# Visualisations of Suicide Rates

Md Faiyam Islam
Student ID: 490470604
The University of Sydney
Sydney, Australia
fisl5259@uni.sydney.edu.au

## I. NATURE OF THE DATA

Suicide is a global phenomenon which is recognised by an international public health problem by the World Health Organisation (WHO) [1]. Understanding the dynamics of suicide rates will provide valuable insights into the mental health and well-being of populations, helping to identify risk factors, design targeted interventions, and monitor the effectiveness of prevention efforts. The dataset analysed in this report is essential for saving lives, raising awareness, and addressing the complex societal issues surrounding suicide.

The "Suicide Rates Overview 1985 to 2021" dataset from Kaggle is a comprehensive collection of data related to suicide rates and factors contributing to suicide from 1985 to 2021 [2]. The dataset contains information about countries and various demographic and socio-economic variables such as gender, age, population, GDP, and generation. This dataset is imperative for studying trend, correlations, and patterns in suicide rates over time, to conduct research and identify potential risk factors and inform suicide prevention efforts. Suicide estimates are generated from death certificate data, where deaths are classified under death code for 'intentional self-harm' in the International Classification of Diseases (ICD) [3]. Individuals who had not intended to die may not be classified as a suicide based on a certain country's legal definition, hence there could be bias in the data.

To further evaluate the nature of suicide rates in this dataset, there are various factors contributing to bias which we need to take into consideration. In a 2012 systematic review spanning 46 years and involving 31 studies, it was revealed that 13 studies indicated underreporting of suicide, ranging from 5-10%, while among the remaining 18 studies, 52% reported over 10% underreporting and 39% reported over 30% underreporting [4]. Social stigma and legal implications are contributing factors towards underreporting of suicide rates as different countries may not choose to disclose true causes of death, resulting in incomplete records. Socioeconomic disparities can also influence suicide rates and the misclassification of suicides as lower income areas may not have access to mental health services compared to developed countries. It is paramount for individuals and organisations to test suicide rates for misclassification, mainly in populations with proportionately undetermined and accidental death rates [5].

## II. CONSUMERS OF THE DATA

### A. *Public Health Officials*

Data involving suicide rates can assist officials to identify trends and patterns across different demographics, generations, geographic regions across time. Public health officials would be amongst the primary uses of this dataset due to their role in analysing information to pinpoint at-risk populations across highly relevant areas and factors contributing to suicide rates. The benefit of determining these factors can guide public health officials to develop and target prevention and intervention programs of suicide. Additionally, social, economic, and healthcare are variables to consider for officials who consume this data to enable and advocate for policies and resources for prevention.

Moreover, public health officials consume this data through visualisations in empirical studies which showcase trends and patterns. A thorough analysis in statistics, visualisations and data analytics will allow them to communicate findings to various stakeholders, including policymakers, healthcare professionals and in general the public. Evidence-based decision making with the help of data analysis is powerful for officials to address the statistically significant variables contributing to suicide rates in society. Suicide rates should focus on contemporaneous factors instead of a probabilistic outlook, hence providing evidence through data is crucial for public health officials [6]. For example, considering the effectiveness of machine learning models, conducting a regression analysis on the correlation between lack of access to mental health services and higher suicide rates in specific regions can be used to advocate for officials to target these more vulnerable regions.

### B. *Mental Health Advocates*

Advocacy roles for mental health workers have shifted from psychiatric hospitals to community services, their consumption of data will allow them to understand the scope of the problem and its impact on society [7]. Mental health advocates can be represented in non-profit organisations, individuals and communities, education campaigns and online platforms. They play a similar role in consuming data compared to public health officials, to analyse data, identify trends in vulnerable populations which informs their advocacy efforts. Mental health workers are typically focused on advocating for improved mental health care, reducing stigma, and providing direct support through resources across individuals and communities. Their role is more related to directly supporting affected individuals and providing them a voice to share experiences rather than using data to implement policies and interventions at a population level. Despite advocates playing a more supportive role in reducing suicide rates through empathy and personal stories, their understanding of data is important to bolster their arguments when working with stakeholders.

