Money Matters: How Socioeconomic Status Impacts Mental Health

Money Matters

Information Visualization Project (COMP3045) – 2022/2023 University of Nottingham Malaysia

Omar I. Taha

(Student ID 20311657) School of Computer Science, University of Nottingham Malaysia, hcyot1@nottingham.edu.my

CCS CONCEPTS • Health Informatics • Visualization • Interaction Design

**Additional Keywords and Phrases:** Mental Health, World Bank Income levels, Institute for Health Metrics and Evaluation, Global Burden of Disease

1. Introduction

The relationship between income level and mental health has long been a topic of interest to researchers and policymakers alike. While studies have shown that money can't buy happiness[1], the question remains: can it buy mental health? This assignment aims to explore the complex relationship between income level, unemployment, and the prevalence of specific mental disorders such as depression and anxiety globally over time.

By utilizing data from the Global Burden of Disease study, this assignment will use various data visualization techniques such as scatter plots, line plots, choropleth maps, box plots, and stacked bar charts to uncover any trends and patterns that may exist in the data. Due to the limits of the dataset, data will be shown from 1990 – 2019.

Most importantly, the visualizations presented in this study will have interactive elements to filter results and view animated changes over time, radically increasing the comprehensibility and usefulness of the data visualization.

The findings of this study can have important implications for policymakers and public health officials in developing effective strategies to address the mental health needs of different income brackets across the globe.

* 1. Methodology

The initial exploration visualizations will be done with R, and the final visualizations will be done with the use of multiple libraries like Plotly (interactive plots) and Shiny (creates dashboards) to make an interactive dashboard to aid in filtering data and enhance comprehensibility of the findings.

1. Data Description

The main dataset used in this report is the Institute for Health Metrics and Evaluation’s Global Burden of Disease dataset[[1]](#footnote-1)[3] (to be referred to as IHME GBD from now on).

* 1. IHME GBD

The IHME GBD dataset provides a massive range of epidemiological data, but in this report, we will only focus on mental health data. Data on the prevalence of mental health disorders globally is collected by combining medical, epidemiological data, surveys and meta-regression modelling where raw data is unavailable.

It is important to note that the data should be interpreted with caution as the uncertainty of data on mental health is generally high. Mental health disorders remain under-reported across all countries, particularly at lower incomes where data is scarcer and there is less attention and treatment for mental health disorders. To fill gaps in knowledge, data is collected from diagnoses made by doctors and the severity of depression reported by patients and the general population using various questionnaires and rating scales[2]. This collation and projection help give more a realistic dataset on world mental health rates.

The table below showcases the general layout of the datasets with examples used in this report after dropping unnecessary columns.

Table 1: General layout of the datasets used with examples

| location | year | val |
| --- | --- | --- |
| Azerbaijan | 1990 | 0.11 |
| Azerbaijan | 1991 | 0.12 |
| Azerbaijan | 1992 | 0.13 |
| United States of America | 1990 | 0.03 |
| United States of America | 1991 | 0.02 |
| United States of America | 1992 | 0.07 |
| … | … | … |

The “val” column in the table above measures the specific metric we are looking at. For example, prevalence of mental disorders as a % of population. This varies for each question. The location and year columns are self-explanatory.

1. Initial Questions and Visualizations

For all of the following questions, an online tool provided by IHME called GHDx will be used to filter the massive amounts of data into only portions of what is required for the proposed questions.

