Economic Inequality and the Global Mental health Burden

Alona Sychevska, 2769408

June 5, 2025

```r
library(readr)
library(dplyr)
library(ggplot2)
library(tidyr)
library(WDI)
library(rnaturalearth)
library(rnaturalearthdata)
library(readxl)
library(naniar)
library(reshape2)
library(sf)
library(viridis)
```

Economic Inequality and the Global Mental health Burden

Alona Sychevska

Tutorial lecturer's name: Chantal Schouwenaar, Jack Fitzgerald

1 Problem Motivation & Literature

Why It's a Recognized Social Problem

Mental health issues such as depression and anxiety are on the rise globally—not only in high-income countries like the Netherlands or the UK, but also across low- and middle-income countries (LMICs). According to the WHO, mental disorders are among the leading causes of disability worldwide, and their burden is projected to grow (World Health Organization, 2023).

A growing body of literature suggests that **economic inequality**, rather than just poverty or absolute income, is a key social determinant of mental health. Studies have found that countries with higher levels of income inequality tend to have higher rates of psychological distress, even after controlling for GDP per capita and other macroeconomic indicators (Wilkinson & Pickett, 2009).

A systematic review of 26 studies—primarily from high-income countries—found that greater income inequality is associated with a 19% increased risk of depression (Patel et al., 2018). While most of this literature focuses on wealthy countries, studies in LMICs have also confirmed that poverty and inequality significantly predict common mental disorders (Lund et al., 2010).

More recent cross-continental studies have demonstrated that income inequality not only exacerbates anxiety and depression in high-income countries, but also in emerging and developing economies, reinforcing the need for a more globally inclusive analysis (Patel et al., 2022).

Theoretical Framing & Mechanisms

Several mechanisms help explain how economic inequality may drive mental health problems:

- **Relative deprivation**: People evaluate their well-being not in absolute terms but in comparison to others. This social comparison can lead to chronic stress and lower self-worth, especially in unequal societies (Smith et al., 2012).
- **Social capital erosion**: Inequality weakens social cohesion and trust, contributing to social exclusion and loneliness, which are known risk factors for mental illness (Kawachi & Berkman, 2000).
- **Bidirectional poverty–mental health link**: Mental illness can both result from and reinforce poverty, creating a feedback loop that is particularly severe in LMICs (Lund et al., 2010).

Gap in the Literature

Most cross-national studies on mental health and inequality focus on wealthy countries or single-region datasets. Far fewer include **LMICs**, despite the fact that these countries now bear a disproportionately large share of the global mental health burden.

Moreover, few studies consider a **broad set of economic variables together**, such as for inequality: Gini index, wealth concentration, gdp per capita.

This project combines all of these, providing a richer understanding of **how multiple inequality-related factors relate to mental health outcomes** across different economic contexts. It uses cross-national data covering income groups from the World Bank and mental health data from the Global Burden of Disease project.

2 Data Sourcing & Description

Data Sources and Credibility

This project integrates datasets: the Global Burden of Disease (GBD) Study 2021 and a multi-source datasets compiled from the World Bank (WDI), WHO, and other internationally reputable institutions that includes socioeconomic variables.

The GBD Study is produced by the Institute for Health Metrics and Evaluation (IHME) at the University of Washington. It is considered the gold standard for global health statistics and is widely cited by academic institutions and policymakers worldwide (IHME, 2022).

The merged compilation of datasets drawing from sources like the World Bank's World Development Indicators (WDI), and World Population Review. These are recognized institutions known for robust methodologies and transparency. WDI, in particular, is frequently used in cross-country empirical studies due to its consistency and broad temporal coverage.

Metadata Overview

Go to ReadMe.txt

Complementarity of the Datasets

GBD offers high-quality health burden metrics (Prevalence), crucial for understanding the impact of mental health disorders, while the multi-sourced dataset offers explanatory variables on inequality and economic performance. These datasets are complementary in structure and purpose: the GBD gives outcome measures, while the merged dataset gives potential predictors. Their integration allows for robust modeling of social determinants of mental health at the national level.

While I initially collected more than two datasets, the final dataset has complementary strengths and ability to support the causality-oriented focus of this project: exploring how inequality and structural factors relate to mental health burdens across countries.

Relevance to the Topic

The project's focus is on the causal relationship between inequality and mental health in low-, middle-, and high-income countries. The selected datasets are ideal because: The GBD dataset provides mental health burden metrics (e.g., Prevalence of depression or anxiety), which are essential for measuring the societal impact of mental disorders across countries. The merged dataset allows for correlational and regression analyses of how inequality indicators (e.g., Gini index, top 10% wealth share) , wealth, economic performance relate to mental health prevalence. Datasets cover a wide range of countries, including those outside the high-income bracket, which aligns with the study's comparative angle.

