BELLEVUE
UNIVERSITY

Bellevue University

DSC 680

Winter 2021

Project Two:

An Analysis of World Suicide Rates

February 2, 2021

Jennifer Ruiz

Abstract

More people around the world are facing mental health crises than ever before. As suicide rates continue to rise globally, research into contributing factors is more important than ever. A wide body of research exist examining patient-centered factors for suicide such as trauma, loss, and socio-economic status. However, the research is lacking on the influences of availability of mental health care.

In this project, the author chose to focus on the availability of mental health care and its impact o suicide rates globally. Using machine learning methods, the author seeks to determine how treatment options and provider accessibility influence rates of suicide. Data visualization methods will be used to explore the relationships between factors and identify potential areas for future research.

Keywords: Regression, K-nearest neighbor, Machine Learning, Python

Introduction

Suicide ranks as a leading cause of death for both men and women in America. This trend is echoed around the globe. Unsurprisingly, suicide rates between men and women vary widely around the world. According to SSM Health (2019), serious life stressors such as "intimate partner problems, legal issues, unemployment/financial problems and health concerns" account for higher rates among middle-aged men. Though suicide is a more prevalent cause of death among middle-aged individuals, it also ranks as a leading cause among seniors and young adults. Bitten (2019) points out that suicide is a leading cause of death among 20-34-year old's in the UK and 27% percent of those individuals are under the care of a physician for mental health concerns. This leads one to question what impact the availability and accessibility of mental health care has on suicide rates around the world.

The World Health Organization collects extensive data regarding mental health services around the world. Information collected includes suicide rates by gender, per capita availability of hospital, inpatient and outpatient treatment, as well as, provider data. Using the WHO data, the author seeks to focus on understanding and identifying trends in access and level of mental health care and their impact on suicide rates.

Research Question

This project endeavors to analyze the impact of multiple factors in suicide rates for understanding and to predict increased risk. To achieve this end, a comprehensive statistical and visualization analysis will be conducted.

Key Research Questions include:

1. What impact do widespread access to mental health facilities have on suicide rates?

2. What countries have the highest rates of suicide?

3. Do lower rates of day treatment mental health facilities predict higher suicide rates?

4. What factors are the greatest indication of suicide rates within a country? What factors
    have the least impact in predicting suicide rates?

Background Information

Methods

Data Understanding

When approaching this project, the author began by examining the individual attributes
contained in the three datasets utilized. All data was coded numerically based on reporting
country. Attributes were visualized during Exploratory Analysis using Python (see appendix one)
to gain insight into relationships within the data.

Data Preparation

Significant data preparation took place to prepare the data for exploration and modeling.
The data contained a high number of missing values for some features. Two features were
removed from the *Facilities* dataset, as more than fifty percent of the values were missing.
Missing value replacement for the remainder of that dataset was done using the median value of
each feature so as not to skew the distribution. The *HR* dataset, which examines facilities
staffing, also contained a moderate number of missing values. All null values within that dataset
were replaced by zero to preserve the data.

Modeling

Data preparation allowed for multiple regression methods to be performed on the data. For all analyses, a test train split was employed on the dataset. Analysis began with a multivariate linear regression analysis. This was followed by a k-nearest neighbor regression and a decision tree regression model. The author chose to a utilize these prediction methods to fully explore the story of the data and create the best fitting model. Each model tested a different set of features to determine those with the best prediction power. All models were created using the sklearn package.