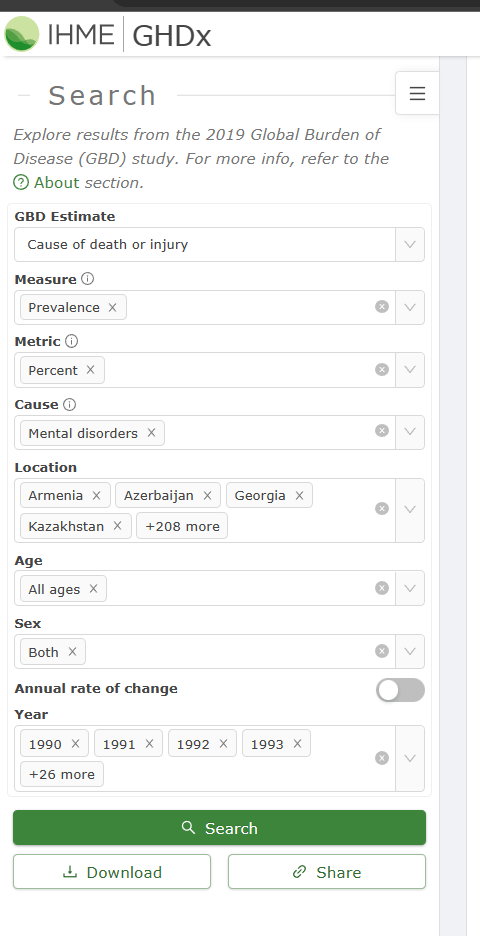


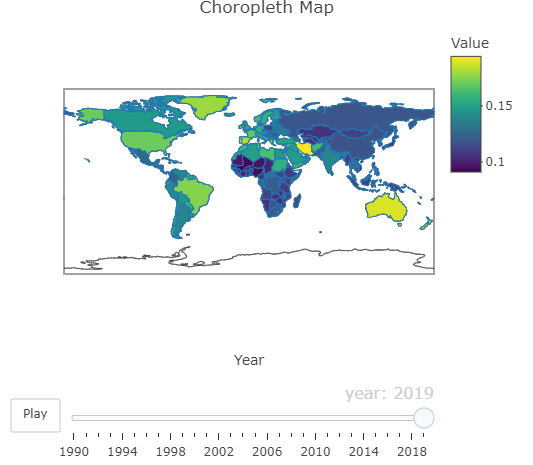
Figure 1: Image of using the online GHDx tool to get the required data

* 1. What is the prevalence of mental disorders globally over time?

The information was obtained from the IHME GBD by looking at the proportion of people with mental disorders in a country's total population over the years 1990–2019. This data includes both genders. The countries used were all countries and territories of the world. The chosen visualization for this question is an animated choropleth map, so we can easily see the changes across the globe over time.

The dataset contains a lot of columns that are not useful to us, so we import the “raw” CSV file that was downloaded from the GBD website into a dataframe, and them chose only the “location\_name”, “year”, and “val” columns.

The dataset contained the country names in the “location\_name” column, for example: “United States of America”. This has to be changed to ISO3C country codes for it to be plotted in a choropleth map with plotly (for example, change “United Arab Emirates” to “ARE”). The countrycodes package was used for this. The Plotly package allows you to easily create animations. This is the resulting visualization:

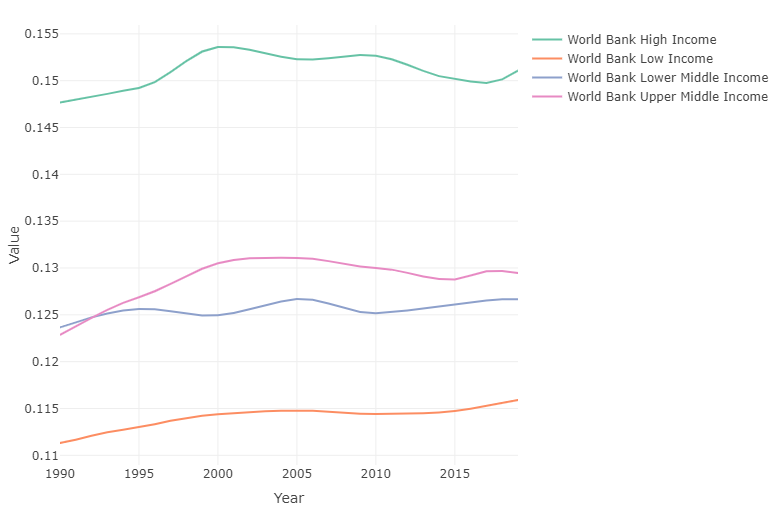


The resulting plot has a play button that makes the countries on the map change color over time, corresponding to the rates of mental disorder prevalence.

* 1. What is the relationship between income level and prevalence of mental disorders globally over time?

