## The University of Nottingham Logo PNG Transparent & SVG Vector - Freebie SupplyInformation Visualiztion Project (COMP3045)

## Project Proposal Report

## “Money Matters: How Socioeconomic Status Impacts Mental Health”

## Omar Ismail (Student ID 20311657)

## 28th March 2023

## Data Sources

* [VizHub - GBD Results (healthdata.org)](https://vizhub.healthdata.org/gbd-results/)

The IMHE Global Burden of Disease 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. In particular, 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. *(insert citation)*

This collation and projection help give more a realistic dataset on world mental health rates.

* [World Bank Open Data | Data](https://data.worldbank.org/)

World Bank Open Data is a comprehensive range of datasets that provide free and open access to the World Bank's collection of development data. The datasets include over 20,000 indicators for over 200 countries and regions, covering topics such as poverty, health, education, and the economy.

With data from member countries, the World Bank ensures the quality and integrity of their vast collection of macro, financial, and sector data. *(insert citation)*

## Problem Description

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 (*insert citation*), 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, the World Bank, and the WID/UNU-WIDER database, 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 main dataset, the Global Burden of Disease, data will be shown from 1990 – 2019, with different limits if other datasets have smaller time ranges.

Most importantly, the visualizations presented in this study will have interactive elements to filter results and view 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.

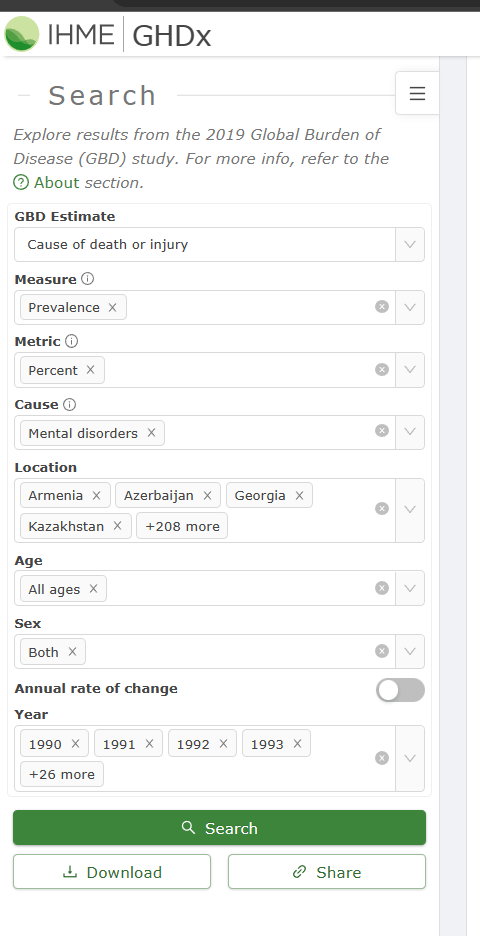
## Methodology

The initial exploration visualizations will be done with R, and the final visualizations will be done with the use of multiple libraries to make an interactive dashboard to aid in filtering data and enhance comprehensibility of the findings.

These are the initial questions to be answered, along with the chosen visualization techniques:

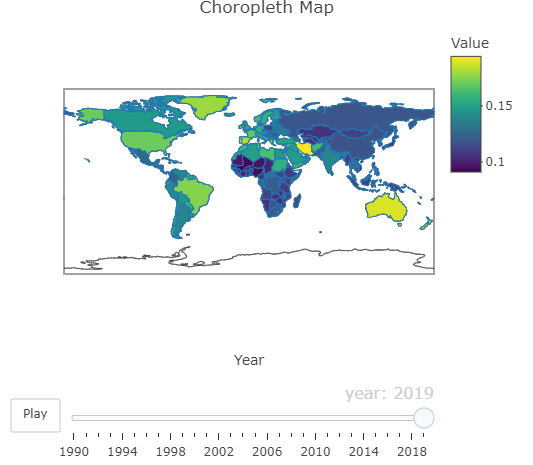
1. **What is the prevalence of mental disorders globally over time? (chloropleth map + line plot)**

The information was obtained from the IMHE GBD by looking at the proportion of people with mental disorders in a country's total population. This data includes both genders. The countries used were all countries and territories of the world, with additional entries for the 4 world regions (Africa, Europe, Asia, and the Americas).



