Final Project Report

Brian Zhou

12/9/2020

Introduction & Background

The histories of colonization and imperialism have profoundly shaped the modern world, leaving a far-reaching footprint in wide ranging fields such as global health. Various authors in postcolonial studies, medical anthropology, and cultural psychiatry have described the *historical trauma* and associated mental health disorders (particularly clinical depression, anxiety, suicide, alcoholism, and other *"postcolonial disorders"*) that are left in the wake of colonial occupation. While there has been a decades-long discussion of how the dehumanizing conditions and traumas of colonization are imprinted into the collective psyche of colonized populations (most notably beginning with *Frantz Fanon*), few quantitative studies have been conducted on the relationship between colonial occupation, violent resistance, and the long-term prevalence of mental disorders. This final project for the Gov 51: Data Analysis and Politics class seeks to address whether the circumstances around colonization and independence in a formerly occupied country affect mental health. Do rates of depression, anxiety, and mental disorders vary based on whether a country was colonized? Does a longer period of independence from a former colonizer lead to better mental health outcomes? Can violent resistance against an oppressor actually restore the humanity of colonized peoples and their long term health, as Fanon suggested?

My hypothesis is two-fold: first, countries that were never colonized (or perhaps acted as the colonizer) are expected to have lower rates of mental disorders. I speculate that colonized countries experienced particular trauma, wars, and violence associated with colonization, and thus the collective mental health of those countries would be negatively. Second, I hypothesize that countries that have had longer periods of independence have lower rates of mental health disorders compared to countries that only recently gained their independence. The likely reason would be that having more time since independence creates more opportunity for the traumas and violent legacy of colonization to heal, and thus results in less mental health problems over time.

My analysis here will explore regressions between whether a country has been colonized (colonial occupation status), how much time has passed since decolonization (years since independence), whether the process of independence for a colonized country was violent, how prevalent mental disorders are today, and other control variables. A positive, statistically significant coefficient for colonial occupation status as a variable for the rate of mental disorders would provide evidence for my first hypothesis, while a negative coefficient would prove that the opposite of my hypothesis is true. Similarly, a negative, statistically significant coefficient for time since independence in a regression for the rate of mental disorders would support my second hypothesis, while a positive coefficient would indicate that mental health disorders actually increase as more time since independence passes. For each regression, if $p < 0.05$ for a given coefficient and there is no statistically significant effect, I will not have sufficient evidence to reject the null hypothesis that years since independence and colonial occupation status do not affect the prevalence of mental health disorders.

Data Sources & Overview

Data Collection

My sample combines data from 2 primary sources: the *Issue Correlates of War (ICOW) colonial history dataset*, which includes information on the independence date, types of independence, colonial rulers, and more for all countries; and the *Global Burden of Disease datasets* that contain measurements of the prevalence of mental health disorders around the world. My unit of analysis is each country listed in the ICOW colonial history data set, each of which was categorized by Paul Hensel's research using historical information This data source was particularly chosen as one of the few datasets available for extrapolating several explanatory variables of interest, including the colonial occupation status of each country and specifically the amount of time that passed since independence, which was recalculated as a numerical variable with the present date subtracted from the date of independence. These variables, combined with specific information on the colonial ruler of each country and the type of independence (violent vs. nonviolent; decolonization, secession, formation, vs. partition) made the ICOW colonial history data set a powerful tool that could be easily cleaned and regressed alongside additional data on mental health disorders.

The remaining data was taken from publicly available data sources at the the Institute for Health Metrics and Evaluation (IHME), an independent population health research center that provides rigorous measurement of the world's most important health problems. Each year, IHME and its partners conduct a survey for the Global Burden of Disease study, which is the most comprehensive observational epidemiological study of over 100 diseases around the world. Specific data on the prevalence of mental health disorders (categorized by depression, schizophrenia, anxiety, etc.) is readily *downloadable* and estimated based on complex, algorithmic *disease modeling softwares*. I chose to focus my outcome variable on the most recent 2019 data for a cross-sectional study of the prevalence of mental disorders, measured as a rate of the total number of cases per 100,000 people, because I hoped the category of mental health disorders would be wide-encompassing and robust enough to capture the overall mental health status of various countries. While using depression, schizophrenia, or more particular disease etiologies as the dependent variable could prove informative, the aggregate measurement for mental disorders includes different types of mental health issues and should hopefully be holistic enough to account for the collective psychic landscape of each country.