Results

Exploratory analysis and the models created produced significant insights into suicide rates globally. Figure one shows the overall suicide rates per country visualized on a world map. Exploratory analysis revealed that Guyana has the highest suicide rates in the dataset, while Barbados has the lowest rates. It also revealed Egypt has the highest number of mental hospitals with 8.31 per one hundred thousand people. The data also revealed that outpatient facilities and health units were more prevalent than mental hospitals throughout the world. Two final areas of interest were only three countries exhibited higher suicide rates among women than men and five countries showed almost equal suicide rates among the genders.
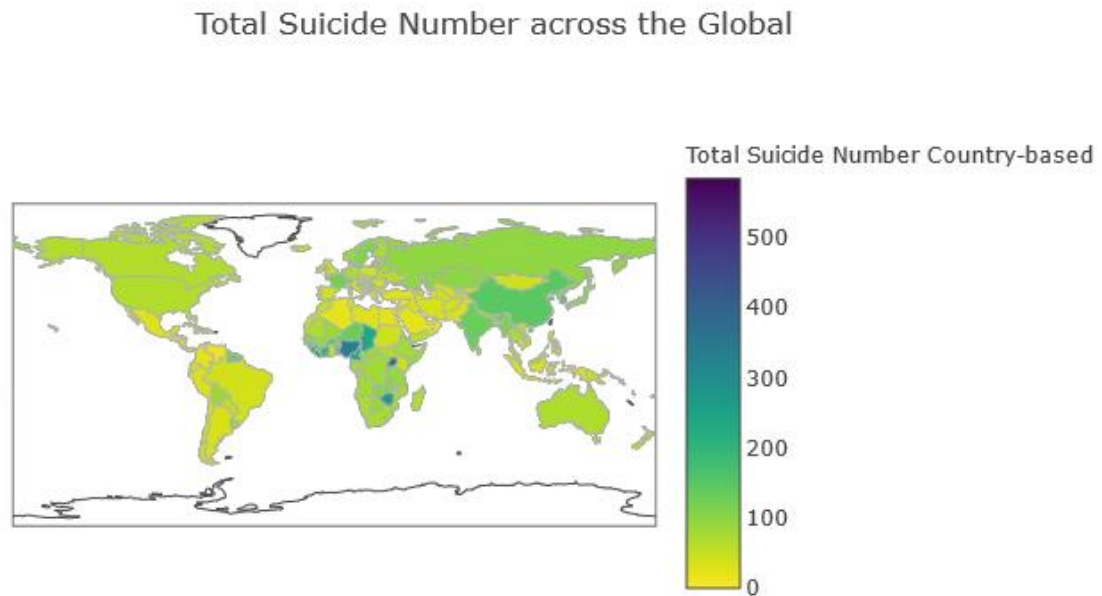
Total Suicide Number across the Global

Figure One **Suicide Rates Around the World**

The multivariate linear regression was run twice, each time with different features. The first analysis yielded an r-squared value of -0.067. The second analysis yielded an r-squared value of 0.03. This indicates that neither group of features have high predictive value for suicide rates.

The k-nearest neighbor regression model was then applied to the data. Results indicated overfitting of the model to the test data, with accuracy at -0.081 and an RMSE of 6.15. Finally, a decision tree regression was applied to the model. This yielded the most promising results of all analyses, with an accuracy of 0.5 on the training data and 0.1 on the test data. Again, the model showed overfitting in the test data.

Discussion

The analysis produced some surprising results and challenges during the analysis were encountered. The performance of the regression analysis yielded limited insight. Given this insight, future analysis should explore additional features that may yield higher predictive power. Exploratory analysis yielded surprising results regarding hospitals, outpatient facilities, and trends in suicide rates around the world.

Conclusion

As mental health and wellbeing continues to be impacted by these unprecedented times, understanding how access and availability of mental health care impacts suicide rates is crucial for prevention. This analysis provided insights regarding global trends in suicide rates and how provider access impacts suicide rates. This is a growing body of knowledge that can be used to impact prevention and government planning for mental health services.

References

Bitten, H. (2019). Suicide in the 21st Century (Part 1). Retrieved on January 10, 2021 from

https://towardsdatascience.com/suicide-in-the-21st-century-part-1-904abe8e1f5c.

Kaggle. (n.d.). Mental Health and Suicide Rates. Retrieved on January 2, 2021 from

https://www.kaggle.com/twinkle0705/mental-health-and-suicide-rates?select=Age-

standardized+suicide+rates.csv.