The information to answer this question was obtained from the IHME GBD by looking at the proportion of people with mental disorders in a country's total population. This data includes both genders. The income levels were split according to the World Bank’s classifications into Low, Lower Middle, Upper Middle, and High incomes. The chosen visualization for this question is a line map, where the x-axis is the years, and the y-axis is the prevalence of mental disorders.

From the dataset, unnecessary columns were removed, except the “location\_name”, “year”, and “val” columns, which contain the income levels, years, and rates of mental illness respectively.

****

* 1. How does the prevalence of specific mental disorders, such as depression or anxiety, differ by income level?

Data will be gotten from the IHME GBD dataset, with the following financial classifications:

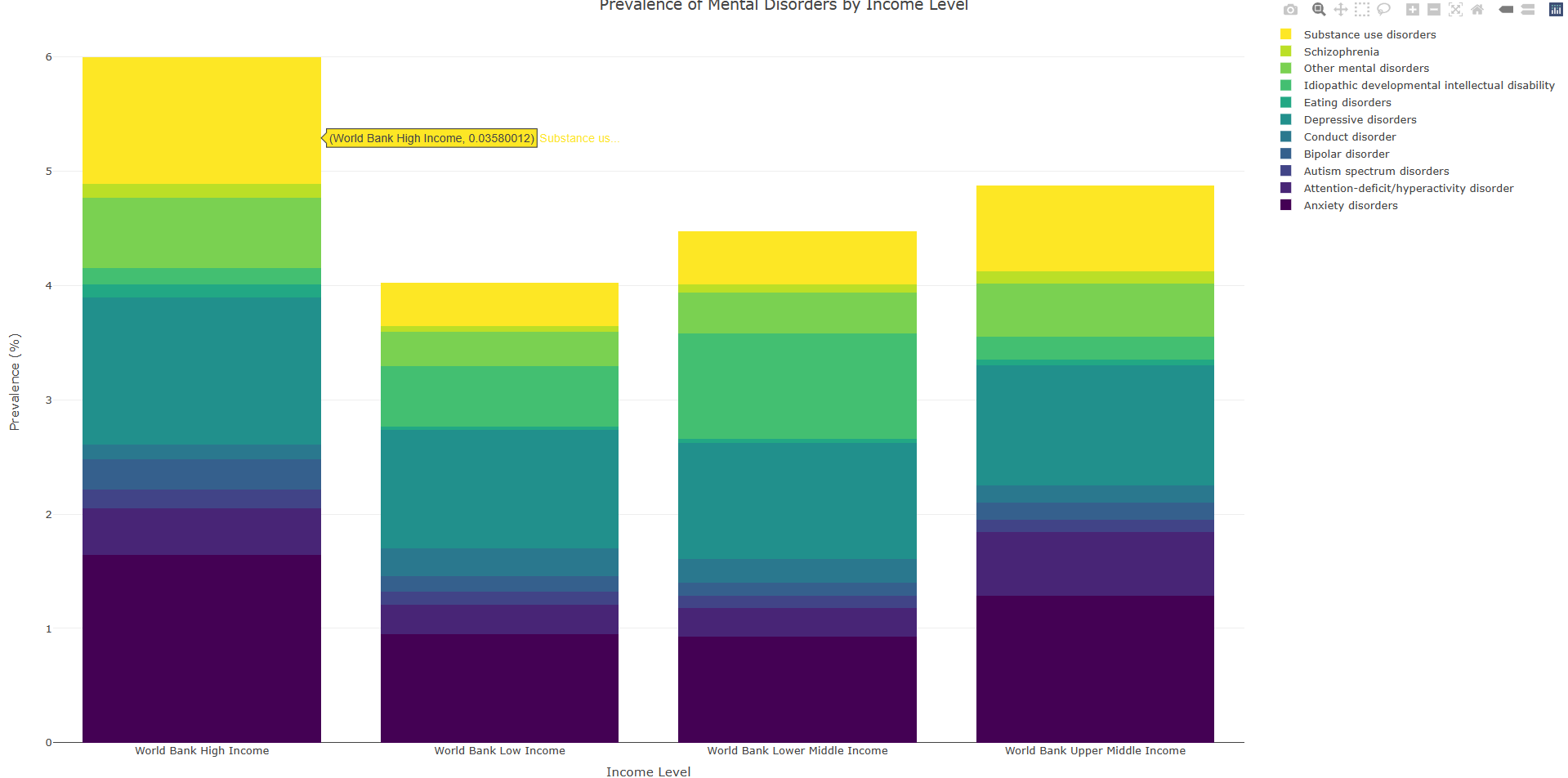
* World Bank Low Income
* World Bank Lower Middle Income
* World Bank Upper Middle Income
* World Bank High Income