To control for possible confounders, I was also able to find estimates on total health spending per GDP, development assistance for public health expenditures, and other measures of global health financing in each country using additional IHME datasets from the *Global Health Data Exchange*. These variables were re-scaled and chosen as possible controls in the regression, considering the obvious impact of financial resources on health outcomes. While broader measurements of a country's economic and health infrastructure, such as GDP or the Gini coefficient, were not included explicitly, I hope that the more specific variables on total health spending, particularly as a percentage of overall GDP, will implictly take into account other possible confounding variables in my regression.

Summary Statistics of Relevant Variables

Below are a few graphical representations of the relevant independent and dependent variables to contextualize the distribution of data in the merged dataframe:

The majority of countries around the world have been colonized. England has the largest colonial footprint, with France, Spain, and Turkey the other predominant colonial powers that colonized many countries.

Figure 1: Distribution of Colonized and Free States by Last Colonial Ruler
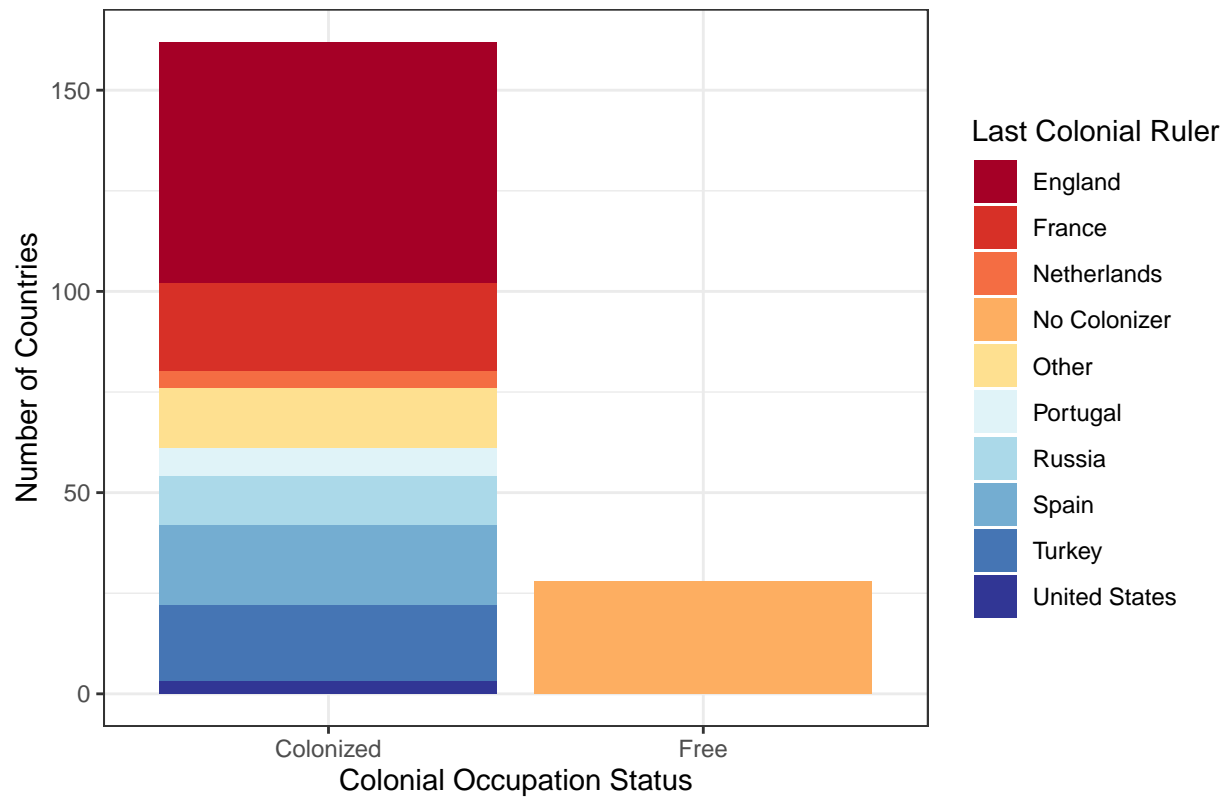
Most countries that were occupied by colonial rulers gained independence through some form of decolonization, the large majority of which were nonviolent processes.