SSM Health. (2019). Why are middle-aged white men more likely to die by suicide? Retrieved

on February 2, 2021 from https://www.ssmhealth.com/blogs/ssm-health-matters/october-

2019/middle-aged-men-more-likely-to-die-by-suicide

Appendix A

Visuals one through thirteen are feature examination during the exploratory analysis process.

Table One

Distribution of Mental Hospitals

Table Two

Distribution of Outpatient Facilities
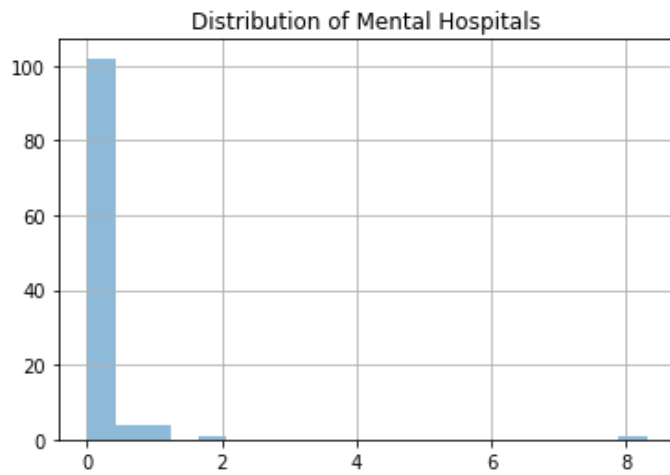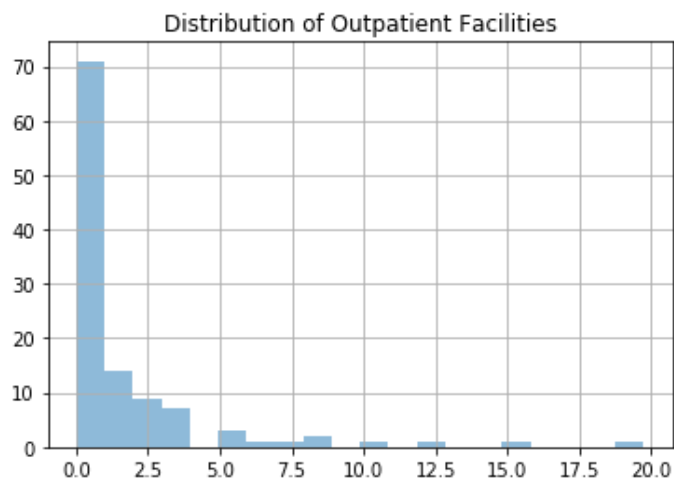
Table Three
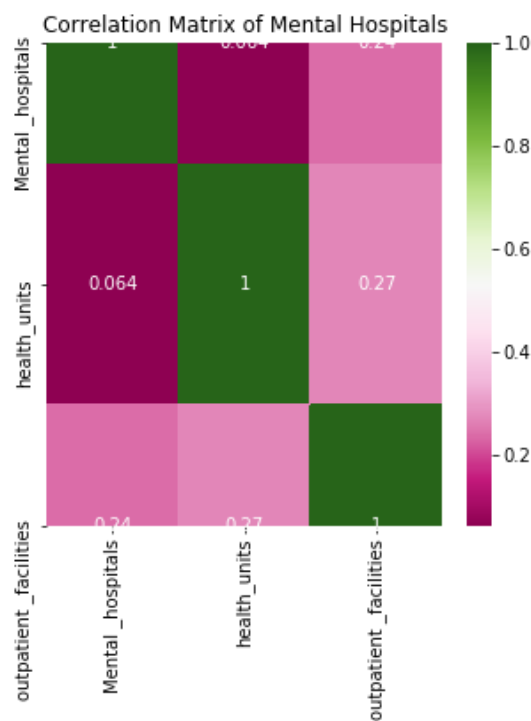
Table Four
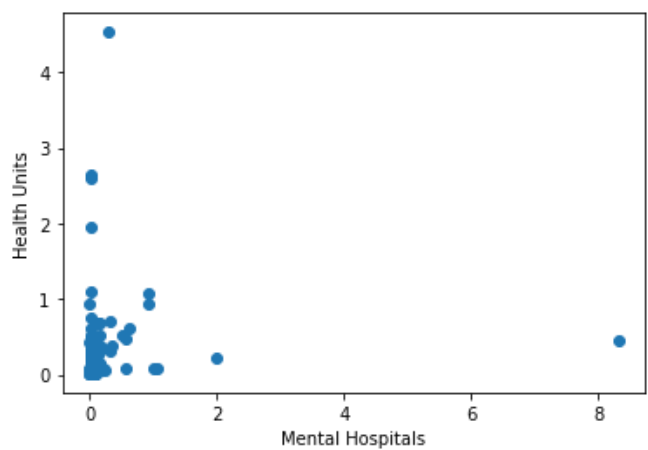
Table Five

