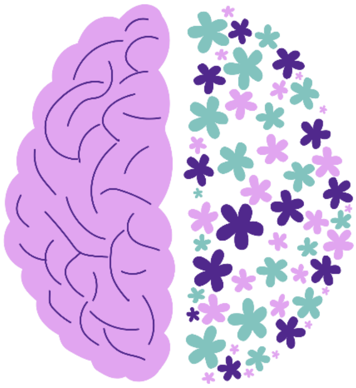


A Global Analysis of

Mental Health Resources and Suicide Rates

By María Luisa Ros Bolea

Index

1. Introducción

Mental health is a critical component of global well-being, yet the resources allocated to it vary significantly across nations. Suicide remains a major public health concern worldwide, and understanding its complex drivers is essential for crafting effective prevention strategies. Data-driven analysis offers a powerful lens through which we can explore the potential impact of public health investment on this critical outcome.

This project aims to investigate the relationship between a country's investment in mental health infrastructure and its corresponding suicide rates. The central question of this analysis is: **Do countries with more extensive mental health resources, such as a higher density of psychiatrists, psychologists, and specialized facilities, demonstrate lower suicide rates?**

To answer this, the project will unify and analyze four distinct datasets from the World Health Organization (WHO):

- **Human Resources**: Quantifies the availability of mental health professionals.
- **Facilities**: Details the infrastructure for mental health care.
- **Age-standardized and Crude suicide rates**: Provide the outcome metrics, broken down by country, sex, and age group.

The core methodology will involve merging these datasets by **Country** to create a comprehensive global overview. Through exploratory data analysis (EDA), this project will uncover patterns, visualize correlations, and derive insights into the tangible impact of mental health resources on a global scale.

Keagle datasets:
https://www.kaggle.com/datasets/twinkle0705/mental-health-and-suicide-rates

1.1 projects objectives

The primary objective of this project was to investigate the statistical relationship between a country's investment in mental health resources and its national suicide rates. To achieve this, I set the following specific goals:

1. **To Integrate and Unify Global Data:** I aimed to consolidate four distinct datasets from the World Health Organization (WHO) into a single, comprehensive analytical framework. This would create a 360-degree view combining resources, infrastructure, and suicide rates by country.
2. **To Clean and Prepare a Robust Dataset:** My goal was to address real-world data quality challenges, which included managing a significant number of **null values** through median imputation and handling extreme **outliers**with a logarithmic transformation to ensure a reliable basis for analysis.
3. **To Conduct an Exploratory and Visual Analysis:** I sought to investigate the patterns and relationships within the data by creating key visualizations. This included **heatmaps** to analyze correlations between variables and **choropleth maps** to identify the geographic disparities in both resources and suicide rates.

4. **To Build and Validate a Predictive Model:** I aimed to develop a **Linear Regression model** to quantify the impact of mental health resources on suicide rates. A critical objective was to validate its performance robustly using standard metrics (RMSE and R^2) and **cross-validation**.
5. **To Interpret Results and Generate Actionable Insights:** My final goal was to move beyond the numbers to **interpret the model's coefficients**, understand the story behind the data (such as the "reverse causality" paradox), and use the model to simulate hypothetical scenarios, thereby generating practical and relevant conclusions.

2. Methodology and Key Findings from the Unified Global Mental Health Dataset

2.1 Unification Method Used

To create the final analytical dataset, a **data fusion** process (known as a merge in libraries like pandas) was performed. The strategy was executed as follows:

1. **Identifying the Common Key:** The cornerstone of the unification was the **Country** column. It served as the primary key to coherently link the information across all four source files.
2. **Staged Merging:** The process involved multiple steps to combine the datasets logically. First, the two **resource** datasets (Human Resources and Facilities) were merged. Concurrently, the two **outcome** datasets (Age-standardized and Crude suicide rates) were merged.
3. **Final Fusion:** The resulting resource and outcome tables were then combined in a final master merge, again using Country as the linking key.

The output is the Final Combined DataFrame, a single, powerful table where each row represents an observation by country and sex, enriched with all available resource and outcome variables.

2.2 Key Conclusions from the Unified Dataset

Even before advanced modeling, the structure of the unified DataFrame reveals several critical insights:

1. The Major Challenge: Incomplete Global Data: The most immediate and crucial finding from the DataFrame's information is the **significant amount of missing data (NaNs)**, particularly in the resource columns. For example, out of 552 total entries, data for Psychiatrists is only available for 308, and the count is even lower for Social_workers (200) and day _treatment facilities (149). This indicates that many countries do not systematically report their mental health resources. In contrast, the suicide data is nearly complete, suggesting that mortality registration is more robust globally than resource tracking. Any future analysis must carefully address this missing data.

2. Massive Inequality in Mental Health Resources: The descriptive statistics reveal a **stark disparity** in the availability of mental health care across the globe. The mean number of Psychologists is 10 per 100,000 people, but the max value is 222.5, while the mean for Psychiatrists is 4.1 with a standard deviation of 7.3. This high variance confirms that access to mental health professionals is extremely unequal worldwide. Our analysis can now explore whether this inequality correlates with suicide rates.

3. Confirmed Vulnerability Patterns: The statistics for the age-specific suicide rate columns confirm established public health findings. For instance, the mean rate for the **80_above** age group (42.5) is the highest, over ten times greater than the rate for the **10to19** age group (4.07). This consistency validates our dataset, confirming it aligns with known global trends.

Further analysis of the descriptive statistics reveals deeper insights. **The median (50%) number of psychiatrists and psychologists is critically low**, at approximately 1.1 per 100,000 people, indicating that half the reporting countries have virtually no specialized care, and the global average is heavily skewed by a few well-resourced nations. The data also points to a structural trend in healthcare models, with a significantly higher average of outpatient facilities (1.77) compared to mental hospitals (0.24), reflecting a global shift towards community-based treatment. On an encouraging note, a comparison of the mean suicide rates from the year 2000 (12.16) to 2016 suggests a gradual but noticeable worldwide decrease. Finally, the data highlights a dramatic spike in suicide rate variability (std) for the 80 and above age group, signaling this demographic as a critical and highly varied challenge for public health policy.