Figure 2: Types of Independence Across Countries
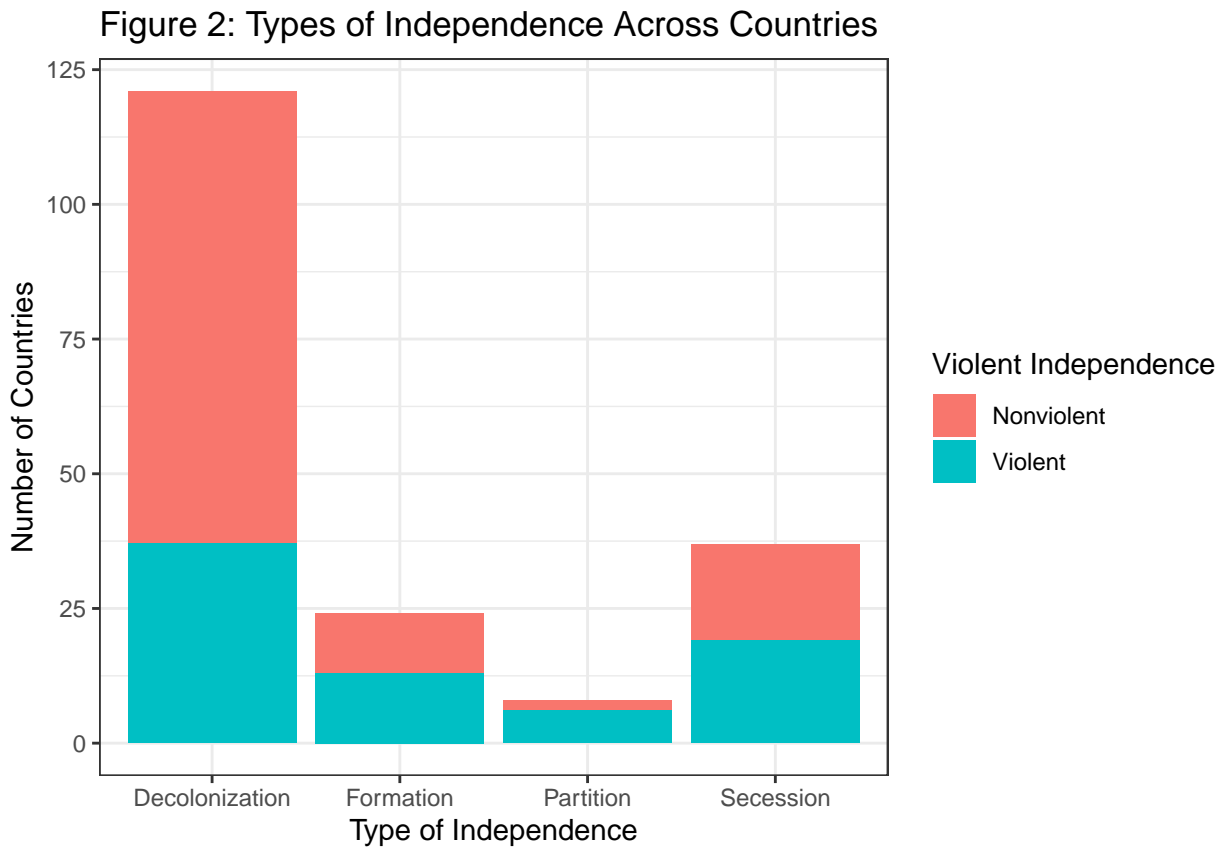
To clarify according to the ICOW codebook, "decolonization" refers to an entity that was a dependency ruled by a foreign power (ex: traditional colonies like India in the British empire); "formation" refers to an entity formed from other entities that have no direct analog (ex: the UK became independent with seceding from any other international actor); "secession" refers to an entity that was part of another state before achieving independence (ex: Republics of Soviet Union leaving the USSR); and "partition" refers to an entity partitioned out of another state with the original state disappearing (ex: North and South Korea).