Table Six

Table Seven

Correlation Matrix of Mental Health Staffing
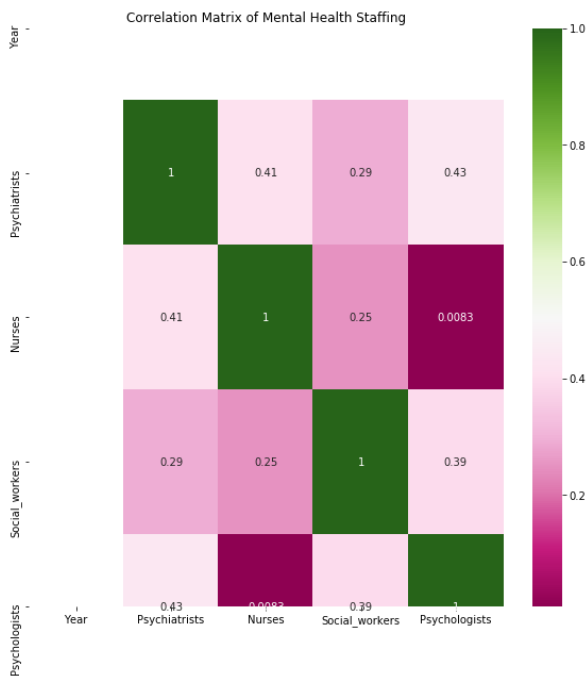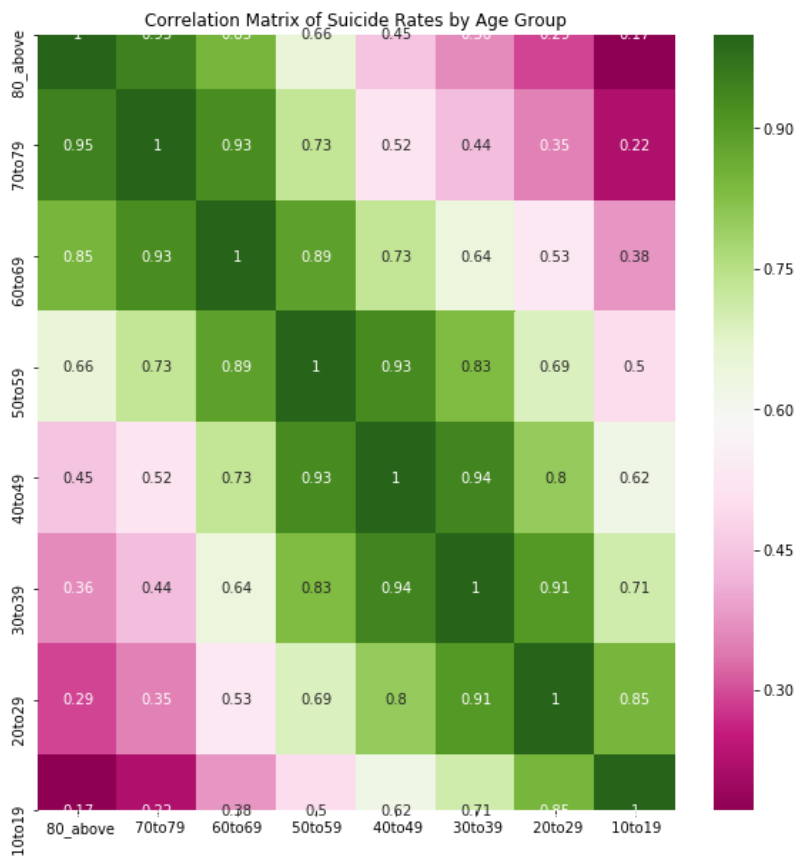
Table Eight