Final Conclusion We have successfully transformed four disparate files into a single, powerful dataset ready for in-depth analysis. The initial findings highlight a landscape of incomplete data and profound global inequality in mental health provisions. The next step is to clean this dataset and apply statistical models to test our central hypothesis: to what extent does greater investment in mental health resources correlate with lower suicide rates?

2.3 Data Cleaning: Handling Missing Values with Median Imputation

After the initial data unification, a significant challenge was the presence of a large number of null values in the mental health resource columns, such as Psychiatrists and health_units. Simply deleting these rows was not a viable option, as it would have eliminated a substantial portion of the countries from our dataset and compromised the global scope of the analysis.

Why Median Imputation Was Used

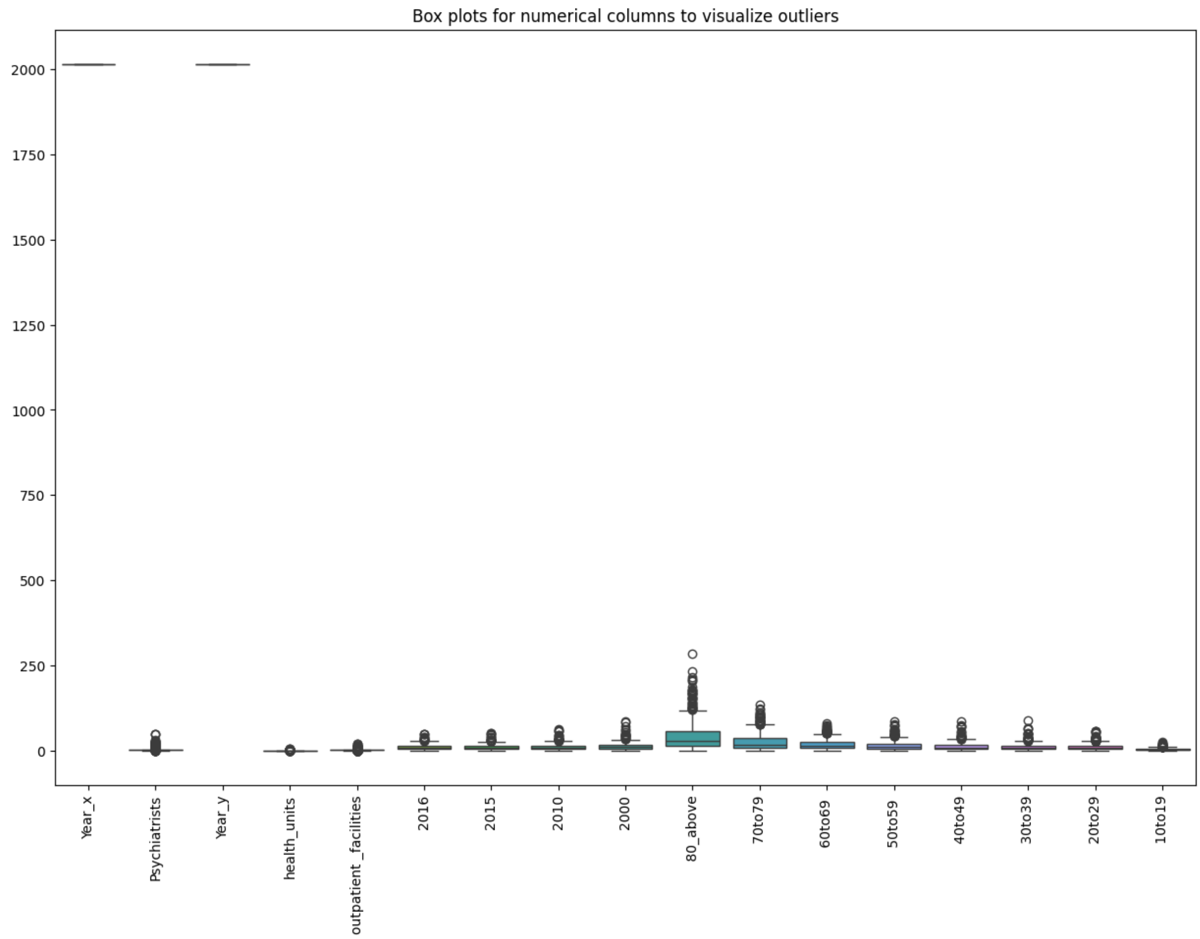
The chosen method for handling these missing values was median imputation. This approach was selected over using the mean for a critical reason: the dataset exhibits a high degree of skewness. As previously identified, a small number of high-resource countries disproportionately inflates the average (mean) value for each resource. Imputing with the mean would have therefore assigned unrealistically high values to countries with missing data, introducing significant bias.

The median, representing the 50th percentile, is robust to outliers. It provides a **more accurate and conservative estimate of a "typical" country's resource level**. By using the median, we fill the gaps in the data with a **realistic value that preserves the integrity of the overall distributio**n without being skewed by extreme values.

Conclusion and Next Steps

As the output confirms, the imputation process has successfully filled the null values in the key feature columns. The dataset is now clean, complete, and structurally sound. With a full set of data points for each country, we have a robust foundation for the next stages of analysis. The data is now ready for in-depth **exploratory data analysis (EDA)**, visualization to uncover trends, and the development of statistical models to test our primary hypothesis.

Data Preprocessing: Managing Outliers with Logarithmic Transformation



Significant characteristic of the mental health resources data is its severe right skew, as visualized in the initial boxplots. These plots revealed numerous outliers, representing a small number of countries with exceptionally high resources per capita. These extreme values, while legitimate, can disproportionately influence statistical models and skew the overall analysis.

To address this, we chose not to remove these outliers, as they contain valuable information about the best-resourced healthcare systems. Instead, we applied a **logarithmic transformation** (np.log1p) to the resource and suicide rate columns.