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 the “location\_name”, “year”, and “val” columns.

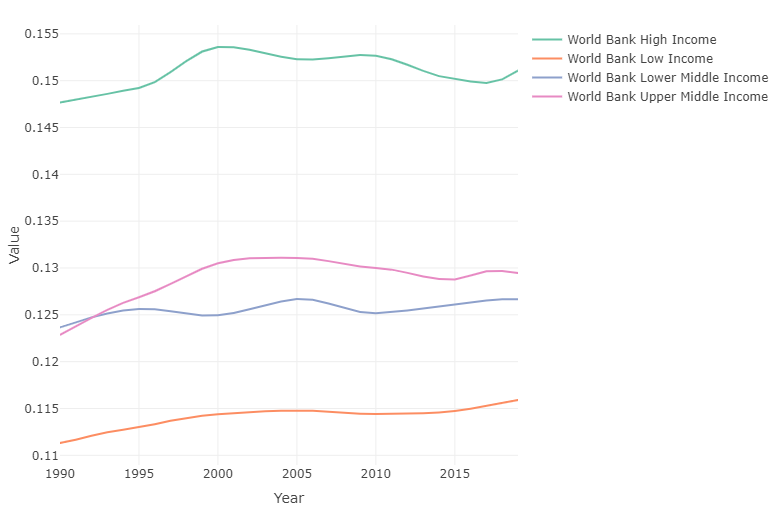
The dataset contained the country names in the “location\_name” column, for example: “United States of America”. This had to be changed to ISO3C country codes for it to be plotted in a choropleth map. The countrycodes package was used for this. Plotly package easily allows animations. This is the resulting visualization:



1. **What is the relationship between income level and prevalence of mental disorders globally over time? (line plot) (data in GBD)**

The information to answer this question was obtained from the IMHE 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.

Extra 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? (stacked bar chart) (GBD)**

