A Graphical Study of Changing Global Suicide Rates Across Time INM433 "Visual Analytics"

Andreas Ioannides

Abstract—This study is going to research the changing number of suicide cases globally over all the years that have passed, by applying specialist data visualization techniques to the data. The objectives of the study are determined by contextual research of the problem, and a thorough investigation of the data used. Previous research work was also consulted for inspiration on potential areas of focus, and relevant methods. Human intuition will be used at every stage of the study to aid key decisions about choices of methods used.The report will cover each area of the investigation procedure. These will be the prior research, data acquisition, planning of strategy, process of analysis, evaluation of results, and critique of work. Various data visualization techniques will be applied to make any trends and patterns in the data more visible, from inferences about the dynamics of global suicide cases will be made. Computational methods will be used in between graphing techniques to make it easier to choose the specifications of the data visualizations constructed. Results will interpreted at each stage using knowledge from scholarly material on the topic as well as previous research findings. In addition to this, an understanding of the different data visualization techniques as well as different data science principles will be used to draw statistical inferences from the data which will inform commentary on observed findings. The success of this work will be determined by how well the objectives of research have been met, and how far has this project has surpassed the work of previous studies.

1 PROBLEM STATEMENT

This research report studies global suicide cases, over time. The focus will be on the geographic and time variation of the data being studied. The suicide rate of all countries is considered a serious mental health concern [1]. There are various influential factors causing suicide, that will be thoroughly analysed in this work . Suicide rates change by time, in different locations because of evolution in society and an increase in wealth [2]. This study will evaluate to what extent different variables influence rates of suicide, by examining distributions within the relevant data. The questions that will be answered to address the main research problem are [3]:

- How has the number of global suicide cases recorded changed over time?
- What is the strength of the relationship between suicide count and economic prosperity metrics for all countries?
- Which categories of people have the greatest likelihood of committing suicide?
- To which extent is the number of global suicide cases influenced by the rate of substance abuse?

The data being studied here, has suicide cases for all countries, from 1980 to 2017. Each variable present in the data can be used to answer a different research question. Data for the human development index and yearly gross domestic product is associated to question 2. Furthermore, data for the sex, age, and generation of the sample is ties to question 3. Finally, data for yearly substance abuse of different countries addresses question 4. All pertinent variables can be compared to the recorded suicide counts of samples to check for relationships. The data has an equal proportion of categorical and numerical variables to analyse demographics of suicide cases and drivers of them. The data also has information on year and country for spatial and temporal analysis of distributions.

2 STATE OF THE ART

Pre-existing research on data for global suicide instances varies in its approach to the topic. Most studies focus on the health science aspect of it, whilst others may look at it from a data science angle. A plethora of research papers use identical or similar data to that being studied in this paper. Research paper [21] uses nearly identical data to that used in this study, and was compiled in a similar way as well, where data from multiple online sources, like Kaggle, were combined to prepare a detailed dataset. The work of the research paper investigates the suicide rates of samples across genders, age ranges, generations, and countries. It effectively analyses the temporal and categorical distributions of global suicide cases using time series graphs and proportional visualizations, respectively. It uses contextual knowledge of the topic in conjunction with predictive machine learning models to recognize patterns and trends within the global suicide data. Research article [19], like this research study, uses data from the world health organization for global suicide cases, which includes demographic and yearly data for worldwide suicide cases. It implements discrete data visualizations and time series analysis of the data, like research article [21]. However, it additionally aggregates years and cohorts of data studies into fewer categories for producing simpler visualizations, as well as converting the data on suicide counts to rates. Research paper [2], uses similar but not identical data to that used in this research study since it omits information on the different cohorts of global suicide cases across samples. Its analysis focuses on both the temporal and spatial aspect of global suicide cases whilst ignoring the characteristics of case subjects. The research paper uses categorical data visualizations and regression lines to present important findings. The use of regression here helps capture the relationship between the number of suicide cases and time. Finally, research paper [1], has data of the same scope as this project but with use of different categorizations. This report uniquely uses tabular representations of the data, in addition to spatial and temporal visualisations for its analysis, which is not effective. Variables like suicide success rate and mental health history of sample participants are not studied. The reason for this being that the volume of data being used in for analysis is very large, and incorporation of additional variables would mean an increase in computational load leading to inefficiency [3]. The previous research projects had limited use of sophisticated data analysis methods such as inferential and descriptive statistics to inform choices of visualizations. Neither did they have multidimensional data visualization techniques to incorporate greater detail into the graphs being plot. This is because the previous work consulted, focused too much on the psychology aspect of global suicides, instead of the data science side. This project will calculate relevant descriptive and inferential statistics at each stage to inform choices for methods of data visualization [3]. Furthermore, this project will combine quantitative and qualitative variables of the data to produce multidimensional intricate visualizations [16]. This will be achieved by using advanced data visualization packages like Bokeh, Plotly, Altair, Matplotlib, and Seaborn [4]. R will also be used for modelling of relationships, between variables to inform the choices of graphs used. The visualizations produced will then be looked at in depth to see which important information can be deciphered.