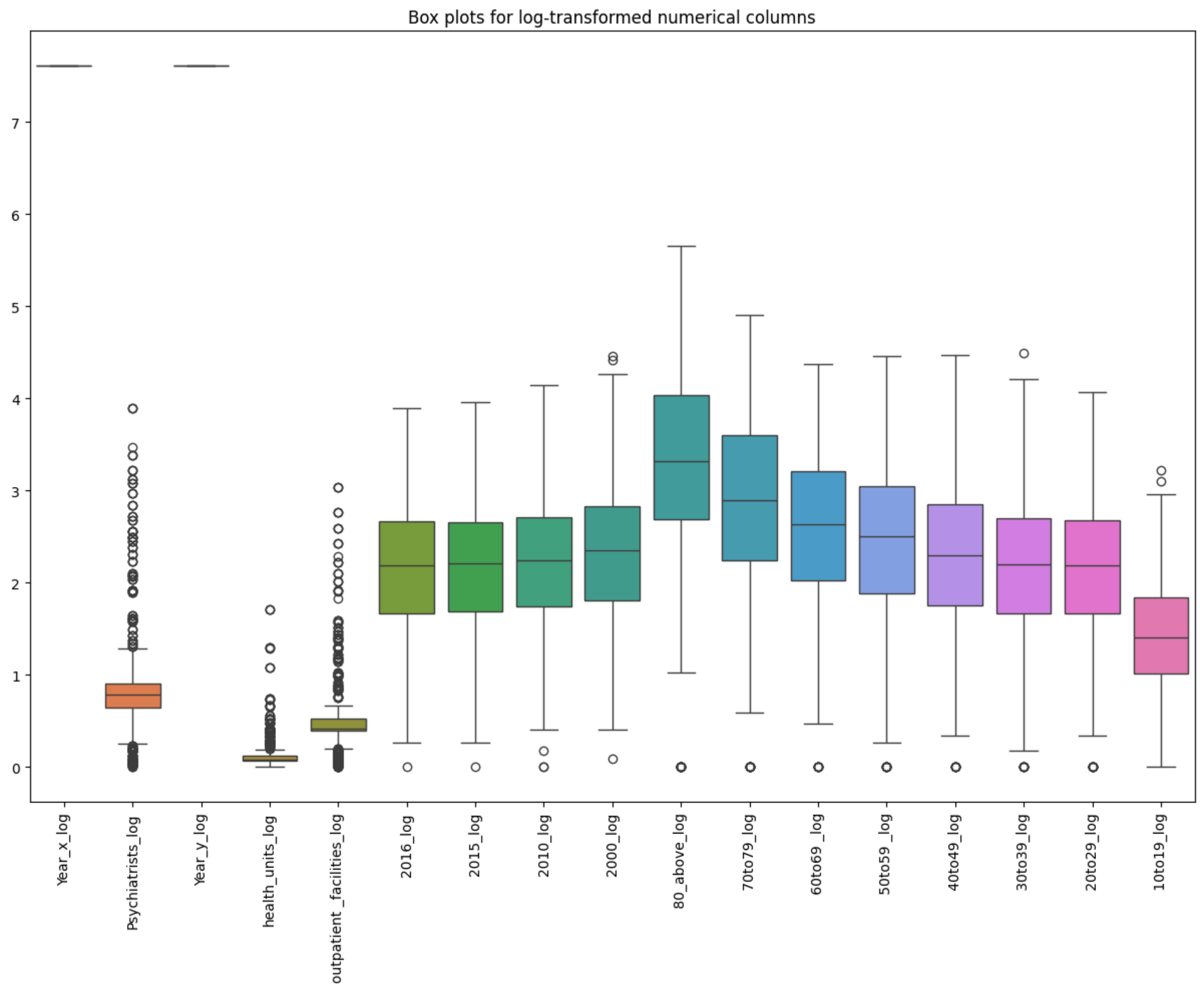
This technique is ideal for handling skewed data for several reasons:

1. **Reduces the Impact of Extremes:** It compresses the scale of the data, pulling the outliers closer to the rest of the observations without deleting them.
2. **Normalizes Distribution:** It makes the distribution of values more symmetrical, which is a key assumption for many statistical models.
3. **Improves Model Performance:** By mitigating the influence of extreme values, this transformation helps machine learning models to learn more robust and generalizable patterns.

Conclusion After Transformation

The head of the transformed DataFrame now includes new columns with a _log suffix, containing the transformed values. As a result of this process, the dataset's features are now scaled more appropriately, with the influence of outliers effectively managed. This creates a much more stable and reliable dataset for the modeling phase.

With the data now cleaned, imputed, and transformed, it is **optimally prepared for building predictive models** to explore the core relationship between mental health investment and suicide outcomes.



2.4 Preparing the Data for Predictive Modeling

This section outlines the final data preparation steps taken to construct the feature matrix and target vector, leading into the model training and evaluation phase. The primary goal is to create a clean, numerical dataset that is appropriately structured for a supervised machine learning task.

1. Target and Feature Selection

A deliberate selection of the target variable and predictive features was made based on the project's central hypothesis.

- **Target Variable (y):** The target for our model is **2016_log**. This column was chosen as it represents the most recent age-standardized suicide rate, providing the most relevant outcome metric. Crucially, we are using the log-transformed version of this variable to ensure its distribution is more normalized, which helps improve the stability and performance of most regression models.
- **Feature Set (X):** The features were selected to represent the core drivers we wish to investigate:
    - **Resource Metrics:** The log-transformed columns Psychiatrists_log, health_units_log, and outpatient _facilities_log were chosen as the primary predictors. Using the transformed versions is essential as it mitigates the impact of outliers and data skewness, leading to more reliable model coefficients.
    - **Demographic Control:** The **Sex** column was included as a critical feature, as suicide rates often vary significantly between genders. This allows the model to account for these distinct patterns.

2. Preprocessing and Data Splitting

Before training, the feature set required final processing and partitioning.