All of the mental disorders measured in the GBD dataset will be considered, they consist of the following:

* Schizophrenia
* Depressive disorders
* Bipolar disorder
* Anxiety disorders
* Eating disorders
* Autism spectrum disorders
* Attention-deficit/hyperactivity disorder
* Conduct disorder
* Idiopathic developmental intellectual disability
* “Other” mental disorders
* Substance use disorders

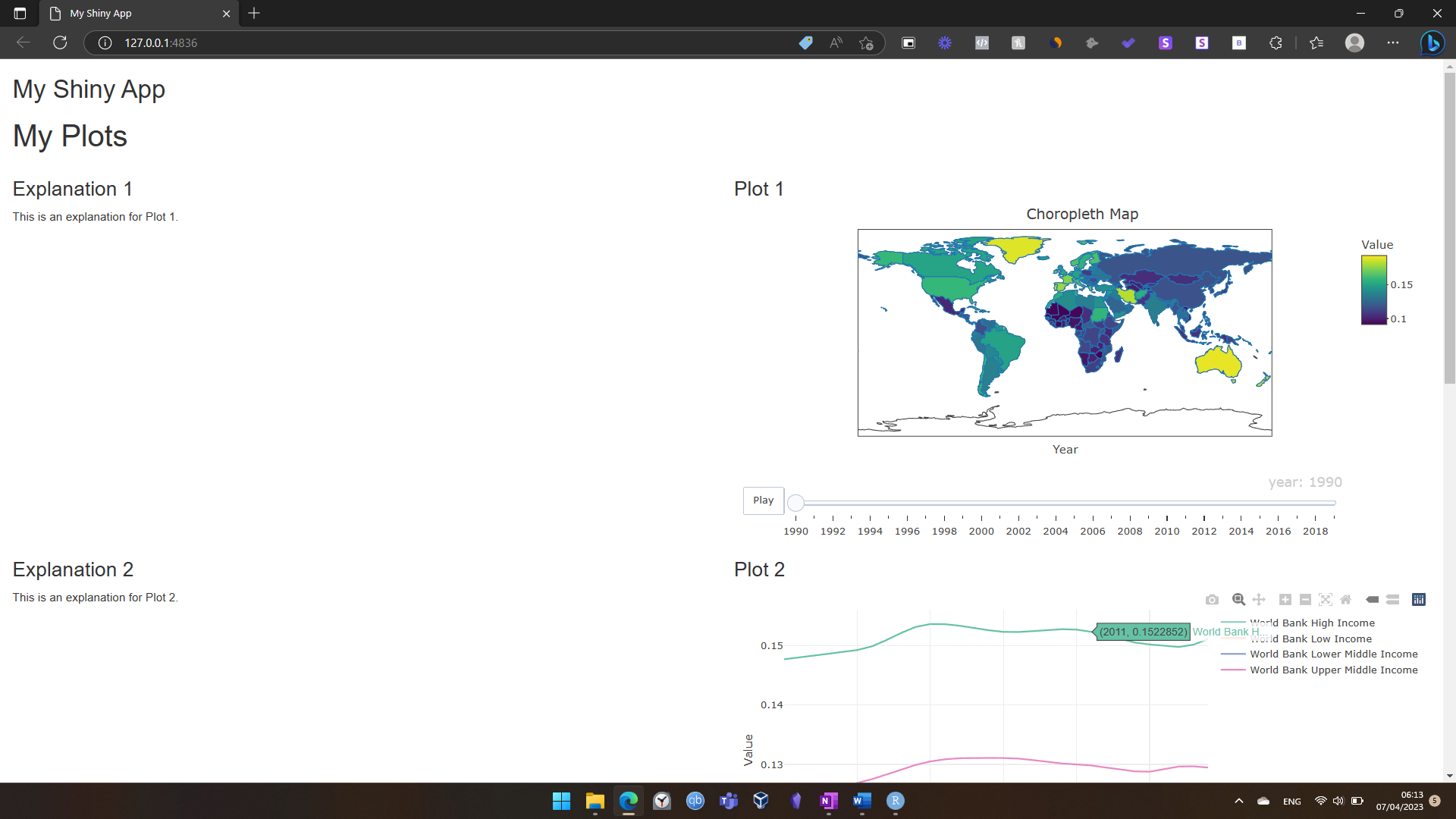
The data we get from the GHDx website provides the income levels under the “location” column. I rename this column to “income” in the code to make it easier.

