanding of the topic can assist in lowering numbers.

SECTION 2: STATEMENT OF HYPOTHESES

While looking for datasets that illustrate the deaths by suicides, I was curious to see how they compared to other deaths globally. Seeing the state of the media and how it is feeding into the minds of millions nationwide, I also wanted to see how suicide rates have changed over time in our country. I developed the following hypotheses based on my curiosity to discover these trends.

Hypotheses:

Null Hypothesis (H0): The suicide rates globally on average are the same as homicide rates.
Alternative Hypothesis (H1): The suicide rates globally on average are not the same as homicide rates.

Null Hypothesis (H0): Suicide rates in the US have been the same as the global average over the course of the years
Alternative Hypothesis (H1): Suicide rates in the US have not been the same as the global average over the course of the years

SECTION 3: DATA AND ANALYSIS METHODOLOGY

Table 1: Annual Number of Deaths by Cause

Entity	Code	Year	Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)	Deaths - Self-harm - Sex: Both - Age: All Ages (Number)
Afghanist.	AFG	2007	3657	1310
Afghanist.	AFG	2008	3785	1330
Afghanist.	AFG	2009	3874	1342
Afghanist.	AFG	2011	4170	1391
Afghanist.	AFG	2012	4245	1413
Afghanist.	AFG	2013	4379	1433
Afghanist.	AFG	2014	4420	1444

Details on the data:

The dataset that was collected for this final project was found on ourworldindata.org and had many variables showing different death causes, along with the year and country. Since I wanted to focus on suicide rates and homicide rates for the purpose of this project, I decided to narrow down my dataset to highlight these variables. To test my hypotheses, I utilized these narrowed-down variables and utilized both Excel and Minitab to gain a better understanding of the dataset. For the analysis, I utilized the 2-sample t-test, linear regression, pivot tables, and line plots to learn about the data.

SECTION 4: RESULTS

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Deaths - Interpersonal violence	6840	10637	40889	494
Deaths - Self-harm - Sex: Both	6840	20776	81126	981

Estimation for Difference

Difference	95% CI for Difference
-10139	(-12292, -7986)

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-9.23	10103	0.000

Table 2: 2-sample T-test between homicides and suicides

Table 3: Line plot to illustrate the average deaths for both homicides and suicides
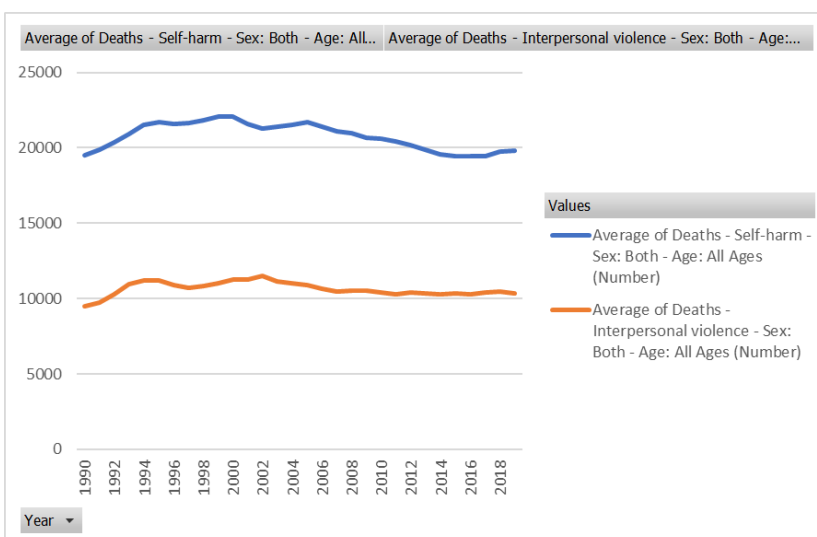
To test my first null hypothesis that both the average rates of suicides and homicides are equal, I first utilized a pivot table in Excel to organize the data and narrow down my variables to the desired output. I then aggregated the values by the mean and plotted a line chart to visualize the trends of both variables over the years. Seeing that there was quite a distinction, I ran a 2-sample t-test which tested that the means for both suicides and homicides were equal. The results yielded a p-value of 0.00, indicating that there is a 0% possibility that these two rates are equal. By looking at the line graph where there is a clear distinction, we can see that suicide rates are approximately double homicide rates globally. These results provide strong evidence to reject the null hypothesis that the suicide rates are equal to the homicide rates globally.

Regression:
Table 4: Regression analysis for Suicide Rates in the USA over time

The regression equation is
Deaths - Self-harm - Sex: Both = - 980583 + 508.6 Year

Model Summary

S	R-sq	R-sq(adj)
1565.38	89.45%	89.07%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	581435336	581435336	237.28	0.000
Error	28	68611781	2450421		
Total	29	650047117			

Fitted Line Plot
Deaths - Self-harm - Sex: Both = - 980583 + 508.6 Year

S	1565.38
R-Sq	89.4%
R-Sq(adj)	89.1%
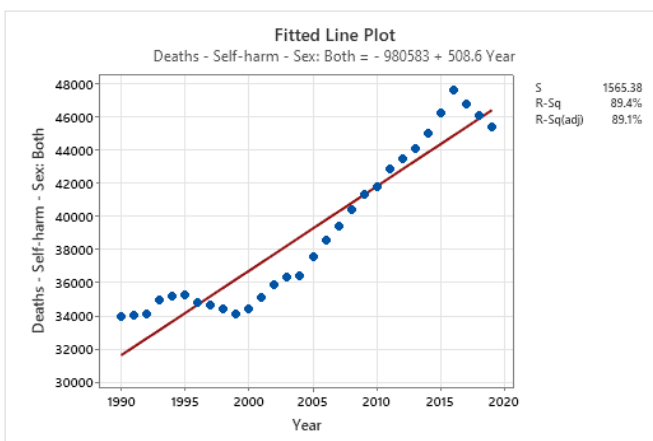
For my second null hypothesis test, I wanted to see the trend of suicides in the USA over the years. I first filtered the country data by creating a subset worksheet in Minitab and selected the desired country, the USA. Interestingly enough, plotting a scatter plot of the deaths over the years, it can be seen that there was a strong upward trend and it yielded an R-squared value of 89.4%, indicative of a strong fitted line. The plot illustrates a strong upwards-sloping

relationship, showing that the rate of suicide in the USA has indeed been increasing, rejecting the null hypothesis that it has stayed the same.

SECTION 5: DISCUSSION AND CONCLUSION
To recap, my analysis of suicide rates was to test for interesting trends regarding suicides both globally and within the United States. I tested my first hypothesis on suicide rates and their comparison to homicide rates globally by using a 2-sample T-test along with pivot tables to create line plots for illustration in Excel. I was able to discover the large distinction between the two variable means and how suicides have significantly higher numbers than homicides worldwide.

For my second hypothesis, I utilized a regression fit line on a scatter plot to determine the trend of suicide rates over the years. My results illustrated a strong upwards trend in suicide rates within the United States, and I was then able to reject my null hypothesis that the rates remained the same.
Some limitation that I encountered during this project was the lack of data that was presented. While there were many countries in the dataset and many causes of death, I wanted to dive deeper into the statistics of suicides in the USA, and when I narrowed it down to just one country, I was left with very little data to work with. One way to combat this would have more data over the years or include data from different states of the country. For the future, I would suggest finding a more specific dataset on the USA, rather than a global dataset.

REFERENCES
Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.