3 Properties of the Data

The original data set used for this research investigation was obtained from Kaggle, an open-source online data repository. It was assembled from four other reliable datasets of Kaggle, UNDP, WHO and World Bank. The data variables include; Country, Year, Sex, Age, Suicide Count, Population, Age Standardized Suicide Rate, Human Development Index, Yearly Gross Domestic Product ($), and Gross Domestic Product per Capita ($). The 'Age Standardized Suicide Rate', is the suicide count per 100k population sample, calculated to mitigate the issue of variance in age. The 'Generation' refers to a group of people who have shared life experiences. The 'Human Development Index' variable had too many null values, so was removed. The data for 'Human Development Index' and 'Substance Abuse Rate' was extracted from other Kaggle data sets and attached to the initial dataset using outer join methods. All rows containing null values in the resultant data table were then removed. Moreover, Principal Component Analysis was also completed on the data to potentially reduce dimensionality, by aiding the feature selection process. The scree plot produced from this dimensionality reduction process had shown that some of the principal components explain a greater amount of the variance in the data than others. The corresponding variables to these principal components were included more frequently in the data visualizations constructed during the study. Nevertheless, the cumulative proportion of explained variance had increased with the number of principal components included, so no data variables were disregarded.

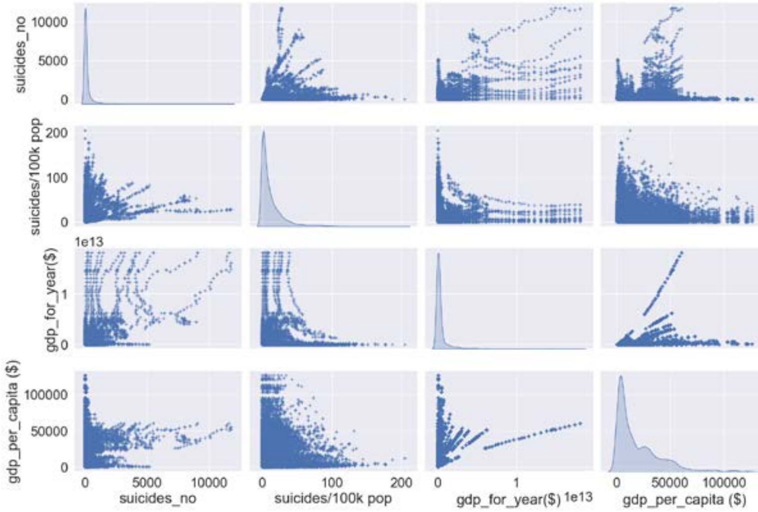

Figure 1: A pair plot containing the distribution of key variables being studied in this research project.

The pair plot was created for performing some exploratory data analysis on the important features of data. It can be seen here that many of the variables are right skewed, meaning that most instances are small in magnitude. It was recognised from the pair plot that the data contains a lot of outliers, which were removed before each graph was plot. Additionally, the fact that most of the quantitative variables present in the data were right skewed was taken into consideration when interpreting obtained during the analysis stage of the project.

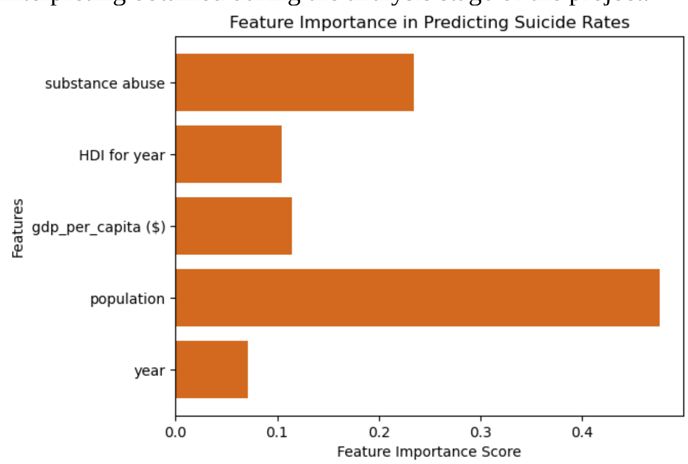

Figure 2: A bar chart showing feature importance of numerical variables.

This bar chart of feature importance displays the most important variables for predicting suicide rates. Here, population has the greatest importance, due to its large score, meaning that there is a strong positive correlation between the population and suicide rates. Rate of substance abuse, gross domestic product per capita, and human development index also play a key role in determining suicide rates. This demonstrates that despite financial success influencing the rate of suicide, other issues are also important. Also, the year variable has the lowest feature importance score, because there is minimal correlation between the suicide rates and year. This graph amplifies important variables for use in constructing visualizations.

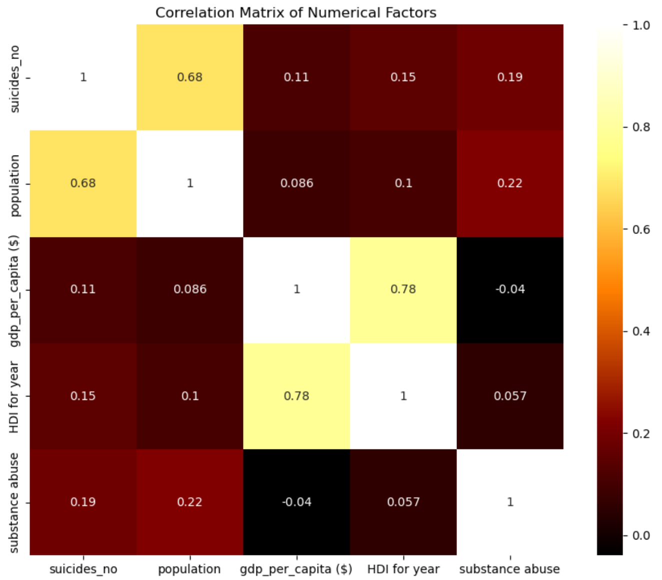

Figure 3: A correlation matrix of all numerical variables.