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

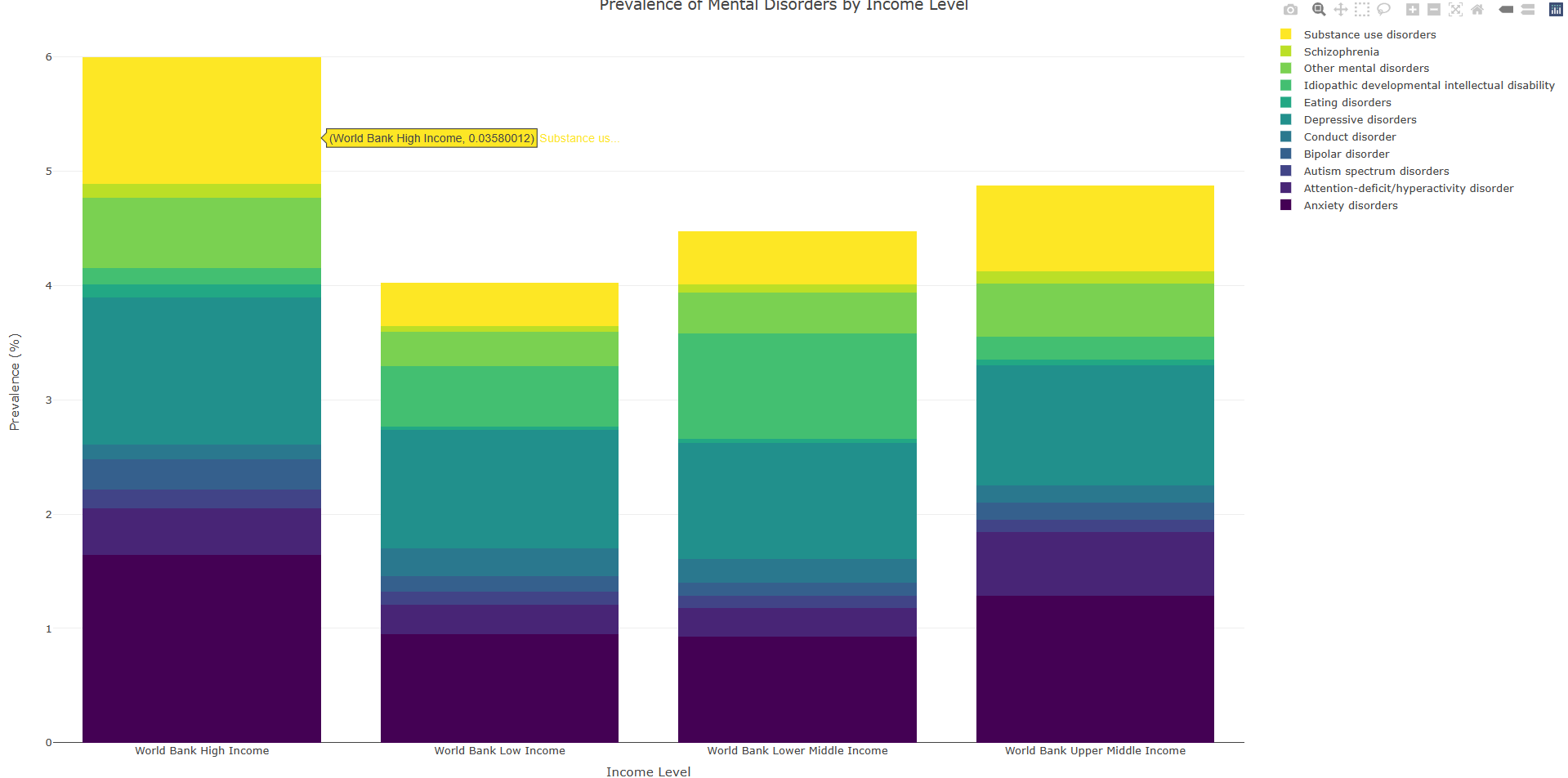
* World Bank Low Income
* World Bank Lower Middle Income
* World Bank Upper Middle Income
* World Bank High Income