Data can be utilised in story-telling which is a powerful tool for advocates to convey a message to stakeholders and the public. Using narratives and visualisations, mental health advocates can take the "front-end" role in communicating insights from datasets involving suicide rates [8]. Data

story-telling is an effective way to inspire actions and identify gaps in mental health support and propose solutions for suicide preventions. Using real-life examples and interactive data visualisations can keep stakeholders interested and attentive throughout presentations.

### C. *Researchers and Academics*

Researchers and academics gain value from consuming this data in relation to initiatives on suicide rates and prevention. The data serves as a primary source for understanding patterns and trends which they can analyse and identify risk factors, correlations, and potential causes of suicide. Addressing the intricacies of the causes of suicides and developing a deeper understanding requires research and time, making it ideal for researchers and academics. They can aid public health officials by informing evidence-based practice and policy, developing, and testing interventions for prevention.

There are a myriad of datasets available to analyse in the context of suicide rates and raising awareness. Comparability of this data between countries is affected as the definition of death by self-harm is ascertained differently across regions [9]. However, researchers and academics should take the responsibility, consume the data, and investigate the inconsistencies and develop reports that contain no misclassification bias. Researchers and academics can leverage their skills and educational background to integrate previous studies in conjunction with their own investigations and provide the most reliable and valid research for stakeholders. This will not only provide value to not for-profit organisations but society overall.

## III. COMMON DATA VISUALISATIONS

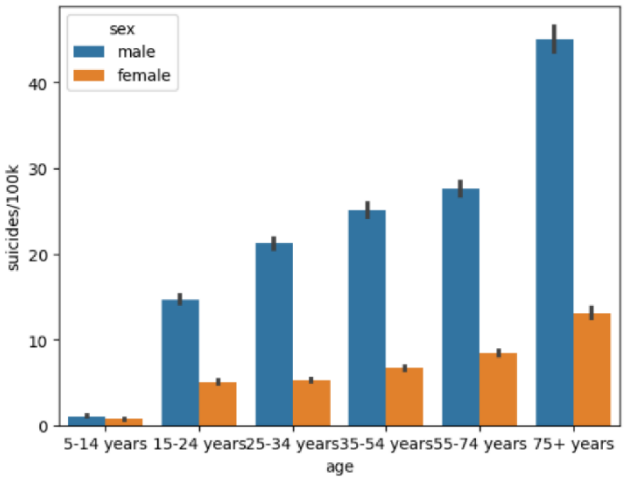


*Figure 1 – Suicide rates (per 100k) by age*

The "Suicide Rates Overview 1985 to 2021" dataset comprises of a wealth of information regarding suicide rates across different countries and regions. To effectively convey insights from this data, a variety of data visualisations can be employed, each tailored to highlight different characteristics of the data. Although common data visualisations largely rely on the data types, we can plot a few variables and provide our initial insights. Some questions that are asked and answered in figure 1 include questions relating to suicide rates and their demographic associations. It can help identify high-risk groups, for instance specific age groups or gender categories have elevated suicide rates, which can inform targeted interventions and prevention strategies. Comparing different age groups and genders can lead to a better understanding of the complex relationship between age, gender, and suicide. It can help researchers, mental health professionals and policymakers identify areas where additional research and targeted interventions are required.

In the context of suicide rate analysis, it is typical to visualise the dataset based on the number of suicides. Here in figure 1, the bar plot generated indicates the age on the x-axis which represents categorical data. The age variable can be further classified as ordinal data, because they have been ordered from youngest group to eldest in figure 1.

In addition, we cannot perform any mathematical operations on these age categories.