The above heatmap shows the correlation between suicide rates and the socioeconomic factors of our dataset. The heatmap is color-coded using the color scale placed right next to it for the purpose of distinguishing the different types of correlation. The heatmap shows us that there is a strong positive correlation between population and number of suicides (0.68) which indicates that countries with larger populations have higher suicide rates. GDP per capita has a strong positive correlation with HDI (0.78) showing that people with higher incomes have better human development. Finally, the substance abuse variable is not correlated with the other variables. This helps the interpretation of results obtained from other analysis procedures.

4 ANALYSIS

4.1 Approach

The objective of this research project is to discover the different factors initiating change in global suicide rates, by recognising patterns and trends in visualizations comparing suicide cases to associated measurements, across many different cohorts sampled. The strategy used to achieve this is outlined in the diagram below. The aim here is to iteratively investigate and organize the data to produce intricate visuals. Furthermore, the goal is to also obtain information that aids the selection of visualization methods. Furthermore, the process of creating the visualizations themselves is also outlined in this section. Finally, the procedure of deriving specific findings from the visualization is also included. The results from each, and every stage of the entire process will used to draw conclusions about the study and fulfil important objectives. This is because all findings explain each other as well as addressing the problems being focussed on in the study. Many different methods are used at all stage of the process to carry out important procedures. For example The distance function available in all programming languages used was applied to inspect for outliers present within the data, as well as many clustering techniques.

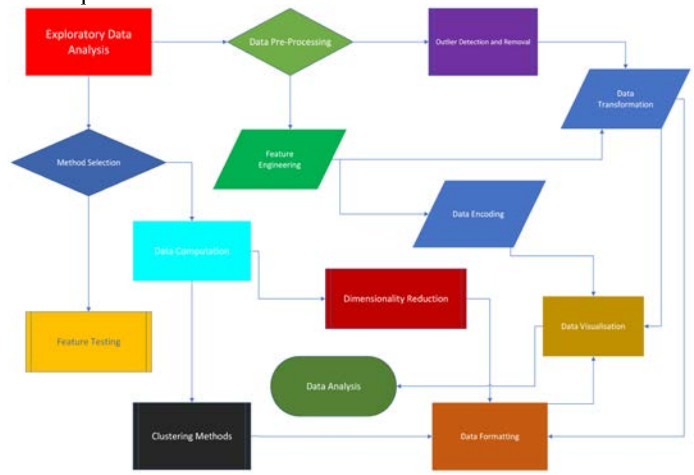

Figure 4: A flowchart diagram, showing the plan used for analyzing the sourced data.

The above flow chart drawing shows the comprehensive process used for analyzing data to answer research questions. The approach used was inspired by previous literature on methods of data visualization. book [3]. Each stage was undertaken meticulously, and each stage had built on the results of the previous one, to achieve the results. The different stages of the process were not completed in strict chronological order. Firstly, exploratory data analysis of the data was completed to aid in later stages of the process. The outcomes of this are shown in the 'Properties of Data' section of the report. Then, using the results of Exploratory Data Analysis, suitable methods of visual analysis and data preparation are selected [15]. Following this, features of the dataset are engineered, and the data is modeled using computational methods. The process of feature engineering is shown in the 'Properties of Data' section of the report. The visuals selected are based on the distributions of data discovered across all variables, as well as relevant subject knowledge of the suitability of different visualizations to certain data types. Then feature testing was performed on all numerical variables of the data, to check for importance. The data was modeled using principal component analysis and hierarchical clustering. The results output from these computational methods were assessed, to inform the selection of good data visualization methods, and referred to later when interpreting results. Before any graphs are plotted, the data is restructured, to make it easier to use. The data restructuring involved formatting and transforming the data into certain columns to be digitally processible. Following the modeling, cleaning, and engineering of the data, detailed visualizations are created using the relevant features for each scenario. At each stage of the analysis process, intuition is used to see whether the methods of visualization chosen are suitable to the data set and research focus. This decision-making process used computational outputs, context, and subject knowledge. The goal was to choose visualizations that represented the distribution of the data best. Lessons were learned from methods used in previous studies referred to in the 'State of The Art' section. The focus when selecting graph types, was their representation of detail and interpretability. In addition to this, the added dimensionality of chosen graphs was used to explain important findings through deriving trends and patterns. The results of these were analyzed to answer the selected research questions.

4.2 Process

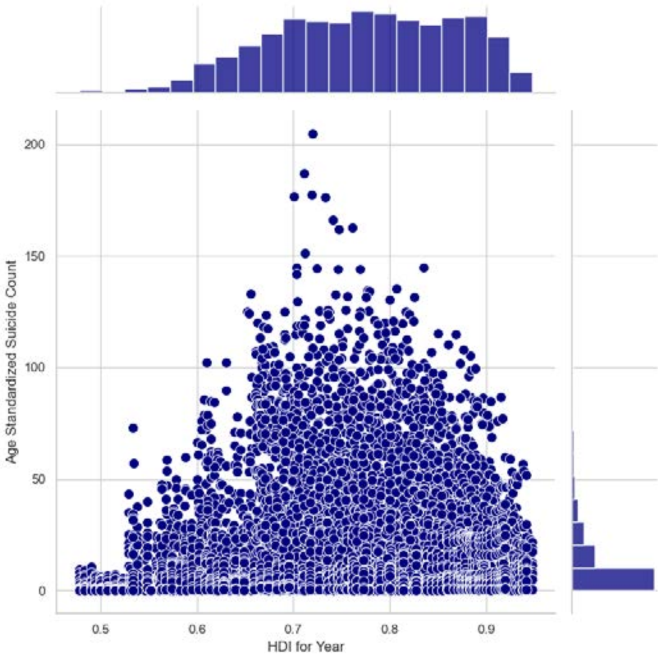

Figure 5: A joint scatter plot and histogram comparing the age standardized suicide count and human development index of different countries.