The first stacked bar chart is successful, here is what it looks like:



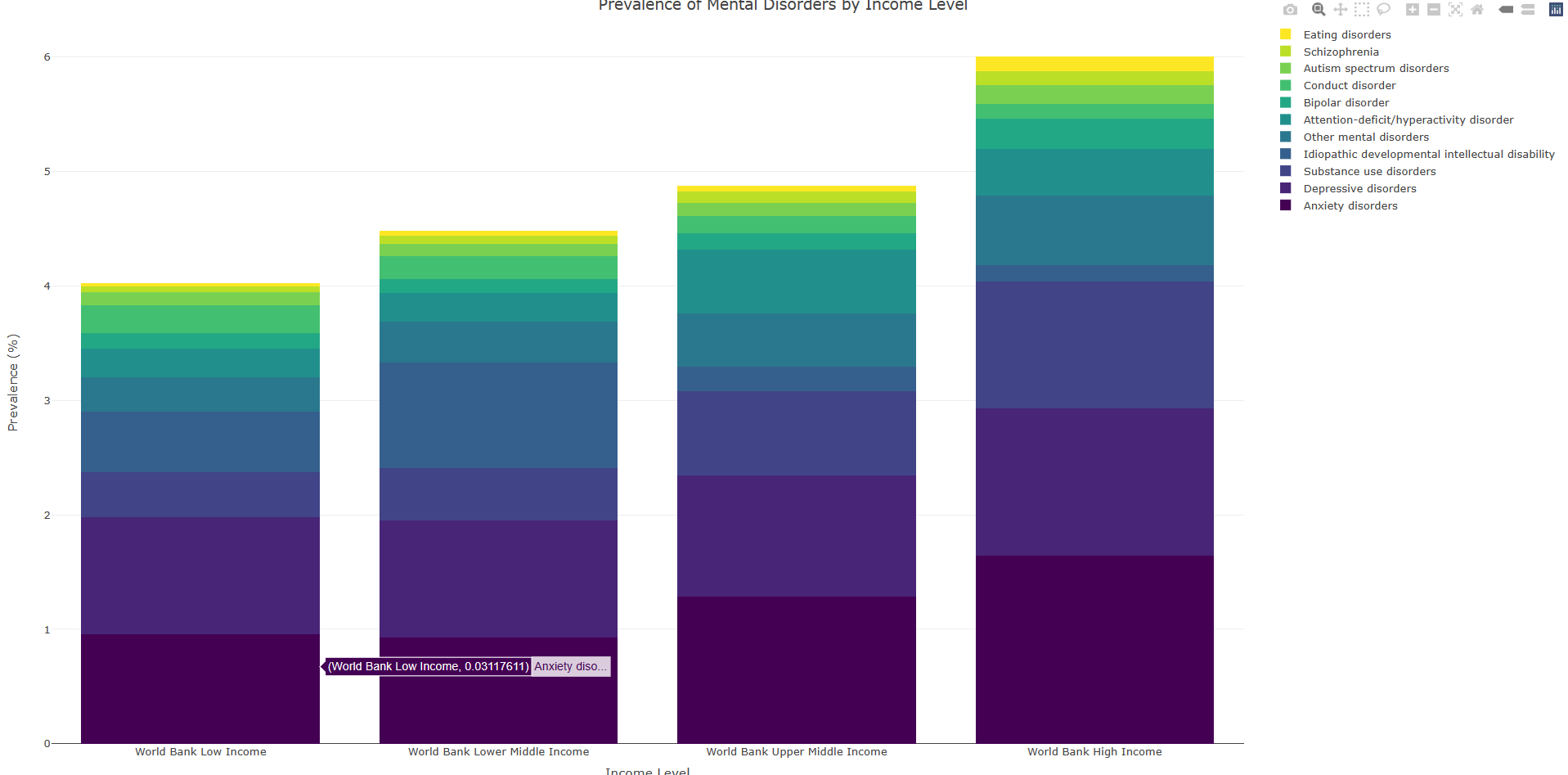
* 1. Putting it all together in a Shiny App