Correlation Matrix of Suicide Rates by Age Group
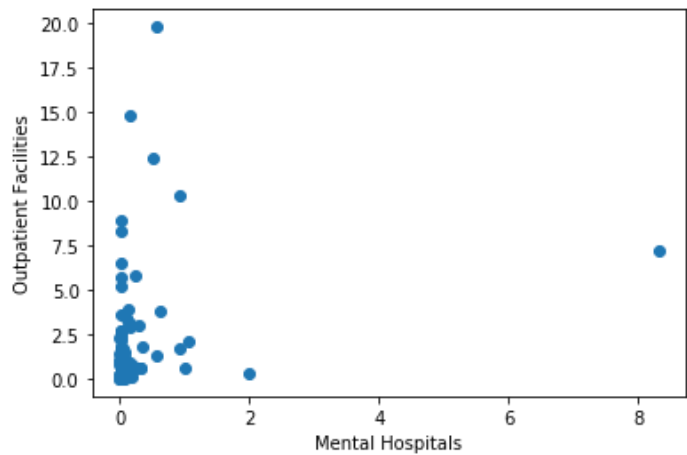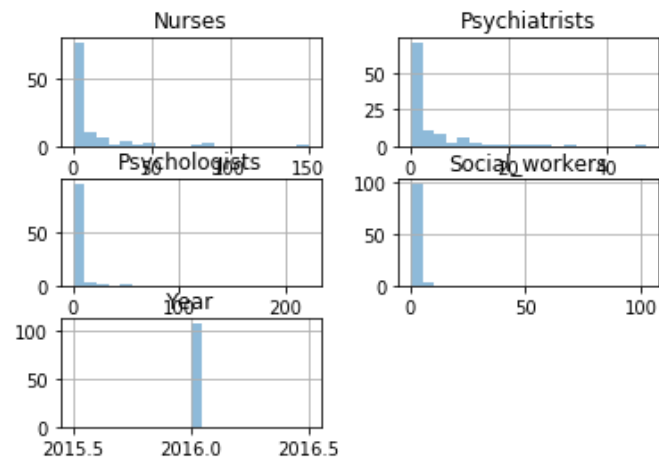
Table Nine