As seen in the above graph there is a strong correlation between the age standardized suicide rate and yearly human development index for all countries sampled, as seen in the scatterplot component of the graph. As there is an increase in the human development index of countries being sampled, the age standardized suicide count also increases. However, it is visible in the attached histograms that both variables of data are skewed to opposite sides of the spectrum. Meaning that, the effect of both variables on one another cannot be considered very big. The age standardized suicide count is left skewed, whilst the yearly human development index is right left skewed. This means that the data has more countries with a high human development index, whilst also having more countries with a low age standardised suicide rate. This means that despite the trends or patterns found in the scatterplot, it can be said that countries that are more economically developed usually have a lower suicide rate. Other equally important metrics of countries economic success will be compared against the suicide count later in the investigation.

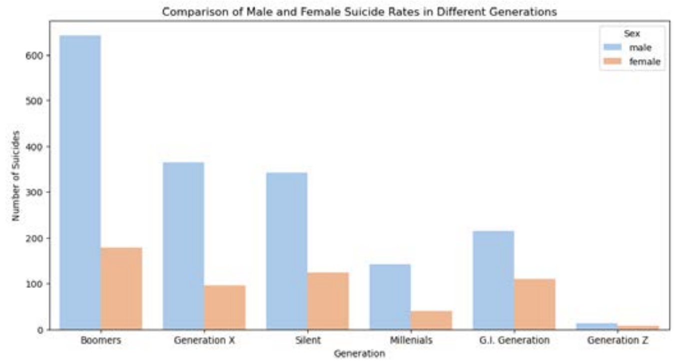

Figure 7: A grouped bar graph showing the total number of suicide cases amongst males and females across generations.

The bar chart shows the total number of suicides for both males and females, across different generations. We can see a decrease in the number of suicides for males, across each generation, except for the GI Generation where there is an increase in suicides. Considering the general trend of female suicides, there are some fluctuations of higher and lower suicides across each generation. As is evident from the graph, the Boomer generation has the highest number of suicides for both males and females, but a significantly higher number of suicides for males. For all the generations presented in the graph, the number of suicides for males. Is higher than the number of suicides for females. We can see that Generation Z, which is the youngest generation, has the least number of suicides for both males and females, in comparison to the rest of the generations. The overall conclusion of this graph is that through the years, global living conditions have improved, with more people having access to good healthcare, necessities, and mental health resources, which led to a decrease in the number of male and female suicides.

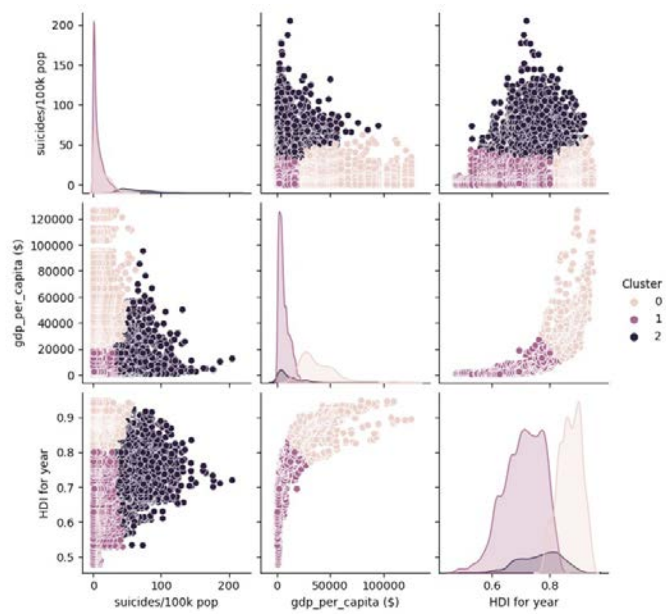

Figure 6: A pair plot of data clustered using the K-Means method.

the above graph is a pair plot representing the results of cluster analysis to visualize the relationships between gross domestic product per capita, human development index, and age standardized suicide rate, for zero, one, and two clusters within the dataset. The scatter plots display a complex correlation between the different economic factors. The scatter plots that compare suicide rates with GDP per capita, as well as the HDI, show a non-linear relationship, indicating that a rise in economic growth and development, does not necessarily correlate with lower suicide rates. Concerning our cluster analysis, cluster 0 represents the countries with lower GDP, lower HDI, and varying suicides. Cluster 2 captures nations with higher income levels as well as higher GDP per capita and HDI. Cluster 2 also shows a range of suicide rates indicating again that higher income and human development do not lead to lower suicide rates. The kernel density estimation plots of the graph show us that most countries experience low levels of suicide rates, and very few countries experience high suicide rates. This is because the histogram is rightly skewed. On the other hand, GDP per capita and HDI have similar skewness indicating that very few countries have either high GDP, or high HDI.