In the y-axis of figure 1 displays the suicide numbers for every 100,000 deaths. The variable "suicides/100k" represents quantitative data, characterised by numerical values that can be measured. The data contained in this variable is typically considered ratio data as it possesses all the properties of interval data whilst having a meaningful zero point. Mathematical operations are possible to be performed for the suicide numbers, for instance, we can calculate the ratio of suicides in one group compared to another or the percentage change in suicides over time.

The "sex" variable is a categorical nominal variable that contains male or female gender. There is no inherent order of this data type, hence there is no order, level of magnitude and is considered distinct. A bar plot is a suitable data visualisation based on our analysis of datatypes for the variables "suicides/100k", "age" and "sex". Comparing suicide rates across age groups and genders provides valuable insights into patterns in suicide rates. Interpreting the trend of the plot exemplifies that the suicide rates seem to be higher on males significantly compared to females and the suicide rates increase through age.

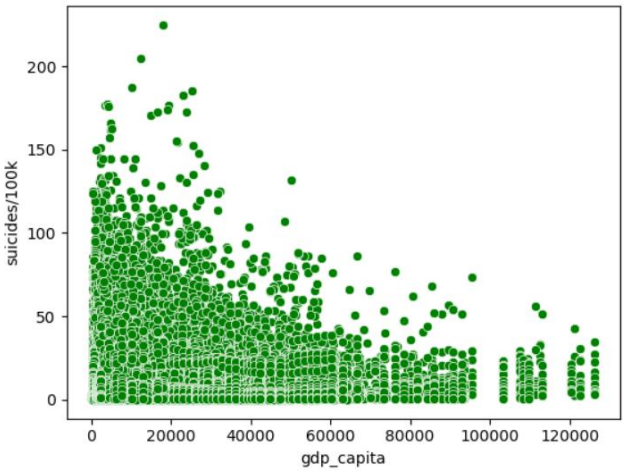


*Figure 2 – Suicide rates (per 100k) against GDP*

Inspecting the datatypes for "gdp_capita" allows us to visualise a scatter plot against "suicide/100k". This plot is

meaningful because it can help consumers of the data showcase the potential correlations between a country's economic prosperity and its suicide rates. The "gdp_capita" variable is considered a quantitative data that is also ratio data that can be measured and quantified. Higher values indicate higher levels of economic prosperity, for example a GDP per capita of $30,000 is higher than a GDP per capita of $12,000. The values in "gdp_capita" has a meaningful zero point, as the value of $0 would represent the absence of economic output per capita.

To justify the visualisation in figure 2, we begin with understanding that "gdp_capita" and "suicide/100k" are continuous quantitative variables which scatter plots are suitable for displaying this relationship. A scatter plot allows identifying patterns and trends in data, for instance this visualisation reveals that the higher the GDP per capita, the lower the suicide rates. Some questions that can be asked and answered include – is there a correlation between GDP per capita and suicide rates? How consistent is the relationship across countries and what are the implications for public policy and mental health programs?