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

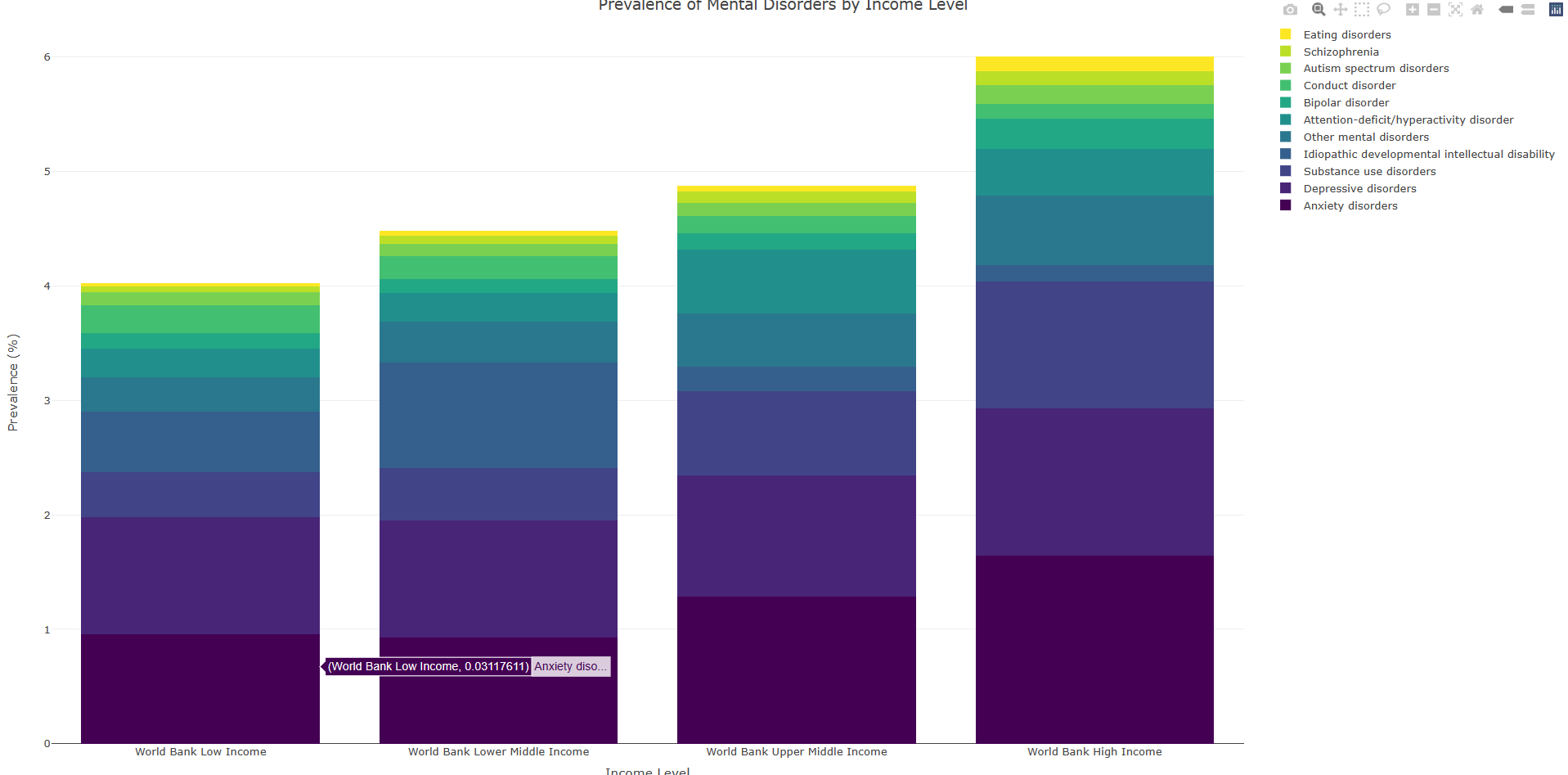
* Schizophrenia
* Depressive disorders
* Bipolar disorder
* Anxiety disorders
* Eating disorders
* Autism spectrum disorders ()
* Attention-deficit/hyperactivity disorder
* Conduct disorder
* Idiopathic developmental intellectual disability
* “Other mental disorders”
* Substance use disorders

The data we get from the GBD 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:

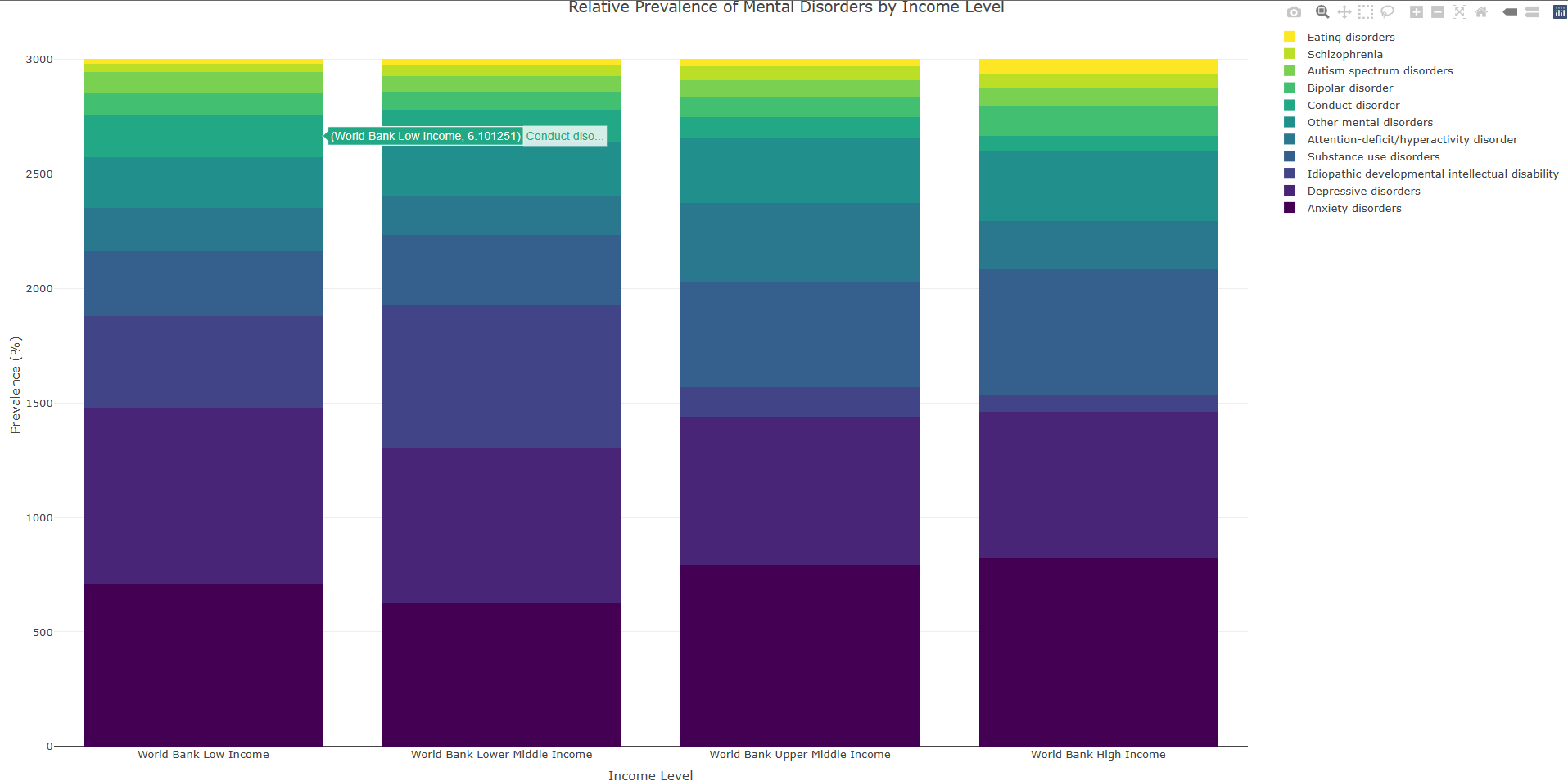


I want to change the layout of the stacks such that the most prevalent mental disorders are at the bottom, and the least are at the top, so it creates a nice visual effect of order. I also rearrange the bars so that they are in order of income level, from Low to High income.