Limitations of the Data

Despite their strengths, datasets have limitations: GBD data, though standardized, aggregate country-level estimates, potentially masking subnational disparities and cultural variation in diagnosis/reporting. The multi-sourced data suffers from missing data for certain indicators in low-income countries, which could bias regression results or reduce sample size. Differences in data collection years (some indicators are 2021, others 2022 or 2023) may introduce temporal misalignment. Mental health prevalence estimates from sources like World Population Review may lack the methodological rigor of epidemiological surveys.

2.1 Load in the data

```r
GBD <- read_csv("../data/Global-Burden-of-Disease-Study.csv")
gni_by_country_year <- read_excel("../data/gni_by_country_year.xlsx")
economic_inequality_gini_index <- read_csv("../data/economic-inequality-gini-index/economic-inequality-g
percent_wealth_10 <- read_csv("../data/10percent_wealth/Income share held by highest 10%.csv")
population <- read_csv("../data/population.csv")
world <- ne_countries(scale = "medium", returnclass = "sf")

# Pull GDP per capita data
  gdp_data <- WDI(
  country = "all",
  indicator = "NY.GDP.PCAP.CD",  # GDP per capita (current US$)
  start = 2012,
  end = 2025
)
```

2.2 Checking the structure

```r
# checking structure of dataset gini
head(economic_inequality_gini_index, 3)
```

```
## # A tibble: 3 x 5
##   Entity  Code   Year Gini coefficient (2017 prices) - I~1 `1039568-annotations`
##   <chr>   <chr> <dbl>                                <dbl> <lgl>
## 1 Albania ALB    1996                                0.270 NA
## 2 Albania ALB    2002                                0.317 NA
## 3 Albania ALB    2005                                0.306 NA
## # i abbreviated name:
## #   1: `Gini coefficient (2017 prices) - Income or consumption consolidated`
```

```r
head(gdp_data, 3)
```

```
##                        country iso2c iso3c year NY.GDP.PCAP.CD
## 1 Africa Eastern and Southern    ZH   AFE 2024             NA
## 2 Africa Eastern and Southern    ZH   AFE 2023       1659.515
## 3 Africa Eastern and Southern    ZH   AFE 2022       1628.025
```

```r
head(GBD %>% select(measure_name, location_name, cause_name, metric_name, year, val), 3)
```

```
## # A tibble: 3 x 6
##   measure_name location_name cause_name       metric_name  year       val
##   <chr>        <chr>         <chr>            <chr>       <dbl>     <dbl>
## 1 Deaths       Israel        Mental disorders Number       2010 0.616
## 2 Deaths       Israel        Mental disorders Percent      2010 0.0000156
## 3 Deaths       Israel        Mental disorders Rate         2010 0.00786
```

```r
head(gni_by_country_year %>% select(Country, "2014", "2015", "2020"), 2)
```

```
## # A tibble: 2 x 4
##   Country     `2014` `2015` `2020`
##   <chr>       <chr>  <chr>  <chr>
## 1 Afghanistan L      L      L
## 2 Albania     UM     UM     UM
```

2.3 Quantifying

Data cleaning

```r
economic_inequality_gini_index <- economic_inequality_gini_index %>%
  filter(Year >= 2012) %>%
  rename("Gini_index" = "Gini coefficient (2017 prices) - Income or consumption consolidated") %>%
  select(-Code, -`1039568-annotations`)
# Creating gini buckets (new var)
gini_cats <- economic_inequality_gini_index %>%
  mutate(gini_bucket = case_when(
    Gini_index < 0.25              ~ "Low",
    Gini_index >= 0.25 & Gini_index < 0.35 ~ "Moderate",
    Gini_index >= 0.35             ~ "High",
    TRUE                         ~ NA_character_
  ))
```

```r
# Select only the columns needed, no duplicates
gdp_data <- gdp_data %>%
  select(country, year, NY.GDP.PCAP.CD) %>%
  rename(gdp_per_capita = NY.GDP.PCAP.CD) %>%
  filter(year >= 2012)
```

```r
gni_by_country_year <- gni_by_country_year %>%
  rename("2016" = "...31", "2017" = "...32", "2018" = "...33", "2019" = "...34") %>%
  pivot_longer(
    cols = matches("^\\d{4}$"),
    names_to = "year",
    values_to = "income_group"
  ) %>%
  mutate(year = as.integer(year)) %>%
  select(Country, year, income_group)
gni_by_country_year <- gni_by_country_year %>%
  mutate(income_group = ifelse(income_group == "..", NA, income_group)) %>%
  filter(year >= 2012)
```

```r
population <- population %>%
  rename(name = Entity, pop_est = "Population (historical)", pop_year = Year) %>%
  select(name, pop_est, pop_year) %>%
  filter(pop_year >= 2012)
```

```r
#transform percent wealth held by Top 10% merge
wealth10_long <- percent_wealth_10 %>%
  pivot_longer(
    cols = matches("^\\d{4}$"),  # Select only year columns
    names_to = "year",
    values_to = "wealth_share_10"
  ) %>%
  mutate(
    year = as.integer(year)
  ) %>%
  filter(!is.na(wealth_share_10)) %>%
  group_by(`Country Name`) %>%
  rename(
    country = `Country Name`
  ) %>%
  select(country, wealth_share_10, year) %>%
  filter(year >= 2012)
```

```r
#renaming col names
GBD <- GBD %>%
  rename(Country = location_name, cause = cause_name, Year = year)
```