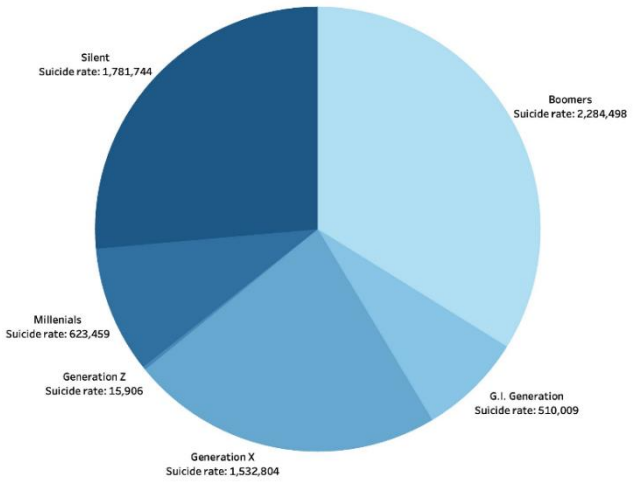


*Figure 3 – Suicide rates by different generations*

To further our analysis in visualising suicide rates, a pie chart has been constructed in figure 3. Questions that are important to ask and answer include – which generation experiences the highest suicide rates? Are certain generations more affected by suicide than others? To what extent do different generations contribute to the overall suicide problem? Although this visualisation may not be the most suitable one compared to bar chart or scatter plot for distribution purposes, the viewer can get a glance of the proportion of suicide rates across different generations. The generation variable is categorical which represents distinct groups that have no inherent order or numerical value. In the context of generations, we can justify this variable being ordinal as we can define a clear order such as Baby Boomers, Gen X, Gen Y and Gen Z in terms of their inherent age groups based on birth years. However, the intervals between generations are not uniform, hence cannot be considered interval or ratio data.

Aforementioned, suicide rates are quantitative ratio data and be used with the "generation" variable to create a pie chart. The caveats of this representation can be interpreted from the loss of information as pie charts display proportions better than quantitative relationships. Simple interpretations such as Boomers clearly having the greatest rate of suicides can be established, however due to the small portion slice of Generation Z, it's difficult for viewers to make the statement of Generation Z having the least suicide rates.

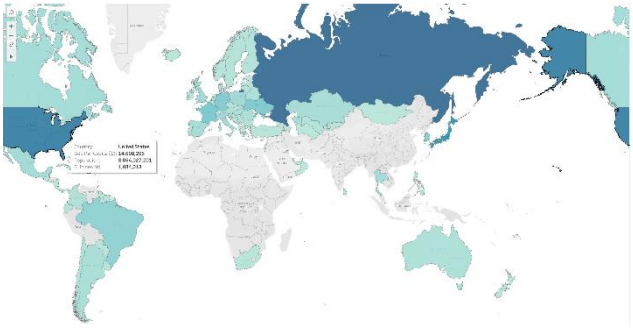

*Figure 4 – Map of suicide rates across countries*

It is imperative to visualise suicide rates across different countries via a map as shown in figure 4. Maps can be a powerful tool to answer a few questions such as, which countries have the highest suicide rates, and what factors might contribute to these differences? Do certain regions exhibit higher suicide rates than others, and are there common regional risk factors? The geospatial nature of the "country" variable allows us to categorise each country with their corresponding number of suicide rates. A blue colour palette has been utilised to convey the severity of suicide rates based on the different shades of blue, where darker shades represent higher suicide rates and lighter shades displays lower rates. In reference to datatypes, the "country" variable is categorical and more specifically nominal as there are no numerical labels and mathematical operations cannot be performed.

For stakeholders, the representation of this data can provide insights in which areas are the most vulnerable towards suicide rates. Overall, the visualisation provides simple interpretations but lacks region-specific suicide rates. Thus, mental health advocates and officials can focus on targeting campaigns and raise awareness towards these countries in alleviating suicide rates.

## IV. Mistakes in Depicting Data for Suicide Rates

Misinterpretations by the consumer of the data is possible due to reporting the number of suicides count across different countries and generations through time. Whether it is deliberate or not, identifying these mistakes can help alleviate misconceptions and allow us to identify any biases.

One typical mistake when depicting data for suicide rates is using scales or axes that are misleading for the viewer. To emphasises differences, some may use a scale that exaggerates the variation in suicide rates. This can lead to overemphasising differences between countries and make it difficult to accurately compare or interpret the data. It's vital to use a scale that accurately represents the data and ensures that differences are not disturbed. For example, the bar and

scatter plot in section III comprises of suicide rates on the y-axis per 100,000 records, without the adjustment of this scale the visualisation would appear different and constructing meaningful interpretation and analysis would be difficult.

An important aspect of visualisation involves their data quality and completeness. Missing or unreliable data can lead to gaps in the visual representation, which can mislead viewers. As mentioned in section I, the issue with representing suicide rates in a dataset are the underreporting rates of suicides from different countries as there is potential social stigma, potential biases can arise. Various countries that have been depicted in the map shown in section III are not present, however the main point is to highlight the counts as they may not be reliable. Underrepresentation of suicide rates, especially towards countries that have certain, more stricter requirements to classify what suicide is can affect the visualisations, causing viewers to inadvertently become misinformed and draw incorrect conclusions.