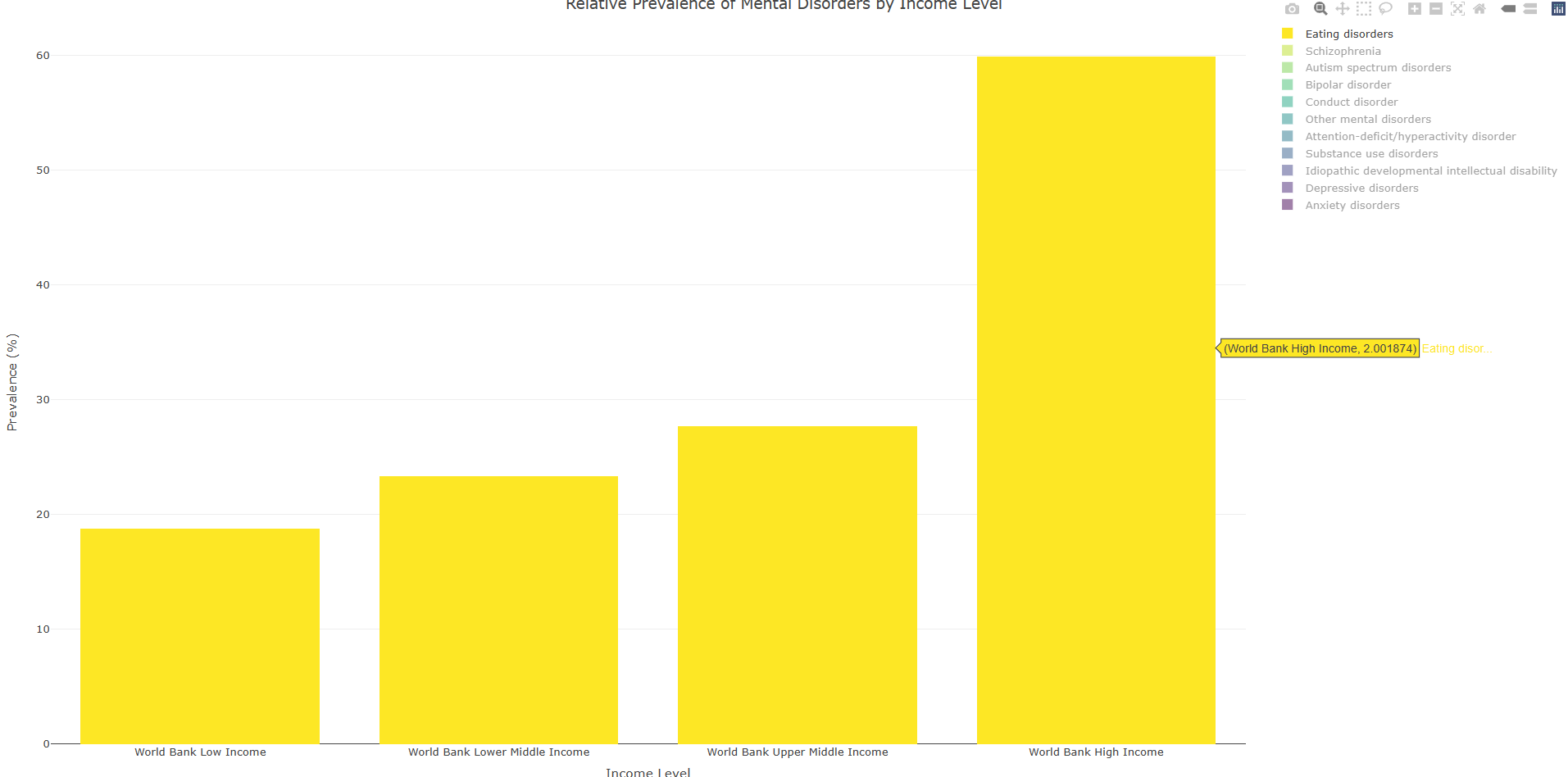


The point of this visualization was to view relative differences in mental disorder prevalence. But each bar is a different height because each income group has different levels of mental disorder prevalence. The visualization in Q2 answers this question (it shows the different rates of prevalence among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

This then means we should compare relative rates of mental disorders by setting them all from a relative/normalized scale. To 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:



This visualization 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 have to 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.

So now we have a Q4 as well.

Putting it all together in a Shiny App

## 

## Discussion/Reflection

## References/Resources

* [Tabs | Dash for Python Documentation | Plotly](https://dash.plotly.com/dash-core-components/tabs)
* [Mental Health - Our World in Data](https://ourworldindata.org/mental-health)
* **Please cite this database version as** UNU-WIDER, World Income Inequality Database (WIID). Version 30 June 2022. <https://doi.org/10.35188/UNU-WIDER/WIID-300622>
* [Global Burden of Disease Study 2019 (GBD 2019) Population Estimates 1950-2019 | GHDx (healthdata.org)](https://ghdx.healthdata.org/record/ihme-data/gbd-2019-population-estimates-1950-2019)
* [Population, total | Data (worldbank.org)](https://data.worldbank.org/indicator/SP.POP.TOTL)

GBD Results tool:

Use the following to cite data included in this download:

Global Burden of Disease Collaborative Network.

Global Burden of Disease Study 2019 (GBD 2019) Results.

Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020.

Available from https://vizhub.healthdata.org/gbd-results/.