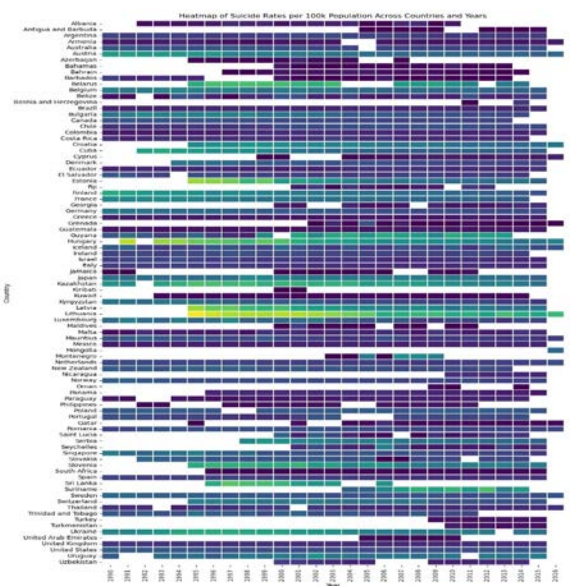

Figure 8: A categorical heatmap showing the total number of suicide cases across countries and years.

The heatmap shows the age standardized suicide rate across the countries and the time frame of the dataset. The heatmap is color-coded using the scale placed right next to it on the diagram. Each the heatmap's rows represents a different country, whilst the columns represent different years. From the heatmap, the country that stands out with the highest frequency of suicide rates was Lithuania during the late 1990s. Countries such as Hungary and Estonia also had relatively high amount of suicide rates during the early 1990s and mid-1990s respectively, in comparison to the rest of the global countries. As is evident from the heatmap, most of the countries have a low to middle suicide frequency. The high suicide frequency across Hungary and Lithuania during the 1990s could be because most Eastern European countries had undergone economic transformations then, involving a variety of social changes that had resulted in political instability, which affected people's mental health. The general consensus here is that global situations which cause uncertainty inadvertently lead to an increase of mental health problems amongst its populace, which results in high suicide rates. Factors which explain this phenomenon are investigated later on in the report.

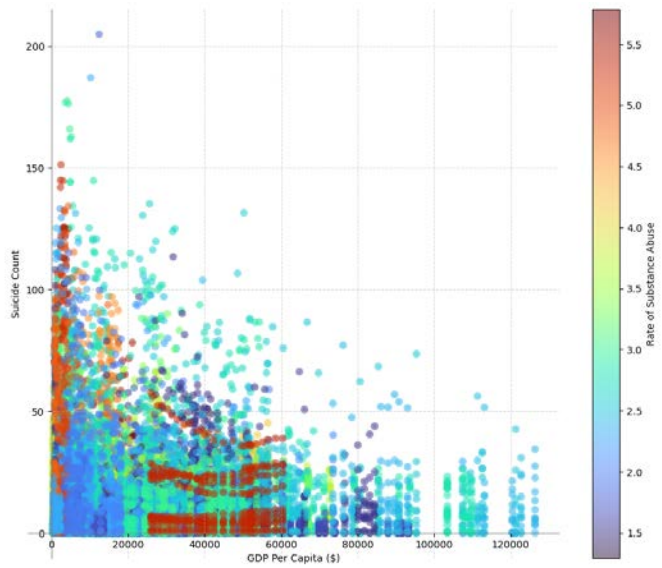

Fig 9: A scatterplot of suicide count against GDP Per Capita ($) coloured by rate of substance abuse.

This coloured scatterplot shows the effect of substance abuse on the rate of change in the number of suicides in conjunction with the increase in GDP Per Capita measured in dollars. Countries with a higher substance abuse rate have more cases of suicides, and a lower gross domestic product per capita. Countries with lower rates of substance abuse generally have a higher gross domestic product per capita and a lower suicide count. This is what can be inferred from the graph presented above. However, there are plenty of unusual data points present in this plot which show otherwise. Here we can see that a lot of countries with low or medium substance abuse rates have suicide counts and gross product per capita, ranging from high to low. The other dimension of the data shows that usually countries with lower economic prosperity have greater rates of substance abuse. Reasons for this could be lack of education on the side effects of substance abuse or lack of support with substance addiction. The results found here are quite significant given the difference in scale between the two variables compared. If the graph was constructed using normalized or scaled data, the relationship between the two variables would be visibly stronger. Later in the work, a regression model is fitted to similar variables representing the same phenomenon to quantify the relationship between wealth generation and suicide rate more clearly.

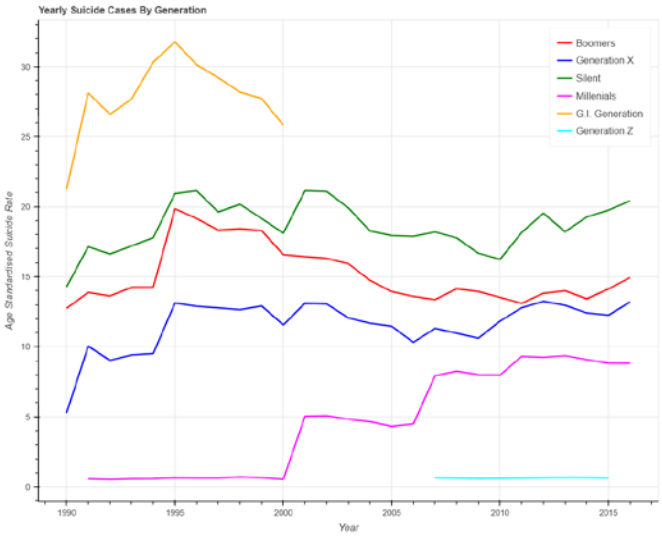

Figure 10: A time series graph showing the change of change in the number of suicide cases for multiple generations of people.