Another mistake is presenting the data without providing adequate context. Suicide rates can be influenced by a myriad of factors, including cultural, social, economic, and healthcare-related factors. Failing to provide context or explanations to variations in rates can result in viewers misinterpreting the data. Contextual information should be included to help viewers understand the complex factors contributing to suicide rates in different countries.

## V. Symbolic Representations of the Visualisations

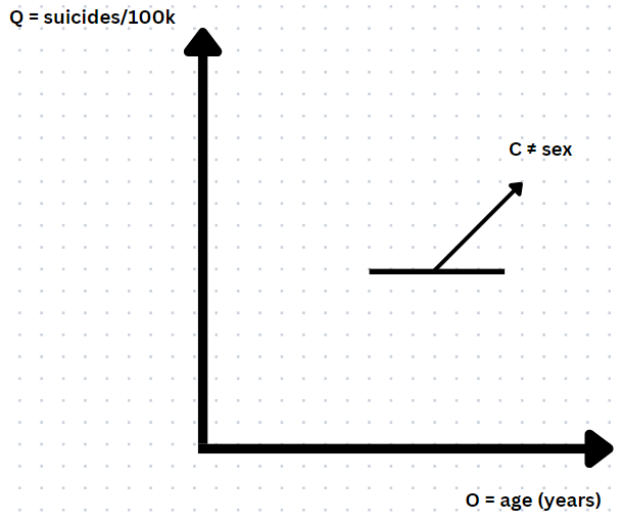


*Figure 5 – Symbolic representation of figure 1*

The plots constructed in section III have been represented using the symbolic representations from Semiology of Graphics for each plot. The report describes in depth of each component in detail. We begin our representation with the bar plot in figure 1, showing suicide counts against different age groups. As suicides per 100,000 counts is quantitative, we denote the graphic using "Q", whereas age being categorical and ordinal in this plot means we denote with "O". As there is elevation above the plane in this bar plot, we provide the corresponding symbol for the retinal variable, alongside the "C" to denote the categorical variable, sex. As sex is a qualitative variable which is differential for sex, we include the selective perception sign.

The symbolic construction of figure 2 is shown in figure 5. Considering datatypes again, the x-axis is quantitative and also classified as ratio data where mathematical operations are possible, we denote it with the "Q" symbol again. Similar to figure 4, we have "suicides/100k" in the y-axis being quantitative data. The difference

here are the implantations that are not differentiated. In this case we use a downward arc arrow with "P" symbol as this is a scatter plot. In addition, as there are components in this visualization containing elements that are considered similar, we denote it using the ≡ symbol.

When comparing figure 4 and 5, these two symbolic representations are equivalent however through different means. The imposition of the representations is largely the same, except figure 4 has retinal variable, whereas figure 3 has point implantation, being scatter plot instead of bar plot. Therefore, different datatypes in variables are crucial when we represent both plots and semiology of graphics.