Generate necessary variables

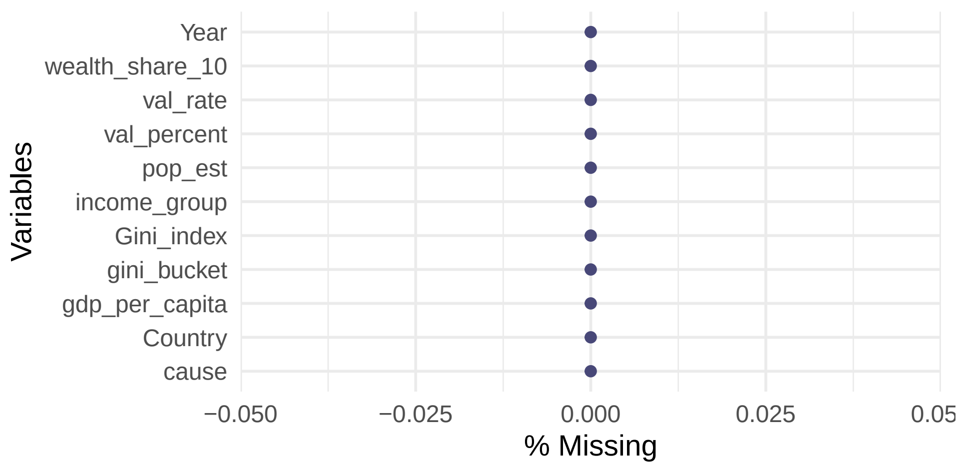
```r
GBD <- GBD %>%
  inner_join(population,
```

```
            by = c("Country" = "name", "Year" = "pop_year"))
# Calculate rate per 100,000 and percent
GBD <- GBD %>%
  filter(metric_name == "Number", measure_name == "Prevalence") %>%
  mutate(
    val_m = round((val / pop_est) * 100000, 2),
    val_rate = format(val_m, big.mark = ",", scientific = FALSE),
    val_percent = (val / pop_est) * 100
  ) %>%
  select(Country, cause, Year, val_rate, val_percent, pop_est) %>%
  #pivot_wider(names_from = cause,
  #             values_from = val_rate) %>%
  filter(Year >= 2012)
```

2.4 Datasets merging

```
# Mental + gini_cats data
merged_data <- GBD %>%
  inner_join(gini_cats, by = c("Country" = "Entity", "Year" = "Year"))
merged_data <- merged_data %>%
  inner_join(gni_by_country_year,by = c("Country" = "Country", "Year" = "year"))
merged_data <- merged_data %>%
  inner_join(gdp_data, by = c("Country" = "country", "Year" = "year"))
# Merge wealth share data
merged_data <- merged_data %>%
  inner_join(wealth10_long, by = c("Country" = "country", "Year" = "year"))
```

```
gg_miss_var(merged_data, show_pct = TRUE)
```



2.5 Describe the type of variables included

```
describe_variables <- function(df) {
  desc <- lapply(names(df), function(var) {
```

```r
    col <- df[[var]]
    var_class <- class(col)

    example_vals <- if (is.numeric(col)) {
      sprintf("mean = %.2f, sd = %.2f", mean(col, na.rm = TRUE), sd(col, na.rm = TRUE))
    } else if (is.factor(col) || is.character(col)) {
      vals <- unique(na.omit(col))
      paste("levels:", paste(head(vals, 5), collapse = ", "), if (length(vals) > 5) "...", collapse = "
    } else if (is.logical(col)) {
      "logical (TRUE/FALSE)"
    } else {
      paste("class:", var_class)
    }

    type_label <- if (is.numeric(col)) {
      "continuous"
    } else if (is.factor(col) || is.character(col)) {
      "categorical"
    } else if (is.logical(col)) {
      "logical"
    } else {
      "other"
    }

    paste0("- ", var, ": ", type_label, " (", var_class, "), ", example_vals)
  })
  cat(paste(unlist(desc), collapse = "\n"))
}
```

```r
describe_variables(merged_data)
```

```
## - Country: categorical (character), levels: China, Mongolia, Belarus, Kazakhstan, Armenia ...
## - cause: categorical (character), levels: Schizophrenia, Attention-deficit/hyperactivity disorder, I
## - Year: continuous (numeric), mean = 2016.37, sd = 2.82
## - val_rate: categorical (character), levels:    366.55,    366.01,    365.29,    364.93,    365.07 .
## - val_percent: continuous (numeric), mean = 2.78, sd = 4.25
## - pop_est: continuous (numeric), mean = 104628265.19, sd = 315184396.73
## - Gini_index: continuous (numeric), mean = 0.34, sd = 0.07
## - gini_bucket: categorical (character), levels: High, Moderate, Low
## - income_group: categorical (character), levels: UM, LM, H
## - gdp_per_capita: continuous (numeric), mean = 30698.76, sd = 24331.04
## - wealth_share_10: continuous (numeric), mean = 26.60, sd = 5.16
```

```r
dep_anx <- merged_data[
  (merged_data$cause %in% c("Anxiety disorders", "Depressive disorders")),
  ]

mental_disorders <- merged_data[
  (merged_data$cause %in% c("Mental disorders")),
  ]
```