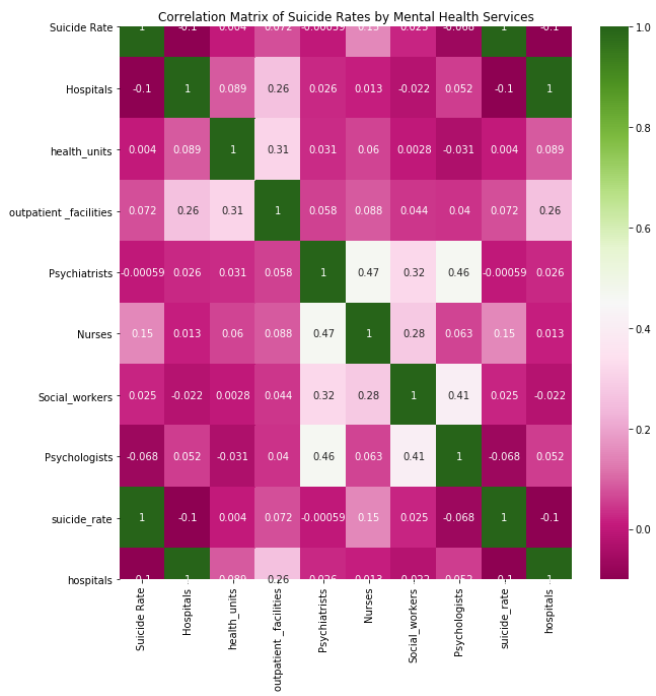Sort the Suicide Number Across the World

Table Ten

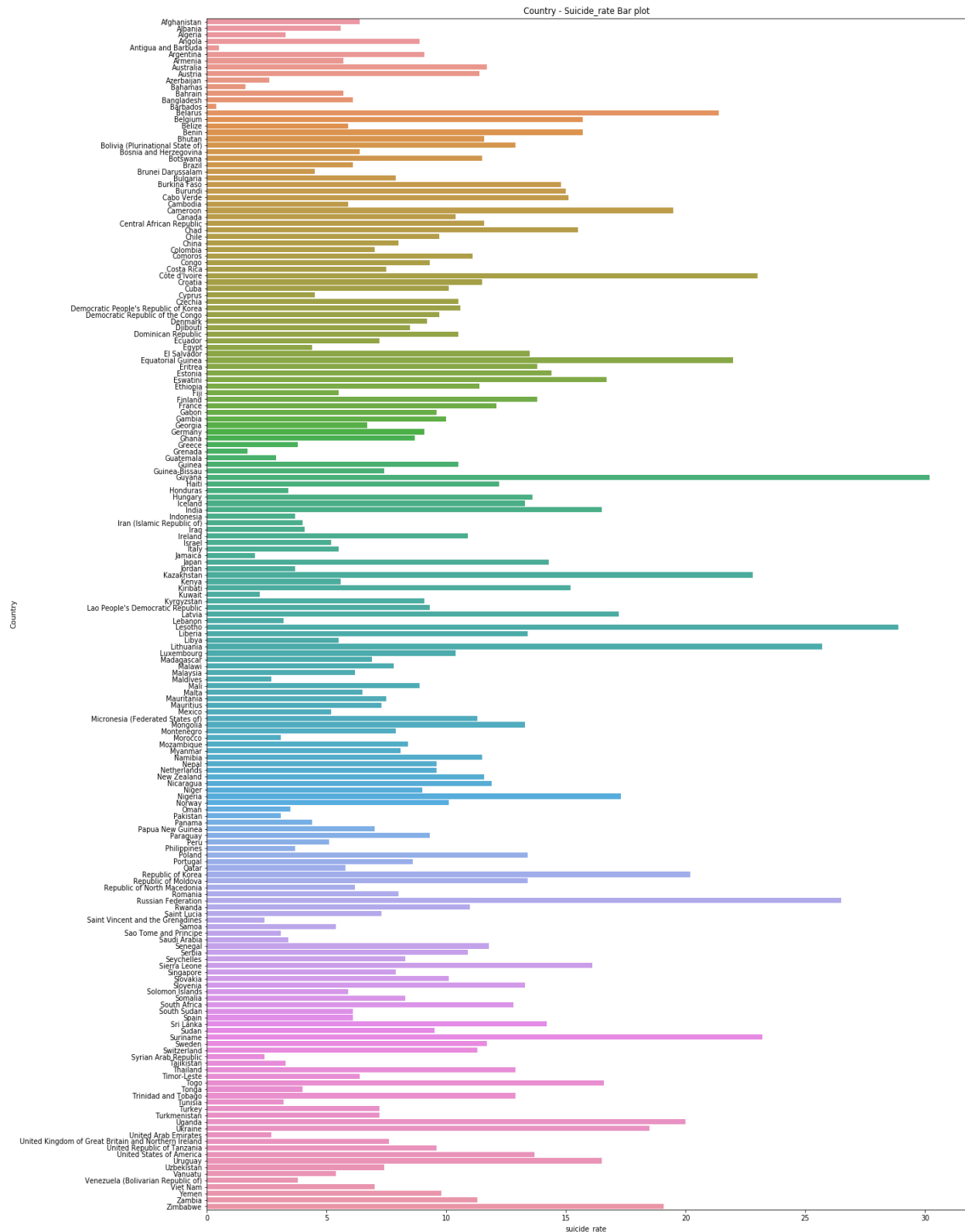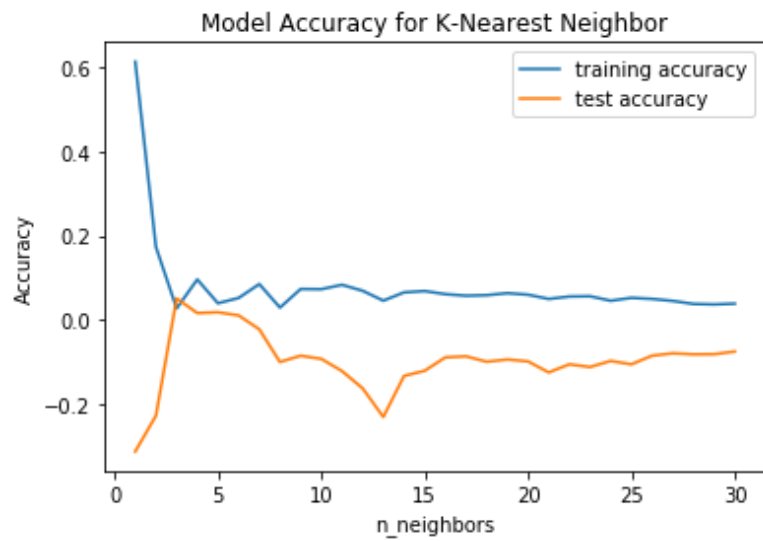Correlation Matrix of Suicide Rates by Mental Health Services