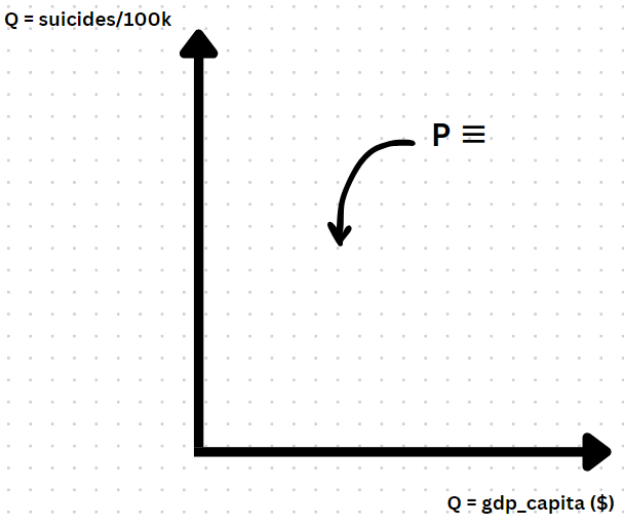


*Figure 6 – Symbolic representation of figure 2*

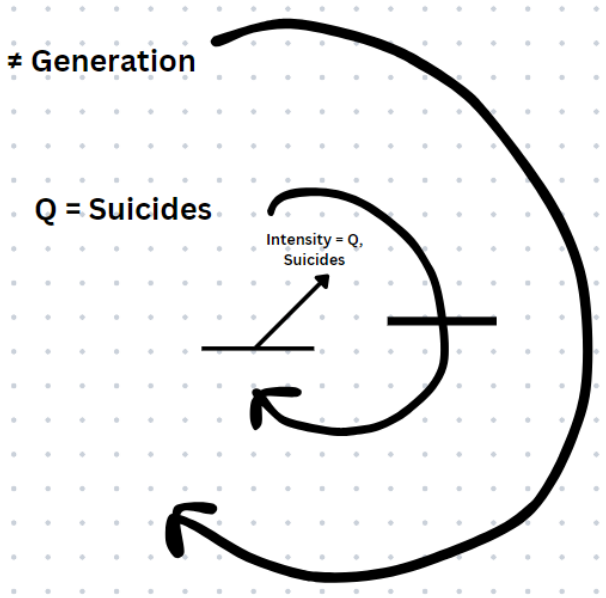


*Figure 7 – Symbolic representation of figure 3*

Now constructing the symbolic representation of the pie chart shown in figure 3 is displayed in figure 6. There are two major components that are noteworthy in the pie chart – suicide rate and the generations. This indicates that we require two circular utilizations, with the larger construction being the generations with the selective perception sign to represent qualitative component. The smaller circular arc represents the quantitative variable for number of suicides for each generation. As there are multiple generations stacked together there is a line intersecting the circular arc. Furthermore, the elevation symbol has been utilized with Intensity representing the suicides that is quantitative. The retinal variable for intensity has been chosen, because the viewer can identify which generation is the most affected by suicide rates based on the greater intensity of the blue palette used in figure 4. Although, the viewer is still able to understand the distribution of suicide rates across generations based on the different portion sizes of the pie chart and the labelled data. The final symbolic representation is shown in figure 8 which captures the semiology of graphics of the map in figure 4. As the map does not explicitly have x and y axes, we have incorporated a "GEO" axis on the bottom left corner to showcase the geospatial nature of the dataset of countries. Due to the axes being nonlinear and non-circular, we represent the arrangement with a continuous axis and a curvature in the corner as shown in figure 8. Additionally, the elevation symbol is used for colour which is equal to the suicide numbers. This is because the colour palette in the map from figure 4 represents the suicide rates across countries. The darker blue shades indicate greater suicide rates than the lighter shades. Since the intensity is differential in this visualization, we infer that there is selective perception, hence we include the corresponding sign near the elevation symbol. Another elevation symbol is utilized, however this time with retinal variables XY as it is a symbolic representation of a map which showcases the Country variable.

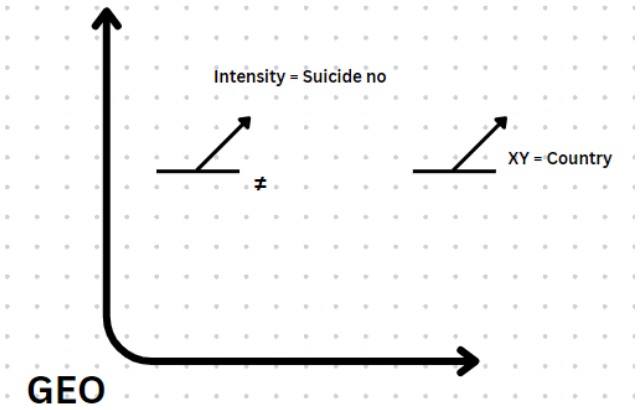


*Figure 8 – Symbolic representation of figure 4*

## VI. Visualisation of the Alternative Symbolic Representation

Extending our analysis of symbolic representations, here is a sample of the visualisation based on figure 5. Below on figure 9, a line chart has been constructed as the alternative visualisation:

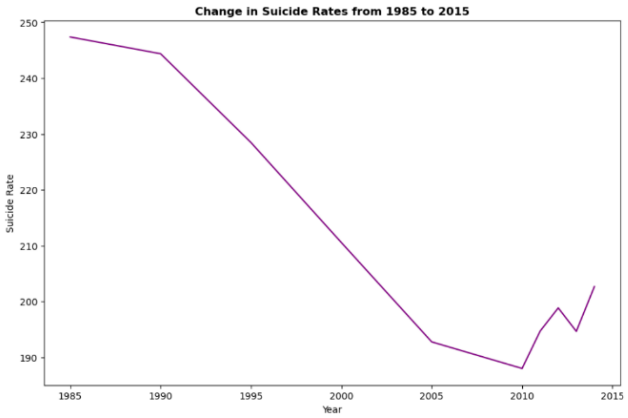


*Figure 9 - Equivalent representation of figure 5*

The y-axis being suicide rates is consistent with the plot obtained in figure 1, both containing quantitative and ratio data. The alternative representation in figure 5 highlights a categorical variable on the x-axis, so we can utilise year as the independent variable to estimate the trend of suicide rates from 1985 to 2015. This plot was constructed using Python and any null values have been removed. This has caused the plot to have a maximum of the year 2015 as most of the null values resided after that year. A line chart being a continuous plot with no points, any null values present have been eradicated.

We also observe that unlike figure 1 which incorporates sex into the bar plot, this representation omits that. To incorporate the sex variable into this plot, there would be 2 lines required and an added legend. When interpreting this plot, we can analyse the overall trend, more specifically the suicide rates seem to have declined significantly from 1985 till 2010. Since then, suicide rates have gradually increased. It is imperative for researchers and academics to scrutinise this pattern, conduct investigations and gain context from these years to understand the causes of the decrease and increase in suicide rates from 1985-2010 and 2010-2015. Moreover, the line chart in figure 9 provides a smooth transition between data points which is beneficial when dealing with continuous data like years. It allows a more intuitive understanding of the data, especially for identifying patterns or cyclical trends.

On the other hand, a disadvantage of this alternative representation is that compared to bar charts, line charts do not provide precise comparisons when identifying exact values for specific years. Despite accounting for null values when constructing this plot, line charts may not make outliers as apparent as bar charts with extreme values not being identified due to the continuous nature of the plot.

## VII. Equivalency of Symbolic Representation of Visualisations

Constructing alternative, but equivalent representations and determining their equivalency is a crucial aspect of working with Semiology of Graphics and data visualisations. Viewers may have varied preferences, background knowledge or cognitive abilities when observing the data. By providing alternative representations and determining their equivalency will allow us to reach a broader audience and ensure that people can understand and interpret the data differently. For example, figure 10 is an equivalent representation for figure 5 and has many similar elements to it. The key differences is observed from the dotted line which symbolises the "sex" variable as it is repeated several times in the bar plot displayed in figure 1. As the age variable is qualitative and differential component we add the ≠ sign. The bottom left corner contains the number of times male, and female grouped together have been repeated. Although this representation is equivalent to figure 5, there are a few alternative components added for the viewer to grasp a more comprehensive understanding of the visualisation itself.