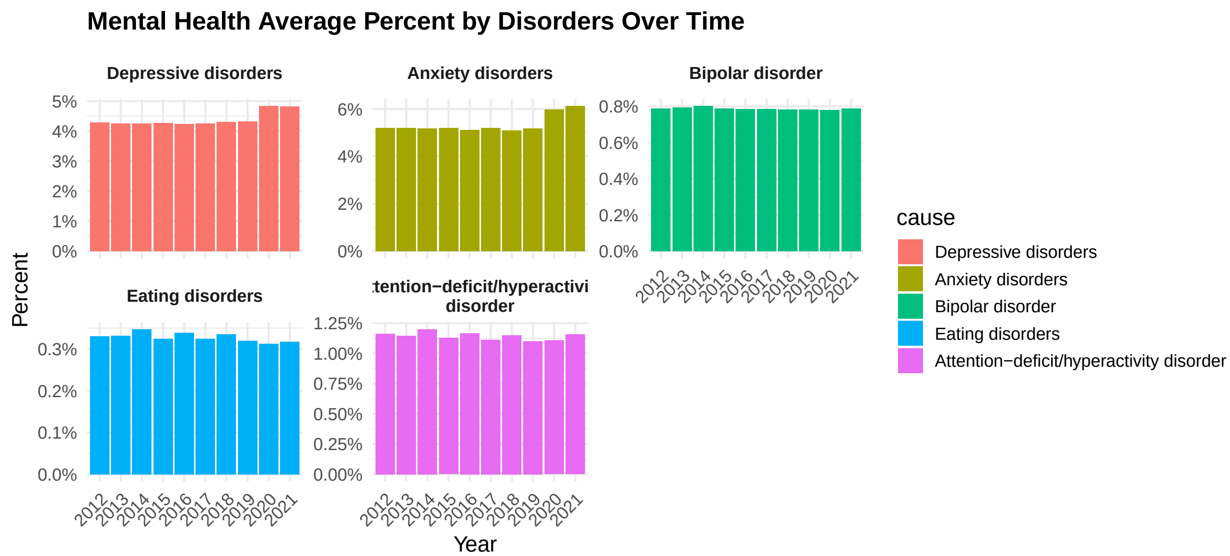
3.1 Visualizations

```r
tendencies <- merged_data %>%
  filter(cause %in% c("Anxiety disorders", "Bipolar disorder",
                      "Attention-deficit/hyperactivity disorder",
                      "Eating disorders", "Depressive disorders")) %>%
  mutate(cause = factor(cause, levels = c("Depressive disorders",
                                          "Anxiety disorders",
                                          "Bipolar disorder",
                                          "Eating disorders",
                                          "Attention-deficit/hyperactivity disorder"))) %>%
  group_by(cause, Year) %>%
  summarise(avg_val_percent = mean(val_percent, na.rm = TRUE)) %>%
  select(cause, Year, avg_val_percent)
```