Of the 115 countries that experienced nonviolent independence, there is a lower average rate of mental health disorders.

Table 1: Summary of Countries for Types of Independence

Type of Independence	Number of Countries	Average Rate of Mental Health Disorders (per 100K)
Nonviolent	115	12482.35
Violent	75	12896.58

Interestingly enough, countries that were formerly colonized have a slightly lower distribution of rates of mental health disorders. However, this could be due to higher reporting and awareness of mental health disorders in countries that were never colonized, which tend to be wealthier (and whiter) nations.

Figure 3: Distribution of Mental Health Burden in Countries With and Without Colonial History
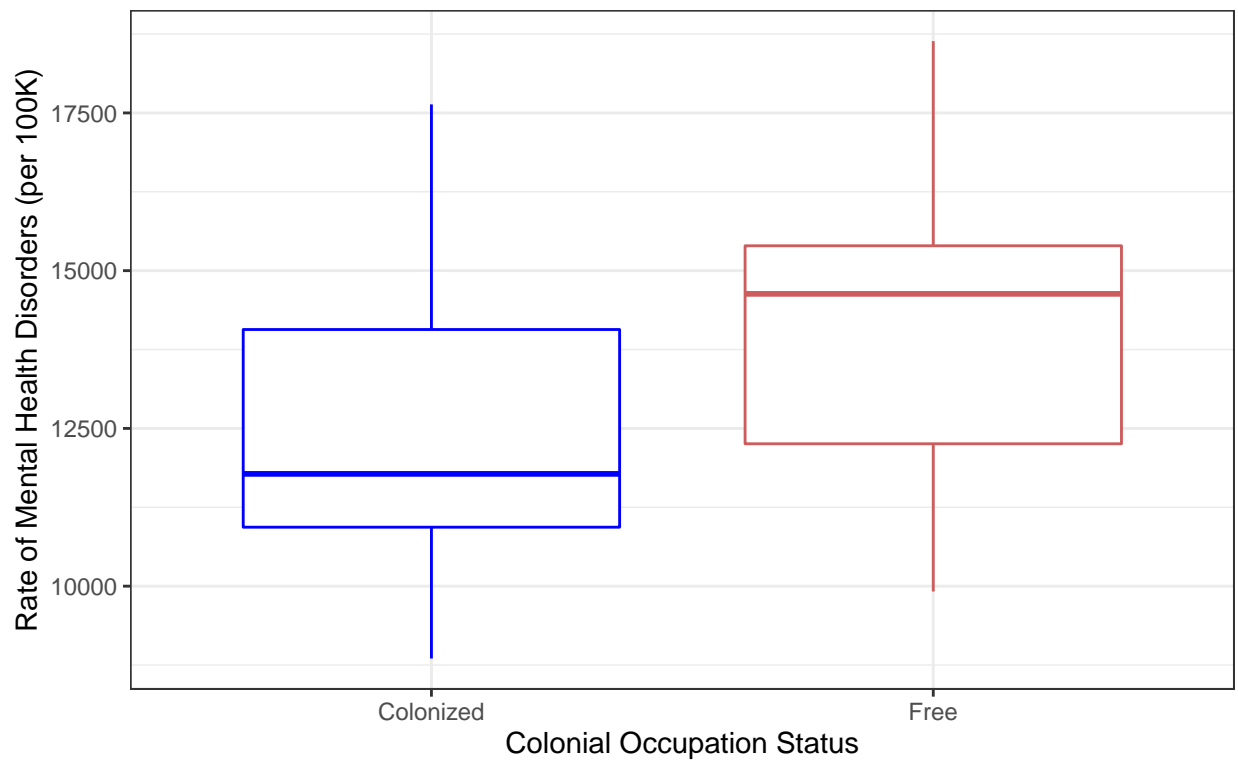
Results