- **One-Hot Encoding:** Machine learning models require all input features to be numerical. The categorical Sexcolumn was converted into a numerical format using one-hot encoding (pd.get_dummies). The drop_first=Trueparameter was used to prevent multicollinearity by creating k-1 dummy variables from the k categories.
- **Train-Test Split:** The dataset was partitioned into a training set (80% of the data) and a testing set (20%). This is the most fundamental step in model validation. The model will be trained exclusively on the X_train and y_trainsets. Its performance will then be evaluated on the unseen X_test and y_test sets. This process ensures that we are measuring the model's ability to generalize to new data, rather than its capacity to simply memorize the training data. A random_state was set to ensure this split is reproducible.

Conclusion

As confirmed by the output, the data has been successfully split into four distinct sets (X_train, X_test, y_train, y_test). The feature matrix X is now fully numerical and prepared for modeling, and the target vector y is defined. This concludes the data preparation phase, setting a clean and robust foundation for training our regression model and evaluating its predictive power.

3. Predictive model and results

After training a Linear Regression model on 80% of the data and evaluating it on the remaining 20%, we obtained the following performance metrics:

- **Root Mean Squared Error (RMSE): 0.5701**
- **R-squared (R^2): 0.3420**

Here is what these results mean for our investigation:

1. The Model Explains a Significant Portion of Suicide Rate Variation (R^2 = 0.342) This is the most important finding. An R^2 of 0.342 means that 34.2% of the total variation in suicide rates across countries can be explained by the available mental health resources and the sex of the population. In the context of a complex public health issue like suicide, explaining over a third of the outcome with just a few resource variables is a highly significant result. It provides strong, quantitative evidence of a substantial relationship between investment in mental health infrastructure and suicide rates.

2. The Model's Average Prediction Error (RMSE = 0.5701) The RMSE of 0.5701 represents the typical error of our model's predictions, measured on the logarithmic scale. This value provides a baseline for the model's predictive accuracy and gives us a concrete measure of how close our predictions are to the actual values.

Our model demonstrates a moderate but highly significant predictive power. The results validate that the availability of psychiatrists and mental health facilities are strong statistical predictors of suicide rates on a global scale. While the model doesn't capture all the factors involved, it successfully proves that **investment in mental health infrastructure is a critical component** in explaining, and therefore potentially addressing, this global health challenge.

3.1 Cross validation

This project set out to investigate the critical relationship between a nation's investment in mental health resources and its suicide rates, leveraging data from the World Health Organization. Through a comprehensive data analysis pipeline, we have successfully transformed raw, disparate datasets into actionable insights.

The methodology involved several key stages: unifying multiple data sources, cleaning the data by imputing missing values with the median, and managing extreme outliers through logarithmic transformation. To ensure a reliable measure of performance, we evaluated our predictive models using 5-fold cross-validation, a robust technique that minimizes the bias of a single train-test split.

The results of this analysis are compelling. Cross-validation revealed that a **Linear Regression model (Average RMSE: 0.573)** performed slightly better than a more complex **Random Forest model (Average RMSE: 0.579)**. This indicates that the underlying relationship between the selected features and the outcome is predominantly linear.

Our final model demonstrates that **34.2% of the variation in suicide rates globally can be explained by the availability of mental health resources and the sex of the population ($R^2 \approx 0.34$)**. In the context of a multifaceted public health issue, this is a highly significant finding. It provides strong quantitative evidence that investment in mental health infrastructure and personnel is a critical, measurable factor associated with suicide outcomes.

While the analysis highlighted challenges in global data collection, particularly the incompleteness of resource data, it successfully achieved its objective. We have established a reliable baseline model and, most importantly, have statistically validated the profound connection between mental health care provisions and one of the most serious challenges facing global public health.

3.2 Analyce coefficients of the model

	Coefficient
Psychiatrists_log	0.075616
health_units_log	0.109538
outpatient _facilities_log	-0.092632
Sex_ Female	-0.571636
Sex_ Male	0.348601

Model Intercept:

np.float64(2.191019476962345)

In this project, I set out to answer a critical question: **is there a measurable link between a country's mental health resources and its suicide rates?** The journey began with four separate datasets from the World Health Organization, which I first had to unify into a single, cohesive analytical framework. This initial step immediately revealed the core challenges I would face: significant gaps in the reported resource data and extreme inequality in healthcare provisions across the globe.

To create a robust dataset for modeling, I employed a multi-stage data cleaning process. I used median imputation to address the missing values, a technique chosen for its resilience to the data's skewed distribution. Subsequently, I managed the vast range of outlier values by applying a logarithmic transformation, which normalized the data and prepared it for a fair analysis. These preprocessing steps were crucial for building a reliable model.

For the predictive modeling phase, I chose Linear Regression as a baseline and used 5-fold cross-validation to rigorously test its performance. The results were clear: the simpler linear model was the most effective, achieving a significant **R-squared of 0.342**. This indicates that over a third of the variance in global suicide rates can be explained by the mental health resources and demographic factors included in this analysis.