The entire Shiny app is put together, with each plot having a short explanation. It will be styled later, here is the resulting Shiny app:



1. Refinement and Proposed Questions
   1. Question 2

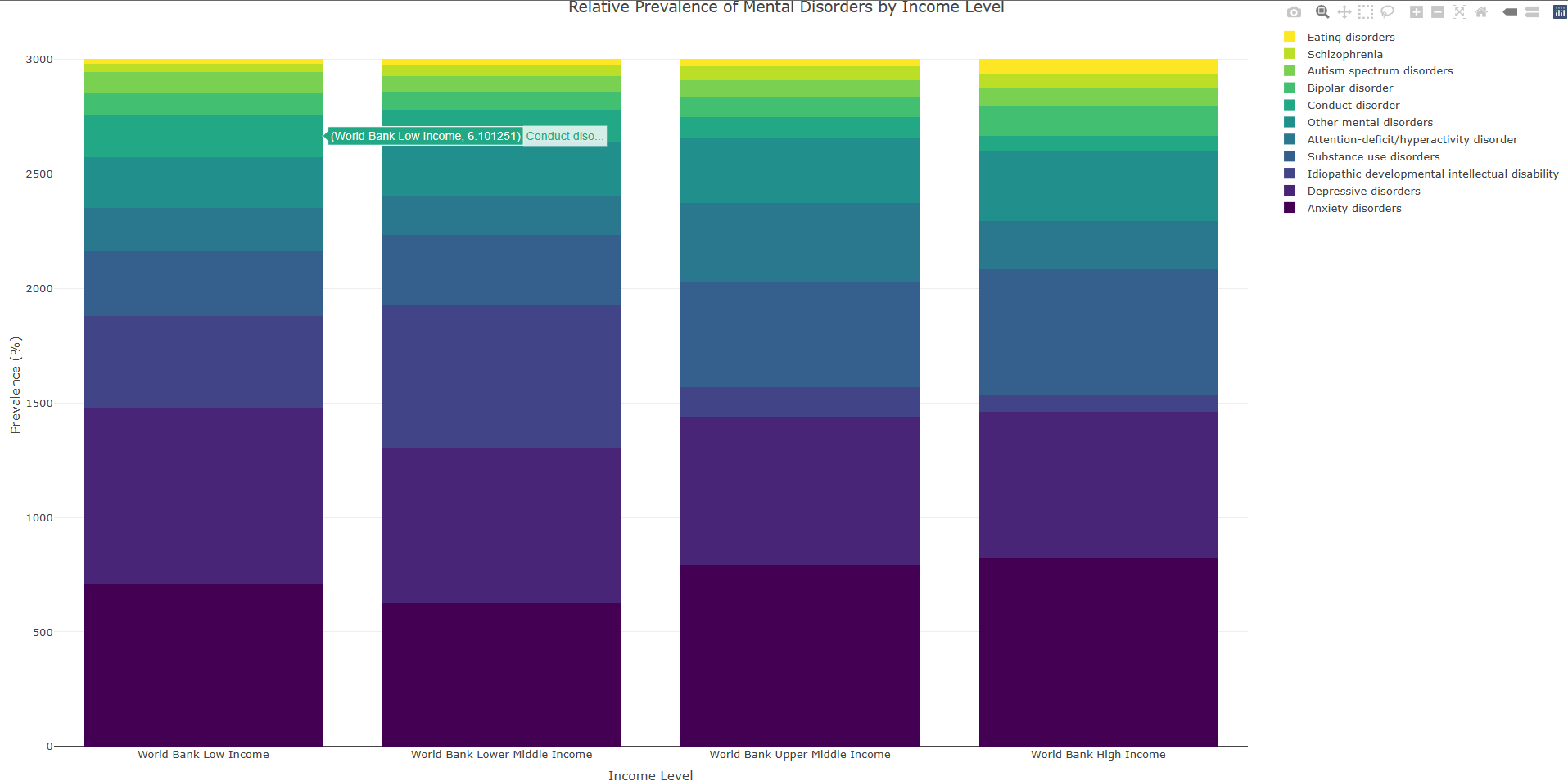
The layout of the stacks should be changed such that the most prevalent mental disorders are at the bottom, and the least are at the top, which results in bigger chunks at the bottom, and smaller chunks on the top. This is more visually appealing. I also rearrange the bars so that they are in order of income level, from Low to High income. This is the resulting visualization:



* 1. Creation of Q4:

The point of the Q3 visualization was to view relative differences in mental disorder prevalence. But each bar is a different height because each income group has different levels of mental disorder prevalence. The visualization in Q2 answers this question (it shows the different rates of prevalence among income groups).

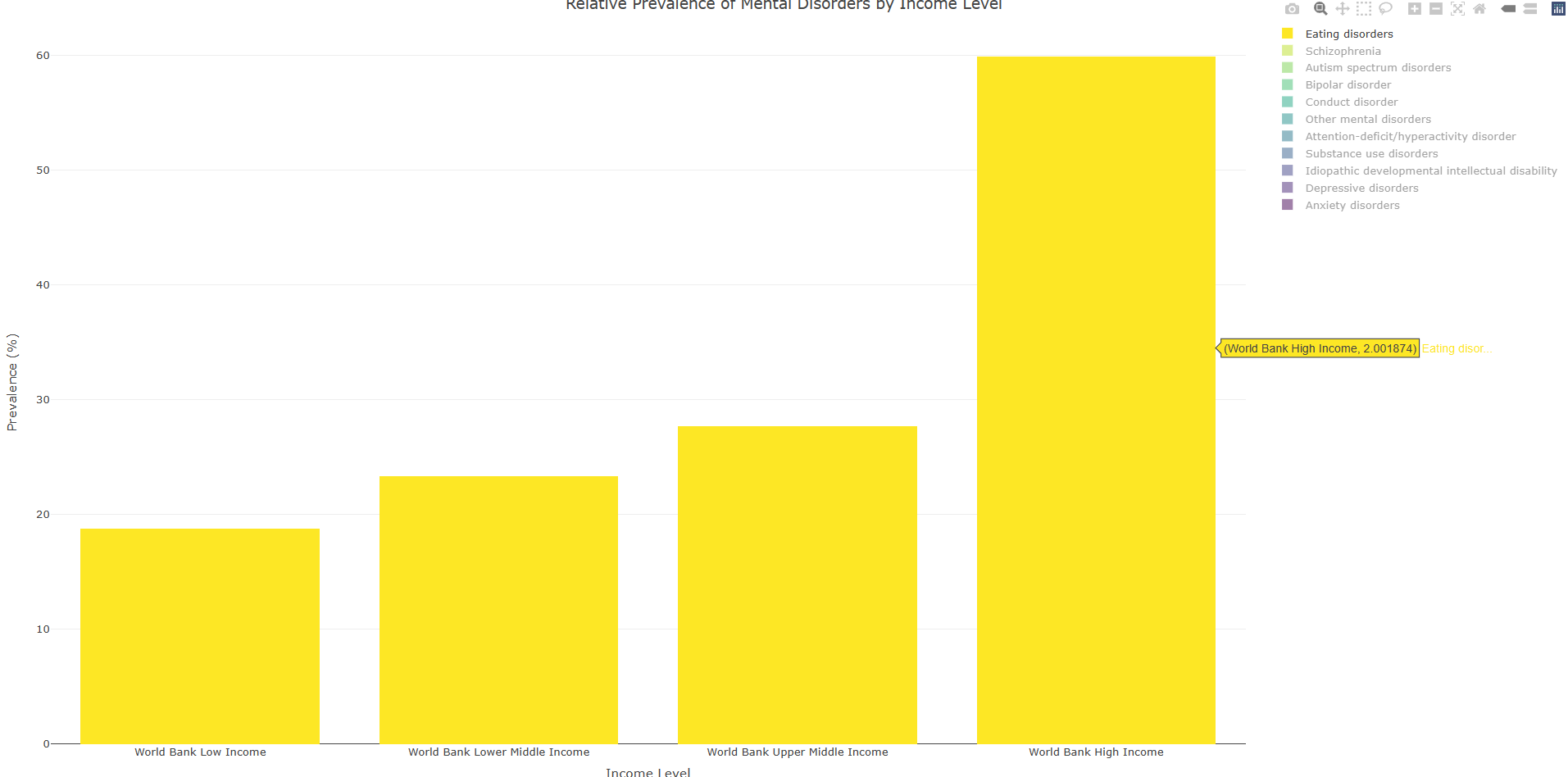
This then means we should compare relative rates of mental disorders by setting them all to a relative/normalized scale. **T**o achieve this, we can normalize the prevalence of mental disorders within each income group to a common denominator. This will now allow us to compare the proportions of mental disorders within each income group, as percentages will be on the same scale (0-100%) for all income groups, regardless of the actual mental disorder prevalence values.  
  