Colonization Is Associated With Higher Rates of Mental Health Disorders

I first present regression estimates of the relationship between colonization and prevalence of mental disorders. Because the main independent variable here is whether or not a country had ever been colonized, each country was sorted into a binary variable for either having a previous foreign occupier (in most cases, a colonizing country) or not. A multiple variable regression model was used, with multiple control variables around global health financing included, in order to determine whether factors such violent independence or various financial considerations may affect mental health disorders. A subset of all the global health financing variables were chosen to represent various confounding forces: total health spending per GDP accounts for GDP and broader economic determinants; development assistance for health spending represents foreign investment and captures some element of international relations; government spending includes domestic investments; and out of pocket expenditures reflects individual payments.

Table 2 suggests that a country that has been colonized has a higher likelihood of taking on a greater mental health burden. Countries that were subjected to colonial rulers have around 1020 more cases of mental disorders per 100,000 people compared to countries that were never colonized. The magnitude of this coefficient is statistically significant, likely causal, and substantial; considering the other control variables, each $1 Billion USD of Development Assistance for Health Expenditures (adjusted to Purchasing Power Parity in 2019) is associated with 788 less mental disorders, which would actually fall short of countering the number of mental health disorders produced by colonial occupation. Of course, this is a crude estimate, but the scale of colonial occupation cannot be ignored here, especially considering the likelihood of under-reporting for mental disorders in previously colonized countries.

Unsurprisingly, the coefficient for violent independence was positive, but the standard error is large enough that we can assume that whether the process of independence is violent or not has very marginal, if any, effect on rates of mental health disorders. This could indicate that additional trauma associated with violent independence has no long term effect on mental health disorders. Indeed, total health spending and development assistance appear to have a larger and statistically significant effect. Interestingly enough, the effect of more total health spending per GDP was higher rates of mental health disorders - a rather counter-intuitive effect that should merit further investigation. Proposing a mechanism for this unexpected result is beyond the scope of this paper, but perhaps higher income countries that have greater health spending tend to have greater rates of non-communicative diseases (including mental health disorders) rather than communicative diseases that characterize lower income countries. This would also explain why development assistance is different and associated with less mental health disorders, considering development assistance goes toward predominantly lower income countries as foreign aid. Lastly, it is worth mentioning that government spending and out of pocket spending appear to have no statistically significant effect on the prevalence of mental health disorders.

Table 2: Relationship Between Colonization and Mental Disorders Regressed by Multiple Variables

	Model 1
(Intercept)	11206.802***
	(449.155)
Colonial Occupation Status	1020.567**
	(440.214)
Violent Independence	130.578
	(298.929)
Total Health Spending per GDP (Expressed as Percent)	21390.825***
	(6230.556)
Development Assistance for Health (Per 1 Billion USD - 2019 Purchasing Power Parity)	-788.474***
	(258.394)
Government Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)	-1.336
	(1.906)
Out of Pocket Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)	5.732
	(5.790)
Num.Obs.	184
R2	0.200
R2 Adj.	0.172

* p < 0.1, ** p < 0.05, *** p < 0.01

More Time Since Independence Is Associated With Higher Rates of Mental Health Disorders

Next, I explore the relationship between the number of years since independence and the rate of mental health disorders. I follow the same methods as in the first regression, but the main variable of interest is now continuous and represents the amount of time that passed since decolonization and independence. I include the same control variables as above, and the data has been subseted to remove countries that were never colonized.

Table 3 shows that each additional year since independence has an effect of increasing mental health disorders by around 12 cases per 100,000 people. This is statistically significant to p < 0.01 and seems counter-intuitive, considering my theory that having more time since independence creates more room for the traumas of colonization to heal, and thus results in less mental health problems over time. While this obviously does not mean that colonized countries were better off under colonial rule, it does suggest that perhaps the colonial footprint in public health infrastructure and collective mental wellbeing lasts long past colonization. Of course, the slight increase in mental health disorders could also be due to increases in population and awareness in mental health over time, which may not be accounted for in this regression.