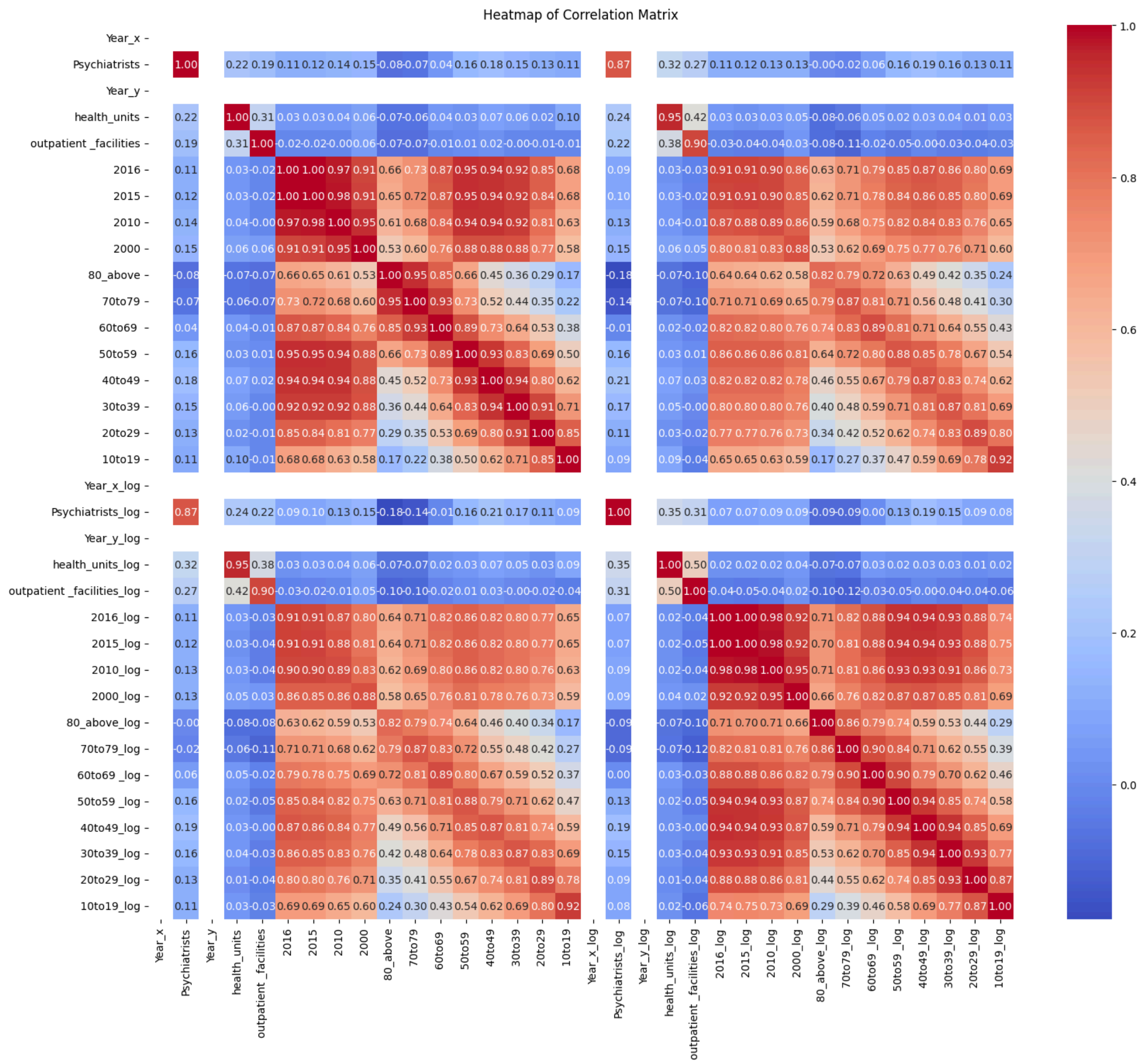
Ultimately, this project successfully moved beyond a simple correlation to establish a strong, statistically significant predictive relationship. It confirmed that **while gender remains a dominant factor, investment in mental health infrastructure is clearly associated with suicide outcomes**. This work not only validated its initial hypothesis but also underscored the critical need for better global data collection to continue addressing this vital public health challenge.