Table Eleven

Country - Suicide_rate Bar plot
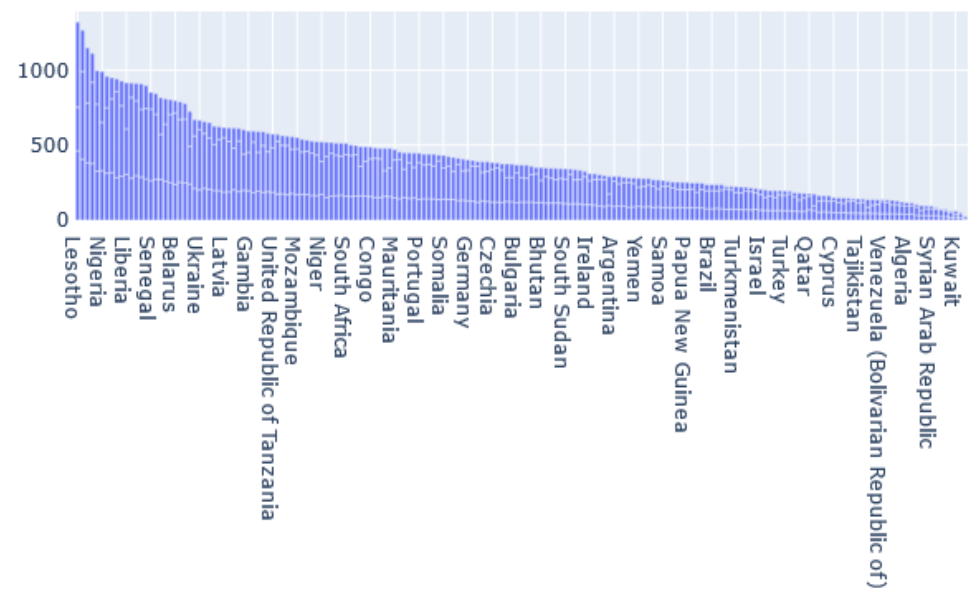
Table Twelve

Table Thirteen