The statistical significance and magnitude of the control variables are relatively unchanged compared to the last regression. Notably, the coefficient for violent independence is negative in this regression, even accounting for the margin of error. Despite the effect of violent independence not being statistically significant, there is still some indication that a violent process of independence is associated with less mental health disorders. This would be consistent with several conjectures in postcolonial theory, including Frantz Fanon's argument that violence used against the colonizer is cathartic and helps the oppressed regain their humanity (and perhaps some of their long-term mental wellbeing).

Figure 2 visualizes the distribution of mental health disorders across the number of years since independence. One thing to note is that the distribution of time since independence is skewed more towards the lower end; there are only a couple countries that have had more than 200 years of independence, which results in a large standard error range for the trend line towards the higher end of years since independence.

Table 3: Mental Disorders in Colonized Countries Regressed by Years of Independence and Other Variables

	Model 1
(Intercept)	10867.898***
	(427.705)
Years of Independence	12.328***
	(2.418)
Violent Independence	-501.292
	(319.548)
Total Health Spending per GDP (Expressed as Percent)	13011.095**
	(5963.905)
Development Assistance for Health (Per 1 Billion USD - 2019 Purchasing Power Parity)	-791.299***
	(256.926)
Government Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)	-2.861
	(2.237)
Out of Pocket Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)	15.405*
	(8.576)
Num.Obs.	158
R2	0.283
R2 Adj.	0.255

$* \ p < 0.1, \ ** \ p < 0.05, \ *** \ p < 0.01$

Figure 2: Mental Health Burden Based on Independence Year
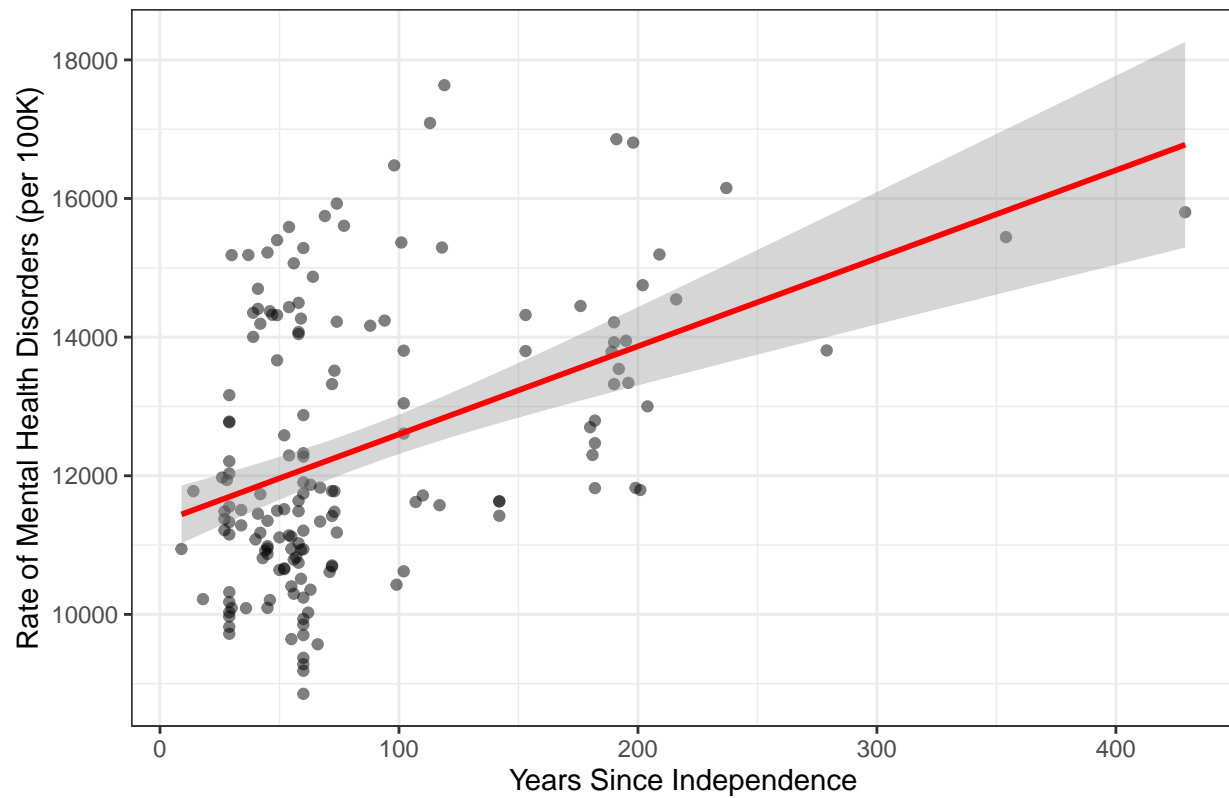
Conclusion