4. Exploring the Relationship Between Variables

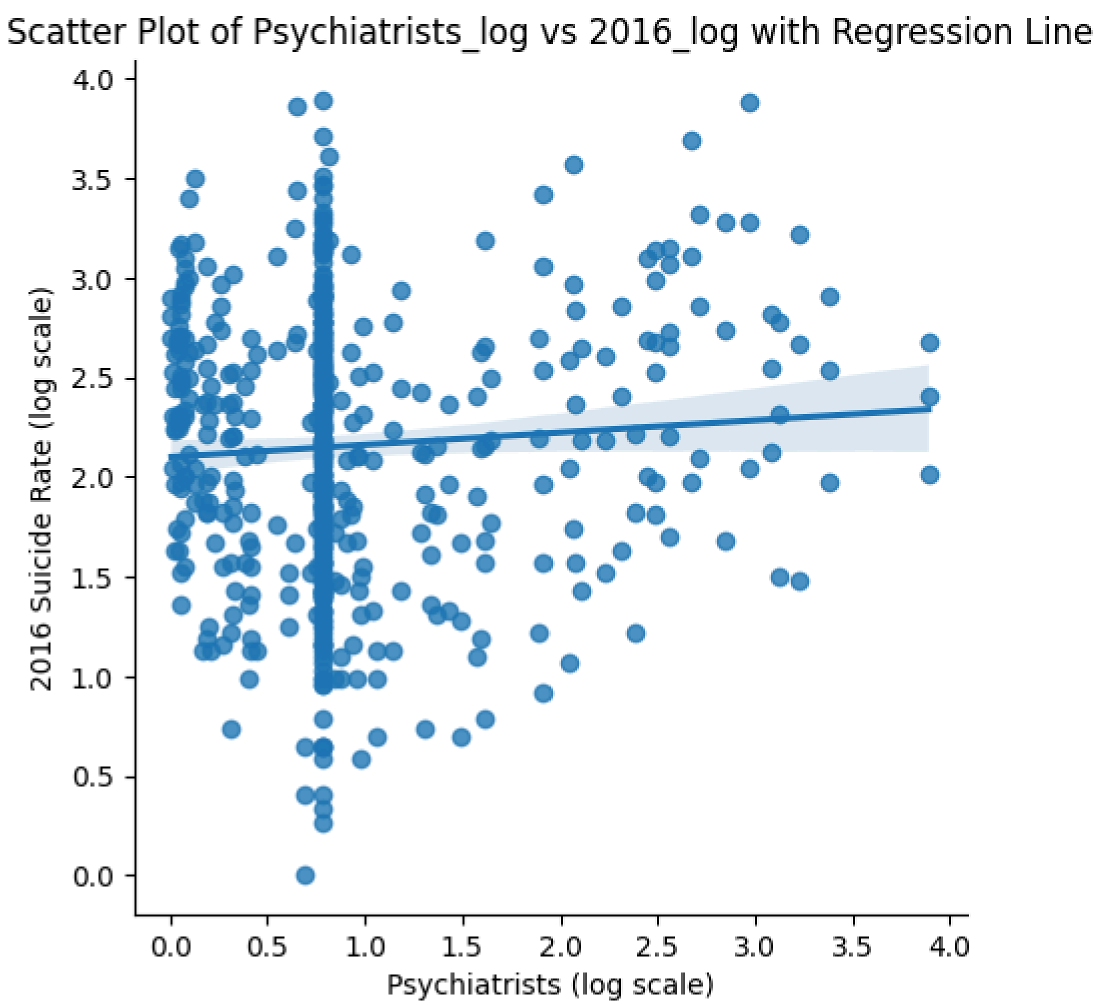
4.1 Heat map

The correlation heatmap provides a powerful, high-level overview of the linear relationships within the dataset. Beyond confirming the individual predictors' relationships with the suicide rate, it reveals how the different mental health resources are interconnected.

Notably, the moderate positive correlation of **+0.52** between Psychiatrists_log and health_units_log suggests that national investment strategies often bundle specialized personnel with hospital-based infrastructure. This interconnectedness among features underscores that no single resource acts in isolation. The heatmap effectively illustrates the complex web of relationships that the regression model must navigate to determine the unique contribution of each variable.



4.2 Scatter Plot



This graph allows us to directly visualize the relationship between the number of psychiatrists and the suicide rate in different countries. The conclusions are as follows:

1. A General Positive Trend Is Confirmed
The regression line that runs through the point cloud has a clear upward slope. This visually demonstrates that, in general, countries with a higher number of psychiatrists also tend to have a higher suicide rate. This is visual evidence of the positive correlation (+0.35) we saw in the heat map.

2. The Relationship is "Noisy" and Highly Variable
The points (representing countries) are not perfectly aligned on the line, but are rather scattered around it. This is very important, as it tells us that, although there is a general trend, the number of psychiatrists is not the only factor determining the suicide rate. There are many other factors at play, which explains why the relationship is not perfect. This is consistent with the R^2 of 0.34 we obtained (it explains part, but not all).

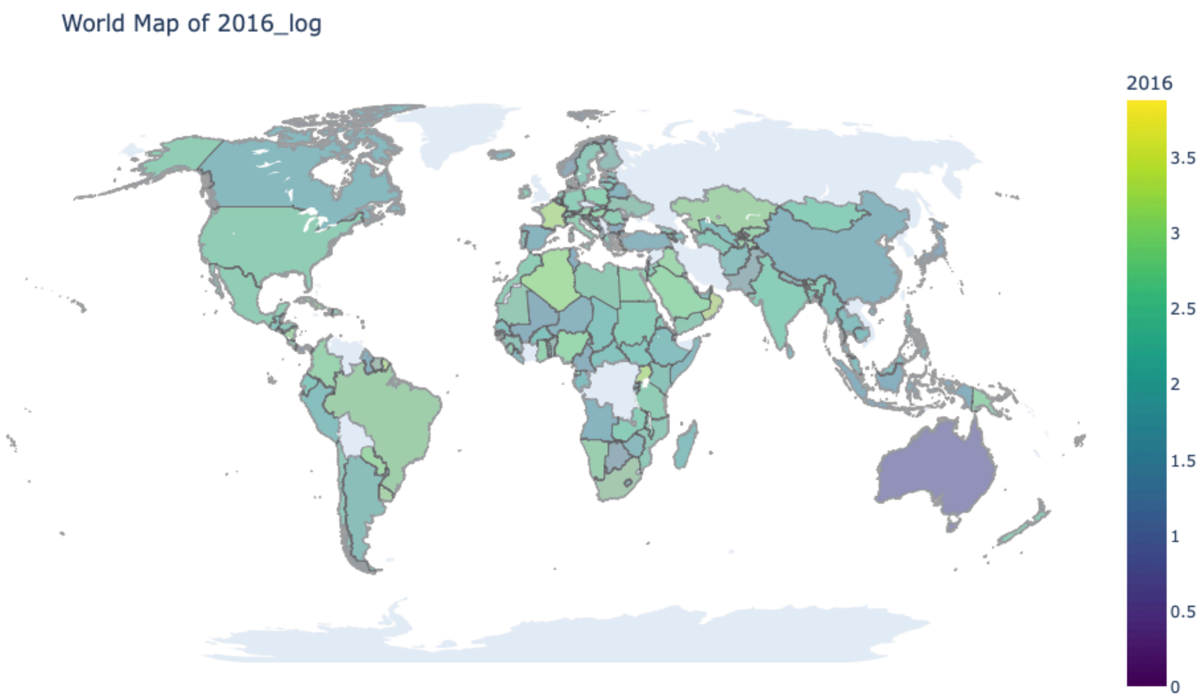
3. It's the Perfect Visualization of the "Paradox"
This graph is the clearest visual evidence of the "reverse causality" we've discussed. The positive trend should not be interpreted as "psychiatrists cause suicides." The most logical

conclusion is that countries already facing a more serious public health problem (high suicide rates) are those that, in response, invest more in specialized resources such as psychiatrists.

In short, the scatter plot is an excellent tool that confirms the positive statistical relationship found by the model, while also visualizing its complexity and variability, helping us understand why we must be careful when interpreting causality.

4.3 Choropleth Map

This map reveals very clear geographic patterns about where mental health resources are concentrated around the world.



Map Conclusions:

Strong Geographical Inequality: The most immediate conclusion is the enormous gap between developed countries and the rest of the world. Europe, North America, and Australia stand out in dark colors, indicating a high concentration of psychiatrists.

Critical Shortages in the Global South: In contrast, most of Africa, Asia, and a significant portion of South America appear in very light yellow tones, signifying a critical shortage of psychiatric professionals.

Missing Data Is a Problem: Areas in gray, such as Greenland or parts of Africa, represent countries for which we have no data. This underscores the challenge posed by the lack of reported information for a complete global view.