3.1.1 Temporal Variations

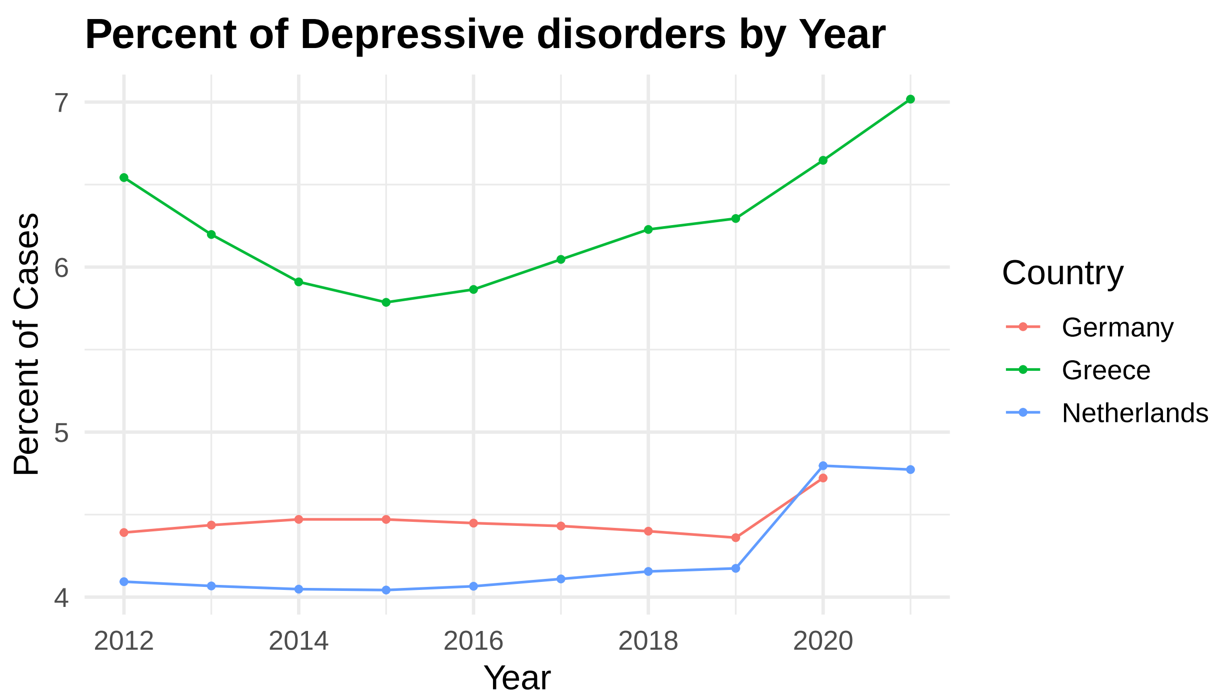
```
## `summarise()` has grouped output by 'cause'. You can override using the
## `.groups` argument.
```

```r
ggplot(tendencies,
       aes(x = factor(Year), y = avg_val_percent, fill = cause)) +
  geom_col() +
  facet_wrap(~ cause, scales = "free_y", labeller = label_wrap_gen(25)) +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  labs(
    title = "Mental Health Average Percent by Disorders Over Time",
    x = "Year",
    y = "Percent"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold", size = 16),
    strip.text = element_text(face = "bold")
  )
```




```
prevalence_depr <- dep_anx %>%
  filter(cause == "Depressive disorders", Country %in% c("Netherlands", "Germany", "Sweeden", "Greece"))
  group_by(Country, Year)

ggplot(prevalence_depr, aes(x = Year, y = val_percent, color = Country)) +
  geom_line(size = 0.5) +
  geom_point(size = 1) +
  scale_y_continuous(labels = scales::label_comma()) +
  labs(title = "Percent of Depressive disorders by Year",
       x = "Year", y = "Percent of Cases", color = "Country") +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold"))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

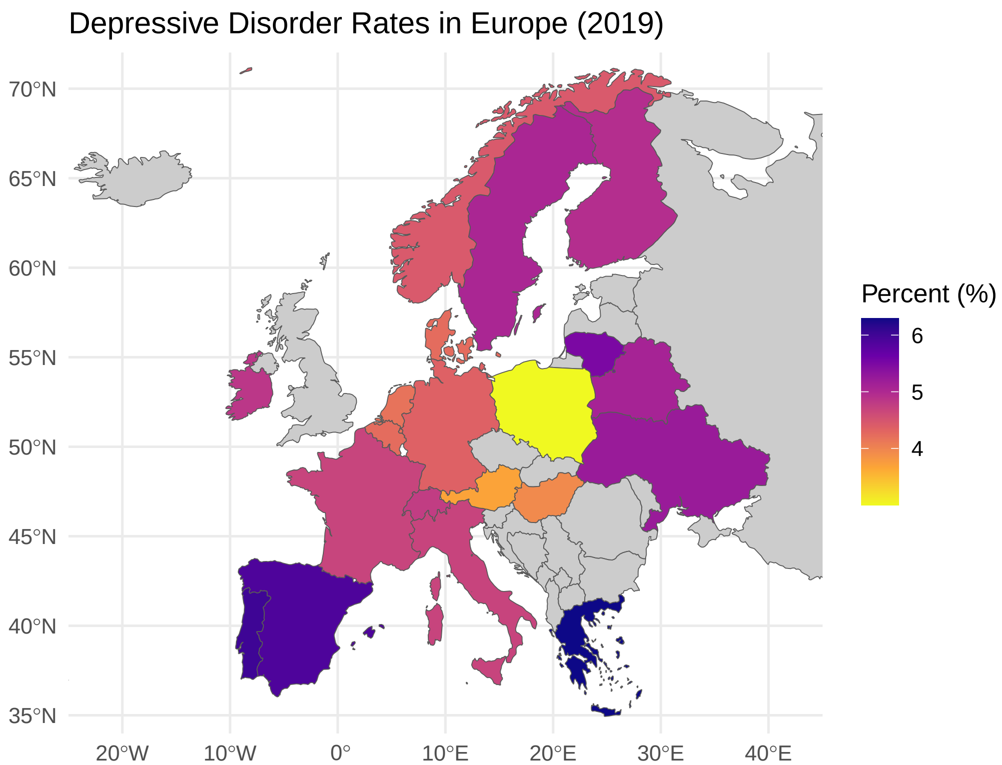