The rate of change in the number of suicide cases across the years is almost the same for all 6 generations of people. However, some of the generation categories show different patterns of change. There is only an increase in the number of suicide cases for millennials between 2000 and 2015, where there are zero suicide cases before that. For the G.I. Generation there is a sharp increase in the number of suicide cases from 1990 to 2000 and then no data available for it after. Additionally, the Boomer generation has a spike in the number of suicide cases in the year 1995, followed by a slow decrease. This is seen as unusual when compared to the Silent and Generation X groups which have mostly gradual rates of increase and decrease for rates of suicide, with the occurrence of occasional spikes in some years. Also, though they follow the same pattern, the time series graph for Generation X is placed above that of the G.I. Generation, meaning that they have more overall suicide cases. The above time series graph is spatial and temporal because the size of data being sampled for the construction of this visualization is reduced by grouping all the rows of data using the country-year key index. Time series modelling could have enabled calculations for averages and rates of change to be computed for a more in-depth analysis. This graph does not show the total suicide count across the age cohorts. This is investigated later on in the project using other methods.

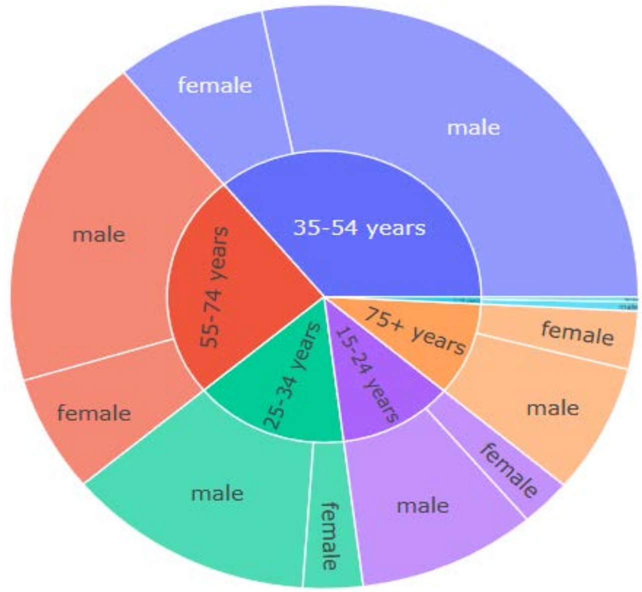

Fig 12. A colourful sunburst diagram of suicide cases across gender and age range.

The graph shown in the image above is a sunburst diagram representing the spread of global suicide cases across the two categories of age and gender. The proportion of males and females in each age group is seen to be very similar, where males are seen to commit suicide more frequently than females. However, it is visible that suicide happens more frequently for people in certain years of their life, in comparison to others. The highest number of suicide cases present in the data is for people between the ages of 35 and 54 years, followed by people between the ages of 55 and 74 years. The least number of suicide cases occur for people between the ages of 5 and 14 years, such that it is not visibly represented on this sunburst chart. People within the age ranges of 15-24 years and people above 75 years have almost the same number of suicide cases. These age ranges can be considered the oldest and youngest years of consciousness for humans. People, between the ages of 25 and 34 years have the third most suicide cases amongst them. It can be concluded from this that people in the middle years of their life are the most likely to commit suicide, whilst those at the beginning or end of their life are not as likely to do so.

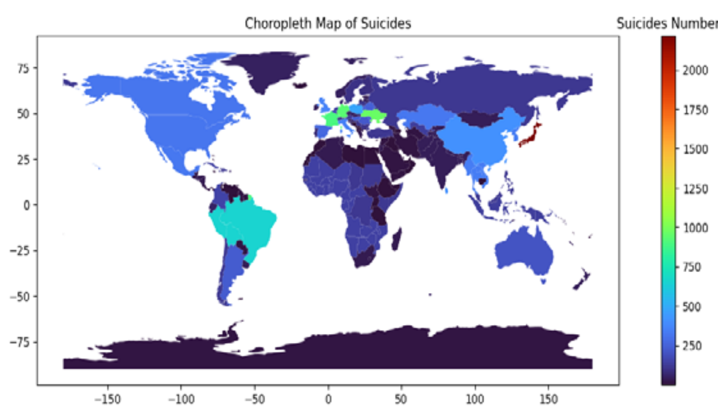

Fig. 11. A mapping of total global suicide cases using color

This graph shows a world map, with all countries colored using a scale determined by the total number of suicide cases across the entire time frame. This map graph was created using both the geopandas and matplotlib packages from python. The value of this graph is that the knowledge gained from this can inform the choices of other graphs used as well as the different inferences made from them. It can be seen in this graph that wealthy countries have had a greater number of suicide cases in comparison to poorer countries. This could be because of the 'complexity of life' present in all these advanced countries. However, it can also be because of the lack of data available for poorer countries. There is a limit to the amount of information we can draw from this choropleth map because of the absence of details on the categorical variance and temporal distribution of the data. However, this graph does help inform the choices of graphs used to visualize other aspects of the data elsewhere in the project, by providing insight into the different factors affecting global suicide rates. Other graphs present in this report will show these elements of the data to give a more holistic view of the study. The graph shows that are near each other within regions, have similar counts of suicide, because of spatial dependence. This graph informs the choices of other data visualization methods used in the investigation.