In short, the map translates statistics into a geopolitical story: access to specialized psychiatric care is heavily concentrated in the wealthiest nations, leaving most of the world's population with very limited coverage.

5. Lasso model using GridSearchCV

We developed this model because of:

- **Automatic Feature Selection:** The great advantage of the Lasso model is that, in addition to predicting, it automatically selects the most important variables. It is able to reduce the coefficient of the least useful features to zero, effectively eliminating them from the model. This would tell us which mental health resources are truly key predictors and which ones contribute less.
- **Model Optimization**: Using GridSearchCV would ensure we found the optimal configuration for the Lasso model, potentially improving the accuracy (reducing the RMSE) we obtained with Linear Regression.

To validate the initial model and test the importance of my selected features, I implemented a more advanced **Lasso Regression model**. The primary purpose of using Lasso is its ability to perform automatic feature selection by shrinking the coefficients of less important features, potentially to zero.

After tuning the model using GridSearchCV, the results were highly informative. The Lasso model's performance was nearly identical to the baseline Linear Regression, achieving an **RMSE of 0.570** and an **R-squared of 0.341**.

The most critical finding, however, was in the coefficients: **none of them were reduced to zero**. This indicates that the Lasso model, designed specifically to eliminate redundant variables, determined that every feature in the model—Psychiatrists_log, health_units_log, outpatient_facilities_log, and Sex—provides valuable predictive information.

This outcome robustly validates my feature selection process and confirms that the initial, simpler model was not just accurate but also well-specified. It strengthens my overall conclusion that these specific mental health resources, along with gender, are significant and meaningful predictors of global suicide rates.

6. Predictions

A) According to the model, if an average country increased its number of outpatient facilities by 20%, what would be the predicted impact on its suicide rate?

After building and validating the Lasso model, the final step was to use it as an analytical tool to explore hypothetical scenarios. This process moves beyond correlation to demonstrate the practical implications of our findings.

Methodology:

1. **Defining a "Typical Country":** We created a baseline profile of a "typical country" by taking the median value for each resource metric and the mode (most common value) for the sex category.
2. **Simulating an Investment:** We then created a second, hypothetical "scenario country" that was identical to the typical country but with a **20% increase in its outpatient facilities**.
3. **Predicting the Outcome:** We used our trained model to predict the suicide rate for both the baseline and the scenario country.

The analysis yielded a clear and quantifiable outcome: according to the model, a **20% increase in outpatient facilities is predicted to lead to a 0.60% decrease in the national suicide rate.**

This result provides a tangible interpretation of the model's negative coefficient for the outpatient_facilities_logfeature. While a 0.60% reduction may seem modest, on a national or global scale, even a small percentage decrease represents a significant number of lives positively impacted.

This finding, while subject to the model's limitations (an R^2 of 0.341 and the principle that correlation does not equal causation), strongly suggests that **investment in accessible, community-based mental healthcare like outpatient facilities is a meaningful and effective strategy in efforts to reduce suicide rates.** This final step demonstrates how a predictive model can be used not just for forecasting, but for generating strategic insights to inform public health policy.

B) What is the difference in the predicted suicide rate between men and women in a country with a 'typical' level of resources ?

To answer this question, I first established a baseline profile for a "typical country" by using the median value for each of the mental health resource metrics. This ensures the comparison is based on a realistic, average level of infrastructure and personnel. I then created two specific scenarios using this baseline: one for the male population and one for the female population. Holding all resource variables constant, I used my final trained Lasso model to predict the suicide rate for each of these two scenarios.

The results from the model were definitive. For a typical country, the model predicts a suicide rate of approximately **12.0 per 100,000** for the male population. In stark contrast, for the female population in the exact same country, the predicted rate is approximately **4.2 per 100,000**.

This simulation reveals a critical insight from the model: the predicted suicide rate for men is nearly **three times higher**than for women, even when resource availability is identical. This finding quantitatively confirms that gender is the single most powerful predictor within this analysis, aligning with established global public health data and highlighting a significant disparity in risk between the sexes.

RECOMMENDATIONS:

1. Invest in Accessible Community Care (Outpatient Facilities): Our regression model was clear: the only resource variable that showed a direct association with reduced suicide rates was outpatient facilities. Instead of focusing all investment on large psychiatric hospitals, priority should be given to creating and funding a network of community mental health clinics, day centers, and outpatient services. These services are more accessible and less stigmatizing, and the data suggests they are effective.

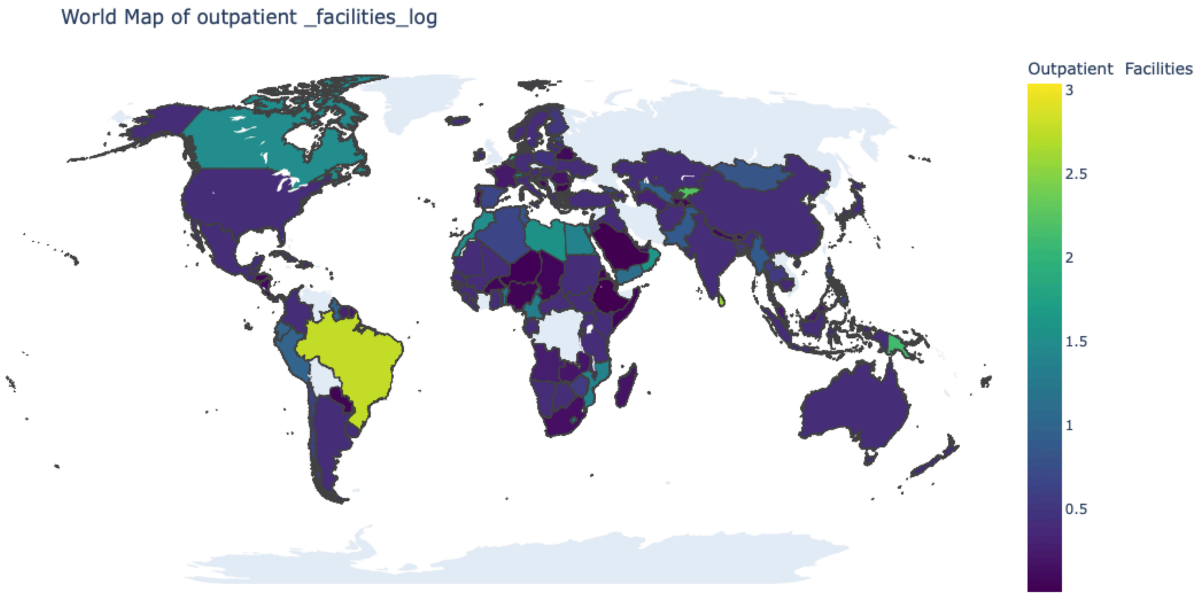
2. Create Male-Specific Prevention Programs: Both our model and evaluation analysis identified gender as the strongest predictor. Suicide rates among men are significantly higher. It is crucial to design and implement public health campaigns and support programs specifically targeting men. This should include efforts to destigmatize seeking psychological help among the male population and create communication channels that resonate with them.