```r
# Load world map
world <- ne_countries(scale = "medium", returnclass = "sf")

# Filter to Europe and remove tiny land masses (area < 10,000 km^2)
europe <- world %>%
  filter(continent == "Europe") %>%
  filter(scalerank <= 6)  # removes tiny countries/islands like Vatican, Monaco, etc.

europe <- europe %>% filter(geometry %>% st_area() > units::set_units(10000, km^2))
# Join with mental health data

dep_map_data <- merged_data %>%
  filter(cause == "Depressive disorders", Year == 2019) %>%
  select(Country, val_percent)

# Join with shapefile
europe_dep <- left_join(europe, dep_map_data, by = c("name" = "Country"))
# Plot the clean European map
ggplot() +
  geom_sf(data = europe, fill = "grey90", color = "white") +
  geom_sf(data = europe_dep, aes(fill = val_percent)) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey80", direction = -1) +
  coord_sf() +  # <--- ensures proper map projection handling
  theme_minimal(base_size = 12) +
  coord_sf(xlim = c(-25, 45), ylim = c(34, 72), expand = FALSE) +
  labs(
  title = "Depressive Disorder Rates in Europe (2019)",
```

```
  fill = "Percent (%)"
)
```

3.1.2 Spatial Analysis (Maps)

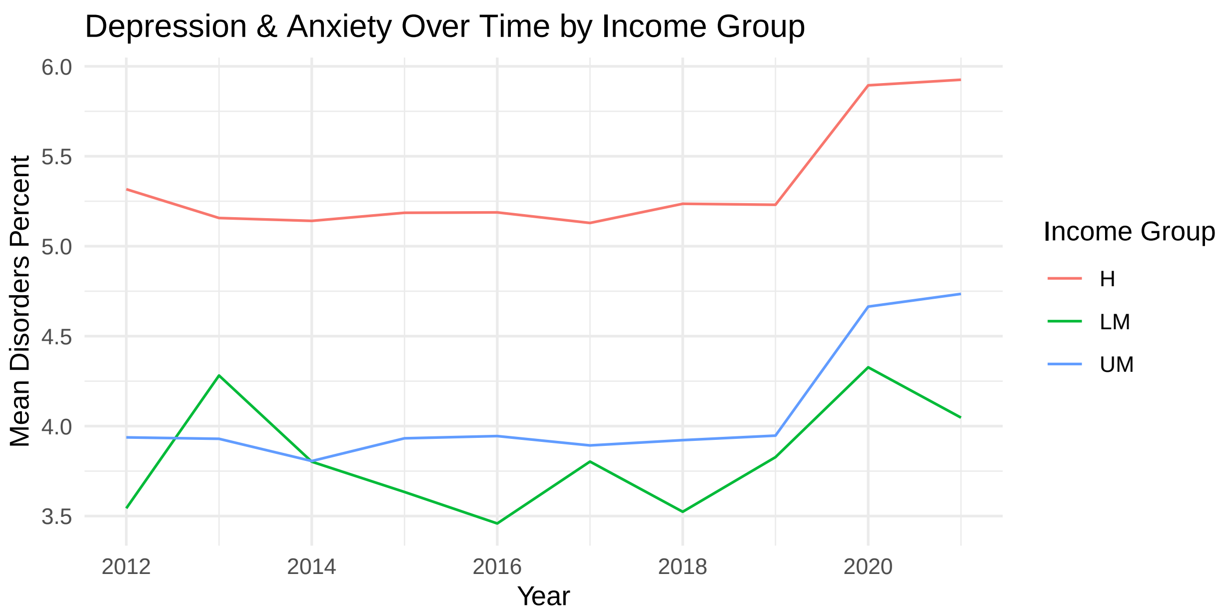
```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```