4.3 Results

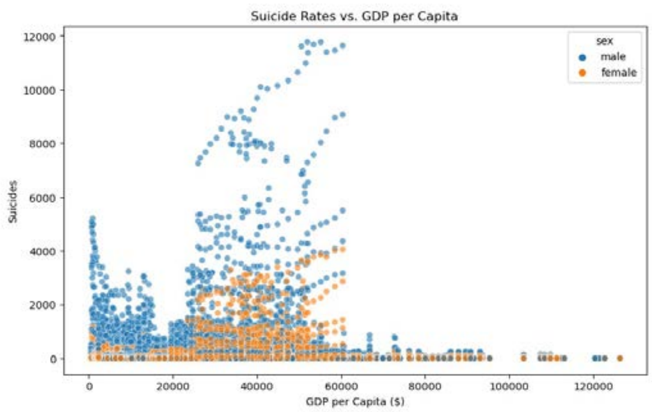

Figure 13: A coloured scatterplot comparing the number of suicide cases against gross domestic product per capita.

The scatter plot shows us a correlation between the number of suicides and GDP per capita, which is more evident in males. Clustering shows that countries with lower GDP per capita have fewer suicides. The male population has higher variability and GPD. Outliers in males show that there are additional factors that lead to suicides.

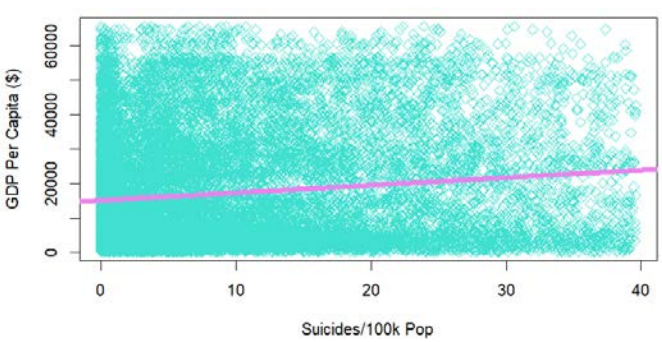

Figure 14: A regression plot of the Age Standardized Suicide Rate against the GDP Per Capita ($).

The regression model above is not a good fit for the data points comparing suicide and economic prosperity. This contradicts the results of earlier investigations in the analysis process. The maximum and minimum residual statistics are not symmetrical, and the upper and lower quantiles are not equidistant. So, therefore, the results are not reliable.

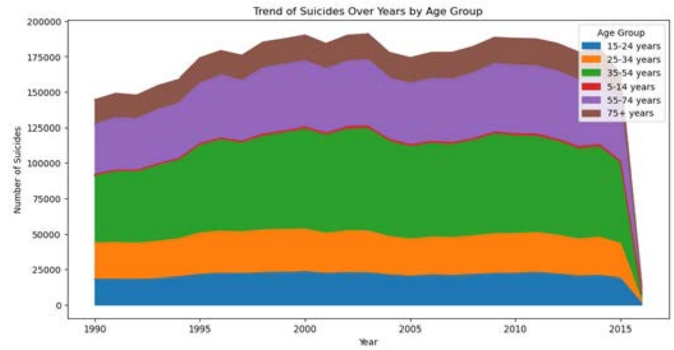

Figure 15: A stacked area graph showing the number of worldwide suicide cases across the years.

The stacked area chart above shows the distribution of the total number of suicides in the years of 1990-2015. The year 2015, shows a decline in suicide cases for all age groups. Ages 35-54 have the highest total suicide count. Ages 5-14 have the least total suicide count, due to factors such as parental control, and lack of self-awareness. To conclude, sharp spikes occur in the total suicide counts, in the years of 2002 and 2015, representing a steep increase and decrease respectively.

5 Critical Reflection

The outcome of this research project has led to introspection on many of its approaches. The success of this project was that it could represent all variables of data using a limited number of graphs. This was done by using detailed and intricate graphs consisting of multiple dimensions to represent the simultaneous effects of different variables on each other. Hence, gaining a deep insight into the contributing factors influencing global suicide cases. A weakness of this study was its lack of explanation for the findings it has discovered. This was because the amount of reliable data available for the study did not include enough key variables for use in explaining relationships between other features of data, derived during the research. For example, data on the mental health history of those involved in suicide cases could not be compared to the cohorts sampled in the study, to examine the effects of having different life experiences. Neither could the abundance of healthcare facilities be compared to the statistics on economic prosperity, to assess the impact of poverty. The intuition used at each stage of the investigation was very important because some results received from computational calculations are subject to interpretation. For example, the dendrogram produced from hierarchical clustering broke into 3 or more clusters depending on the height it was cut. Thus, subject knowledge of the process was needed to decide which height range to consider. Apart from computational methods, human intuition was also used when deciding on the dimensions and types of graphs to plot. Inferences were made from simple graphs to inform the selection of variables for more complicated graphs [3]. The choice of visual representations played an important role in effectively extracting information from the data to answer the research questions selected. This is because different graphs showed different aspects of the distribution of the data being studied. Some had focused on frequency or average, whilst others looked at the strength of the relationship between selected variables. Each of these was important in addressing the different focuses of the research problem. The graphs representing the frequencies and averages of suicide cases were very important in determining the distribution of suicide incidences between categories and locations. The relationship graphs were key in determining the level of impact all variables had on the incidence of suicide cases, in general, and over time. The global coverage of the dataset meant that smaller details, which are usually considered in the study of suicide cases, could not be investigated. For example, other studies to this one had looked at the prevalence of reachable mental health support, which was not included in this study. The entire work was completed using the programming languages RStudio and Python, however, data visualization software with a graphical user interface such as Orange, Tableau, or Microsoft Power BI could have been used to improve the efficiency and effectiveness of graphing techniques [11]. For example, it was difficult to produce static data visualizations combining data on the spatial and temporal distributions using coding, because of issues with adding dimensionality to visualizations. This could not be done using coding because of the risk of potentially reducing the interpretability of the visualization. To conclude, the findings from this research have added to those found in previous studies. Hence, this project can be considered a success because of the additional knowledge it has helped to discover. In this regard, some of the components of this research study are novel because they use specialist data analytics techniques not used in other formal research studies. Previous technical investigations on the selected dataset are limited to data science blog posts online.