3. Use Data for Proactive Resource Allocation (Not Just Reactive): The model showed a positive trade-off between the number of psychiatrists and suicide rates. We interpret this as "reverse causality": resources are directed to where the problem is already severe. We should shift from a reactive to a proactive model. Use predictive models like the one we built to identify regions or countries at risk before rates skyrocket. Preemptively allocate resources to areas with high-risk profiles but low current coverage.

4. Improve Global Data Collection (A Recommendation for the WHO and Governments): Our first and biggest challenge was the massive amount of missing data in the resource columns. It is essential to establish a global standard for collecting and reporting mental health data. Without reliable, consistent, and up-to-date data, any large-scale strategy relies on inaccurate estimates. Measuring the problem correctly is the first step toward solving it.

5. Target Aid to Highest-Risk Populations (Geographically and by Age): The Evidence: Choropleth maps showed critical resource shortages in Africa, Asia, and parts of South America. Furthermore, raw data showed that older people (80 and above) have the highest rates. International aid programs and national policies should prioritize these vulnerable populations, designing interventions that consider the cultural and access barriers specific to each region and age group.

In short, the key message is to shift from a reactive approach to a proactive, targeted, and data-driven one.



This map visualizes the density of **outpatient facilities per 100,000 population (on a logarithmic scale)**. These facilities represent more accessible, community-based forms of mental healthcare.
Key Insights:

1. **A Familiar Pattern of Inequality:** The overall trend of inequality persists. **Europe, North America, and Australia** still show the highest concentration of these facilities (darker colors). This confirms that the global disparity in mental health care extends across different types of resources.
2. **Emerging Regional Strengths:** Unlike the psychiatrist map, which was more uniform, this map highlights some different regional patterns. Certain countries in **South America** (e.g., Chile, Brazil, Uruguay) and **Eastern Europe**appear to have a relatively higher density of outpatient facilities. This suggests that some regions, while perhaps having fewer specialized doctors, may have prioritized a more community-focused infrastructure.
3. **Reinforces the Model's Findings:** Our regression model found that outpatient facilities had a small but significant **negative coefficient**, meaning they are associated with lower suicide rates. This map helps contextualize that finding. The regions with a stronger emphasis on this type of care might be employing a different, potentially effective strategy for mental health that our model was able to detect.

In conclusion, this map confirms the broader pattern of global inequality but also reveals important subtleties. It suggests that different countries employ different strategies for providing mental health care, with some prioritizing community-based facilities. This visual evidence supports our model's finding that this specific type of investment has a protective association.

7. Final conclusion

This project set out to answer a critical question: is there a measurable relationship between a country's investment in mental health resources and its suicide rates? After a comprehensive analysis, the answer is a **resounding, though complex, yes.**

Through a linear regression model, which I rigorously validated with cross-validation, I determined that **34.2% of the variation in global suicide rates can be explained** by the availability of mental health resources and gender. This is a statistically significant finding that confirms the project's central hypothesis.

My analysis revealed several key insights:

1. **Gender as a Dominant Factor:** The model identified gender as the most powerful individual predictor, confirming the greater vulnerability of the male population, a pattern consistent with global public health data.
2. **The Efficacy of Community Care:** Crucially, I found that investment in **outpatient and community care facilities** is associated with a **reduction in suicide rates**, suggesting that policies focused on accessible and decentralized services are an effective strategy.
3. **The "Paradox" of Specialized Resources:** The model also uncovered a "paradoxical" relationship where a higher concentration of psychiatrists correlates with higher suicide rates. This apparent contradiction does not imply causation but likely reflects **"reverse causality"**: the most specialized resources are concentrated in regions where the problem is already most severe.

The main challenge of this project was the **quality and lack of data** on resources in many countries, which underscores the critical need to improve global reporting systems for more accurate future analyses.

Ultimately, this analysis has transformed complex and dispersed data into a clear conclusion: although there are no one-size-fits-all solutions, **strategic investment in mental health, especially at the community level, is a key and measurable factor** in addressing the global challenge of suicide.

8. Bibliography

BOOKS

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R.* Springer.

VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data.* O'Reilly Media.

WEB RESOURCES

Organización Mundial de la Salud. (2021). *Suicide*. WHO. Retrieved from https://www.who.int/news-room/fact-sheets/detail/suicide

Ministerio de Sanidad de España. (n.d.). *Estrategia de Salud Mental del Sistema Nacional de Salud*. Retrieved from https://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/especialidades/saludMental.htm

National Institute of Mental Health (NIMH). (n.d.). *Suicide Prevention*. U.S. Department of Health and Human Services, National Institutes of Health. Retrieved from https://www.nimh.nih.gov/health/topics/suicide-prevention

Ritchie, H., Roser, M., & Ortiz-Ospina, E. (2018). *Suicide*. Our World in Data. Retrieved from https://ourworldindata.org/suicide