3.1.3 Sub-population analysis

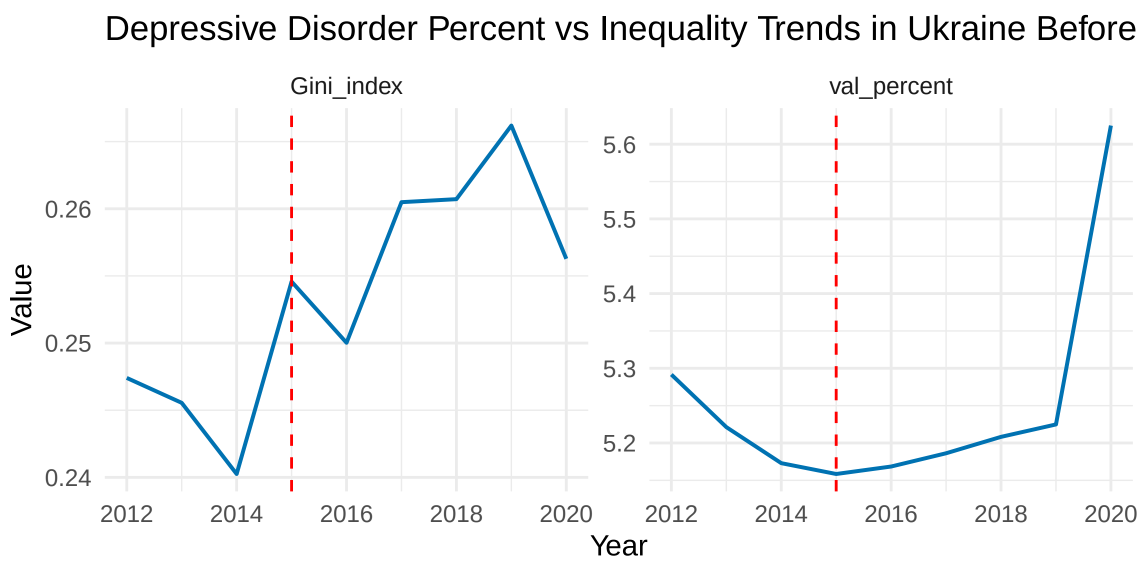
```
# Analyze the relationship between two variables.
# Trend in disorder over time by income group
dep_anx %>%
  filter(!is.na(income_group)) %>%
  group_by(Year, income_group) %>%
  summarise(mean_disorder = mean(val_percent, na.rm = T)) %>%
  ggplot(aes(x = Year, y = mean_disorder, color = income_group)) +
  geom_line(size = 0.5) +
  labs(
    title = "Depression & Anxiety Over Time by Income Group",
    x = "Year",
    y = "Mean Disorders Percent",
    color = "Income Group"
  ) +
  theme_minimal()
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



3.2 Event analysis

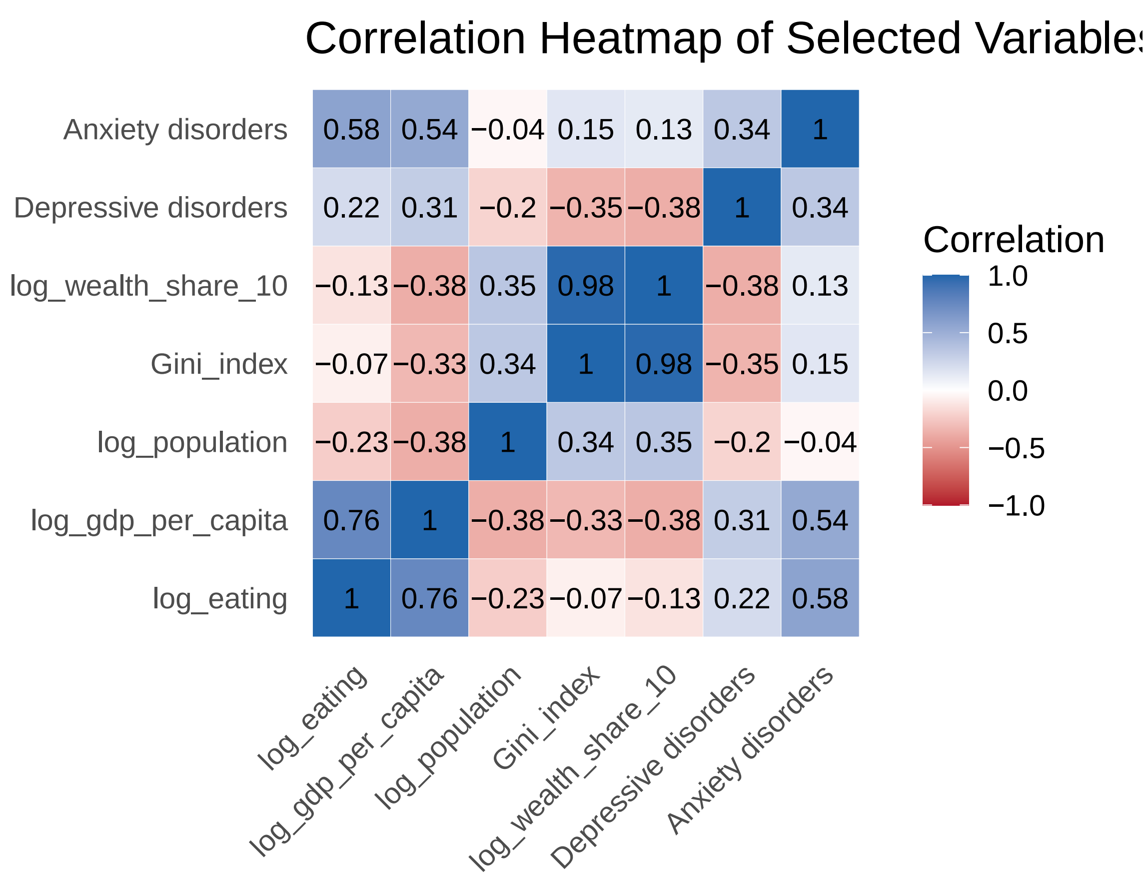
```r
# Reshape Ukraine-only data to long format for multiple indicators
ukraine_long <- merged_data %>%
  filter(Country == "Ukraine", cause == "Depressive disorders") %>%
  select(Year, val_percent, Gini_index) %>%
  pivot_longer(cols = -Year, names_to = "variable", values_to = "value")