Colonization produces long-term effects and structural issues, including for public health systems that must contend with legacies of historical trauma and strive for collective well-being. This study is one of the first to quantify the relationship between colonization, violent independence, and years since decolonization with the modern-day prevalence of mental disorders. My first hypothesis that countries that were never colonized fare better in mental health is supported by the statistically significant finding that colonization caused over 1000 additional cases of mental disorders per 100,000 people in 2019. Next, my second hypothesis that more time since independence would lead to lower rates of mental health disorders was disproven. My analysis actually shows that each year of independence actually increases the number of cases of mental health disorders by 12 per 100,000 people. Although both of the above coefficients in my regression are significant to $p < 0.05$, we should still be cautious about causal claims, considering possible confounders such as increases in population or awareness of mental health issues over time. Other assumptions about the accurate reporting and diagnoses of mental health disorders would also have to be made in order to validate a causal link between colonization and prevalence of mental disorders. Lastly, it is important to remember that there are challenges and subjective judgments that arise when trying to categorize complicated issues like colonial history into a quantitative dataset. It is difficult to judge which foreign power(s) to mark as the "colonizer" and determine an exact date of independence; for instance, both Morocco and Tunisia are coded as being independent for over 200 years, despite being occupied by France and experiencing a substantial push towards decolonization in the mid-1900s. Such intricacies are erased in forcing the twists and turns of history into a numerical data-frame, so further investigations should take into account different periods of occupation and colonization for each country, while comparing rates of mental health disorders overtime (particularly following the first few years after decolonization and independence). This study represents an important start to quantitative examinations of colonization and mental disorders, but future studies should build on these preliminary findings to investigate causal relationships and explore the impact of colonial empires on global health.