Table of word counts

Problem statement	312.5; 301
State of the art	625; 606
Properties of the data	625; 590
Analysis: Approach	625; 602
Analysis: Process	1875; 1863
Analysis: Results	250; 242
Critical reflection	625; 604

References

[1] "who-msd-mer-19-3-eng.pdf." Accessed: Dec. 26, 2023. [Online]. Available: https://platform.who.int/docs/librariesprovider20/default-document-library/resources/who-msd-mer-19-3-eng.pdf?sfvrsn=1fef22be_2.

[2] M. Ilic and I. Ilic, "Worldwide suicide mortality trends (2000-2019): A joinpoint regression analysis," World J. Psychiatry, vol. 12, no. 8, pp. 1044–1060, Aug. 2022, doi: 10.5498/wjp.v12.i8.1044.

[3] N. Andrienko, G. Andrienko, G. Fuchs, A. Slingsby, C. Turkay, and S. Wrobel, Visual Analytics for Data Scientists. Springer Nature, 2020.

[4] C. O. Wilke, Fundamentals of Data Visualization. Accessed: Dec. 28, 2023. [Online]. Available: https://clauswilke.com/dataviz/

[5] Y. Chung et al., "Role of visual analytics in supporting mental healthcare systems research and policy: A systematic scoping review," Int. J. Inf. Manag., vol. 50, pp. 17–27, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.04.012.

[6] "2022 National Veteran Suicide Prevention Annual Report, VA Suicide Prevention, Office of Mental Health and Suicide Prevention, September 2022".

[7] L. Ongeri et al., "Community suicide rates and related factors within a surveillance platform in Western Kenya," BMC Psychiatry, vol. 22, no. 1, p. 7, Jan. 2022, doi: 10.1186/s12888-021-03649-6.

[8] K. Thomas and D. Gunnell, "Suicide in England and Wales 1861–2007: a time-trends analysis," Int. J. Epidemiol., vol. 39, no. 6, pp. 1464–1475, Dec. 2010, doi: 10.1093/ije/dyq094.

[9] C. Chen, W. Härdle, and A. Unwin, Handbook of Data Visualization. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-33037-0.

[10] I. M. Tøllefsen, E. Hem, and Ø. Ekeberg, "The reliability of suicide statistics: a systematic review," BMC Psychiatry, vol. 12, no. 1, p. 9, Feb. 2012, doi: 10.1186/1471-244X-12-9.

[11] P. Varnik, "Suicide in the World," Int. J. Environ. Res. Public. Health, vol. 9, pp. 760–71, Mar. 2012, doi: 10.3390/ijerph9030760.

[12] C. O. Wilke, Fundamentals of Data Visualization. Accessed: Dec. 28, 2023. [Online]. Available: https://clauswilke.com/dataviz/

[13] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual Analytics: Definition, Process, and Challenges," Mar. 2008, doi: 10.1007/978-3-540-70956-5_7.

[14] W. Cui, "Visual Analytics: A Comprehensive Overview," IEEE Access, vol. 7, pp. 1–1, Jun. 2019, doi: 10.1109/ACCESS.2019.2923736.

[15] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 2nd ed. in Springer texts in statistics. New York: Springer, 2002.

[16] S. Bachmann, "Epidemiology of Suicide and the Psychiatric Perspective," *Int. J. Environ. Res. Public. Health*, vol. 15, no. 7, p. 1425, Jul. 2018, doi: 10.3390/ijerph15071425.

[17] J. Bertolote and A. Fleischmann, "A Global Perspective in the Epidemiology of Suicide," Suicidologi, vol. 7, Jun. 2015, doi: 10.5617/suicidologi.2330.

[18] S. R. Midway, "Principles of Effective Data Visualization," Patterns, vol. 1, no. 9, p. 100141, Dec. 2020, doi: 10.1016/j.patter.2020.100141.

[19] T. Junior, A. Magdy, and S. Lemuels, "Detecting Suicide Rate By Using Network Science Specifically Data Science and Machine Learning Used. Developed by TABLE OF CONTENTS," Dec. 2021.

[20] A. K. Das, S. Dutt, and S. Chandramouli, "Machine Learning," Mach. Learn..