The bars in the clustered stacked bar chart will now represent the relative proportions of mental disorders within each income group, allowing for easier comparison across income groups:



1. Discussion
   1. Visualization Design

The stacked bar plots may suffer from the overplotting problem, but it is crucial that we can see all tracked mental disorders so we have the full picture. This where plotly swoops in to save the day, and this is exactly why interactivity is so important.

Plotly allows us to filter out mental disorders simply by clicking on them. Let’s say I want to compare only the difference in eating disorders between the groups. All I must do is deselect all disorders except “eating disorder”, and this is what we get:

****

This shows a stark difference in eating disorders between income groups

* 1. Development Process

Several stages were involved in the development of the data visualization project, including data acquisition, data cleaning and preprocessing, exploratory data analysis, visualisation design, and implementation using the R programming language and associated packages.

The dataset for the IMHE GBD 2019 was obtained from the official IMHE GBD website and was cleaned and preprocessed to address missing data (some small countries had NA values). To gain insights and identify patterns in the data, iterative visualisation design techniques were used, which determined appropriate visualisations, layout, colors, and interactivity. The visualisations were then implemented using R packages such as ggplot2, plotly, and Shiny.

Design choices were overall more difficult than the implementation itself.

1. Conclusion

Finally, this study examined the relationship between income level, unemployment, and mental disorders globally using data from the Institute for Health Metrics and Evaluation's Global Burden of Disease dataset. The findings show that people of higher income are more likely to face mental health problems, and the findings emphasise the importance of targeted interventions and policies to address mental health disparities across income groups. The interactive visualisations presented in this study improve the findings' comprehension and usability, making them useful for informing policy and practise. More research in this area is needed to better understand the multifaceted relationship between income level and mental health and to inform targeted interventions to improve mental health outcomes for all people, regardless of their income level.

REFERENCES

1. Daniel Kahneman and Angus Deaton. 2010. High income improves evaluation of life but not emotional well-being. Proceedings of the National Academy of Sciences 107, 38 (September 2010), 16489-16493. DOI:https://doi.org/10.1073/pnas.1011492107
2. Mental Health - Our World in Data. *Our World in Data*. Retrieved April 9, 2023 from https://ourworldindata.org/mental-health#how-is-prevalence-defined-and-measured
3. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020. Available from https://vizhub.healthdata.org/gbd-results/

1. Available from https://vizhub.healthdata.org/gbd-results/ [↑](#footnote-ref-1)