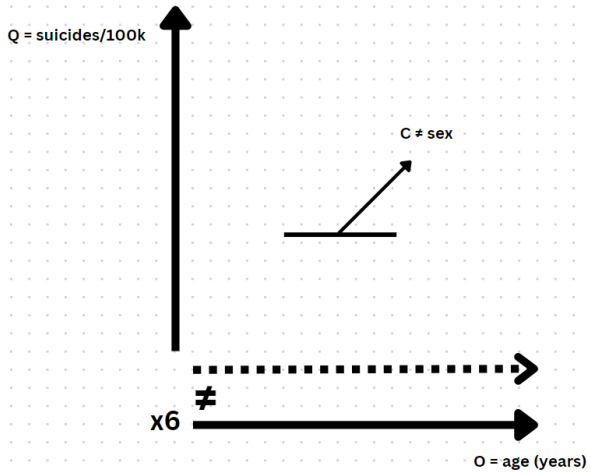


*Figure 10 Equivalent representation of figure 5*

Another alternative symbolic representation can be constructed as shown in figure 11.

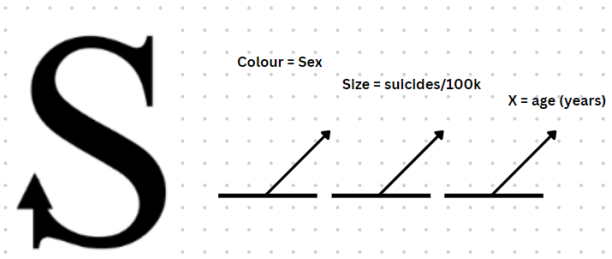


*Figure 11 - Equivalent representation of figure 5*

Here, the S symbol is used as the axes are not linear nor circular with an arrow pointing vertically as the dimension of suicide rates is observed to be in vertical order. Elevation components are also observed, with sex represented using two different colours, size of the visualisation is determined from suicide rates and the x-axis having elevation in the age groups in ascending order. In reference to specific datatypes, the construction of alternative symbolic representations of visualisations mentioned in previous sections acknowledges viewers to have diverse preferences

and needs in making the data more valuable and usable to a broader audience.

## REFERENCES


[1] Anderson, M., & Jenkins, R. (2005). The Challenge of Suicide Prevention. Disease Management & Health Outcomes, 13(4), 245-253. https://doi.org/10.2165/00115677-200513040-00003

[2] Gowda, O. (n.d.). Suicide Rates Overview (1985 to 2021). Kaggle. Available at: https://www.kaggle.com/datasets/omkargowda/suicide-rates-overview-1985-to-2021/data (Accessed: 21 October 2023).

[3] Dattani, S., Rodés-Guirao, L., Ritchie, H., Roser, M., & Ortiz-Ospina, E. (2023). Suicides. Published online at OurWorldInData.org. Retrieved from https://ourworldindata.org/suicide

[4] Olson, R. (2021). The accuracy and reliability of suicide statistics: Why it matters. Centre for Suicide Prevention. Available at: https://www.suicideinfo.ca/local_resource/accuracy-suicide-statistics/#:~:text=How%20accurate%20are%20suicide%20statistics,found%20more%20than%2030%25%20underreporting. (Accessed: 20 October 2023).

[5] Mohler, B., & Earls, F. (2001). Trends in adolescent suicide: misclassification bias?. American journal of public health, 91(1), 150–153. https://doi.org/10.2105/ajph.91.1.150

[6] Large M. M. (2018). The role of prediction in suicide prevention. Dialogues in clinical neuroscience, 20(3), 197–205. https://doi.org/10.31887/DCNS.2018.20.3/mlarge

[7] Saha G. (2021). Advocacy in mental health. Indian journal of psychiatry, 63(6), 523–526. https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_901_21

[8] Cote, C. (2021, November 23). DATA STORYTELLING: HOW TO EFFECTIVELY TELL A STORY WITH DATA. Harvard Business School Online. https://online.hbs.edu/blog/post/data-storytelling

[9] OECD. (2023). Suicide rates (indicator). https://doi.org/10.1787/a82f3459-en