ggplot(ukraine_long, aes(x = Year, y = value)) +
  geom_line(color = "#0072B2", size = 0.7) +
  geom_vline(xintercept = 2015, linetype = "dashed", color = "red") +
  facet_wrap(~ variable, scales = "free_y") +
  labs(
    title = "Depressive Disorder Percent vs Inequality Trends in Ukraine Before and After Conflict (2014
    x = "Year",
    y = "Value"
  ) +
  theme_minimal()
```




```r
disorders_wide <- merged_data %>%
  filter(cause %in% c("Depressive disorders", "Anxiety disorders", "Eating disorders")) %>%
  select(Country, Year, cause, val_percent, gdp_per_capita, pop_est, Gini_index, wealth_share_10) %>%
  pivot_wider(names_from = cause, values_from = val_percent)

merged_data_log <- disorders_wide %>%
  mutate(
    log_eating = log(`Eating disorders` + 1),
    log_gdp_per_capita = log(gdp_per_capita + 1),
    log_population = log(pop_est + 1),
    log_wealth_share_10 = log(wealth_share_10 + 1)
  )
```

```r
# Calculate correlation matrix
cor_matrix_log <- merged_data_log %>%
  select(log_eating, log_gdp_per_capita, log_population,
         Gini_index, log_wealth_share_10,
         `Depressive disorders`, `Anxiety disorders`
      ) %>%
  cor(use = "complete.obs") %>%
  round(2)
# Melt the correlation matrix
cor_df_log <- melt(cor_matrix_log)

# Create heatmap
heatmap <- ggplot(cor_df_log, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(
    low = "#B2182B", high = "#2166AC", mid = "white",
    midpoint = 0, limit = c(-1, 1), space = "Lab",
    name = "Correlation"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
```

```r
    panel.grid = element_blank()
  ) +
  coord_fixed() +
  geom_text(aes(label = value), color = "black", size = 4) +
  labs(
    title = "Correlation Heatmap of Selected Variables",
    x = NULL,
    y = NULL
  )
# Display
print(heatmap)
```



4 Discussion

Result: get a sharp, labeled heatmap with intuitive coloring:

Red = strong negative correlation

Blue = strong positive

White = neutral

4.1 Discuss your findings

5 Reproducibility

5.1 Github repository link

https://github.com/qgelena/Quantifying_a_social_problem/tree/main

5.2 Reference list

- Kawachi, I., & Berkman, L. F. (2000). Social cohesion, social capital, and health. *Social Epidemiology*, 174–190.
- Lund, C., Breen, A., Flisher, A. J., Kakuma, R., Corrigall, J., Joska, J. A., … & Patel, V. (2010). Poverty and common mental disorders in low and middle income countries: A systematic review. *Social Science & Medicine, 71*(3), 517–528. https://doi.org/10.1016/j.socscimed.2010.04.027
- Patel, V., Burns, J. K., Dhingra, M., Tarver, L., Kohrt, B. A., & Lund, C. (2018). Income inequality and depression: A systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry, 17*(1), 76–89. https://doi.org/10.1002/wps.20492
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., … & UnÜtzer, J. (2022). The Lancet Commission on global mental health and sustainable development. *The Lancet, 392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736%2818%2931612-X
- Smith, K. E., Bambra, C., Hill, S. E., & Watt, R. G. (2012). Health inequalities and the social determinants of health: What works? *Journal of Public Health, 34*(4), 523–529. https://doi.org/10.1093/pubmed/fds052
- Wilkinson, R., & Pickett, K. (2009). *The Spirit Level: Why more equal societies almost always do better*. London: Allen Lane.
- World Health Organization. (2023). *Mental health*. Retrieved from https://www.who.int/health-topics/mental-health