Code

```r
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(stargazer)
library(modelsummary)
# Import the dataset on colonial history.
coldata <- read_csv("data/ICOW Colonial History 1.1/coldata_renamed.csv") %>%
  select(Name, IndDate, ColRuler, IndViol, IndType) %>%
  rename(country = Name)

# Import the mental health disorder rates dataset.
mentaldisorders <- read_csv("data/GBD Data/Mental Disorder_Prevalence_Percent_IHME-GBD_2019/Rate_Prevale
  select(location_id, location_name, val) %>%
  rename(country = location_name)

# Import the global health financing data.
financing <- read_csv("data/GBD Data/Global Health Financing/Dataset_IHME_RETROSPECTIVE_FINANCING_1995_
  select(location_name, location_id, year, the_total_ppp_mean,
         ghes_total_ppp_mean, oop_total_ppp_mean, dah_total_ppp_mean,
         the_per_cap_mean, the_per_gdp_mean) %>%

  # Filter to the most recent year: 2017.
  filter(year == 2017) %>%

  # Rescale monetary variables in billions of USD.
  mutate(dah_total_ppp_mean = dah_total_ppp_mean / 1000000,
         ghes_total_ppp_mean = ghes_total_ppp_mean / 1000000,
         oop_total_ppp_mean = oop_total_ppp_mean / 1000000)


## Merge the datasets.

# Some countries are missing in the "financing dataset" - keep the values in the
# "mental disorders dataset with a right_join and use missing values for the
# countries that are missing.
gbd_data <- right_join(financing, mentaldisorders, by = "location_id")


# Merge GBD Data with colonial data set.
newdata <- inner_join(coldata, gbd_data, by = "country") %>%

  # Categorize countries that did or did not have a colonial ruler.
  mutate(colonized = ifelse(ColRuler > 0, "Colonized", "Free")) %>%
    # Should I take out other "colonized" Western countries?
    # filter(country != "Canada",
        #   country != "Australia",
        #   country != "United States") %>%

  # Calculate the time since independence
  mutate(indep_date = as.integer(str_sub(IndDate, start = 1L, end = -3L)),
         years_indep = 2020 - indep_date) %>%
```

```r
  # Rename variable for colonial ruler in terms of country.
  mutate(ColRuler = case_when(
    ColRuler == 230 ~ "Spain",
    ColRuler == 210 ~ "Netherlands",
    ColRuler == 220 ~ "France",
    ColRuler == 2 ~ "United States",
    ColRuler == 640 ~ "Turkey",
    ColRuler == 365 ~ "Russia",
    ColRuler == 200 ~ "England",
    ColRuler == 235 ~ "Portugal",
    ColRuler == -9 ~ "No Colonizer")) %>%
  mutate(ColRuler = ifelse(is.na(ColRuler) == TRUE, "Other", ColRuler))

# Filter a subset for just colonized countries and also when the type of
# independence was decolonization.
colonized <- newdata %>%
  filter(colonized == "Colonized")



# Barplot for colonized countries.
ggplot(newdata, aes(x = colonized, fill = ColRuler)) +
  geom_bar() +
  labs(title =
"Figure 1: Distribution of Colonized and Free States by Last Colonial Ruler",
       x = "Colonial Occupation Status",
       y = "Number of Countries") +
  theme_bw() +
  guides(fill = guide_legend(title = "Last Colonial Ruler")) +
  scale_fill_brewer(palette = "RdYlBu")



# Barplot for types of independence.
newdata %>%
  mutate(IndViol = ifelse(IndViol == 1, "Violent", "Nonviolent")) %>%
  mutate(IndType = case_when(
      IndType == 1 ~ "Formation",
      IndType == 2 ~ "Decolonization",
      IndType == 3 ~ "Secession",
      IndType == 4 ~ "Partition"
  )) %>%
  ggplot(aes(x = IndType, fill = IndViol)) +
    geom_bar() +
    labs(title = "Figure 2: Types of Independence Across Countries",
         x = "Type of Independence",
         y = "Number of Countries") +
    guides(fill = guide_legend(title = "Violent Independence")) +
    theme_bw()
```

```r
# Show summary table of mental health disorder rates, subsetted by
# violent/nonviolent independence.
sum_table <- newdata %>%
  mutate(IndViol = ifelse(IndViol == 1, "Violent", "Nonviolent")) %>%
  group_by(IndViol) %>%
  summarize(countries = n(),
            rate = mean(val))

knitr::kable(sum_table, caption =
                 "Summary of Countries for Types of Independence",
             col.names =
                 c("Type of Independence", "Number of Countries",
                   "Average Rate of Mental Health Disorders (per 100K)"))



# Boxplot for Mental health rates in colonized and free countries.
ggplot(newdata, aes(x = colonized, y = val, color = colonized)) +
  geom_boxplot(color = c("blue", "indianred")) +
  labs(title = "Figure 3: Distribution of Mental Health Burden in Countries
       With and Without Colonial History",
       x = "Colonial Occupation Status",
       y = "Rate of Mental Health Disorders (per 100K)") +
  theme_bw()

# Fit regression based on colonization.
fit_col <- lm(val ~ colonized, newdata)

# Fit multivariable regression on colonization.
fit_col_multiple <- lm(val ~ colonized + IndViol + the_per_gdp_mean
                        + dah_total_ppp_mean + ghes_total_ppp_mean
                        + oop_total_ppp_mean, newdata)

model = c(colonizedFree = "Colonial Occupation Status",
          IndViol = "Violent Independence",
          the_per_gdp_mean =
            "Total Health Spending per GDP (Expressed as Percent)",
          dah_total_ppp_mean =
"Development Assistance for Health (Per 1 Billion USD - 2019 Purchasing Power Parity)",
          ghes_total_ppp_mean =
"Government Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)",
          oop_total_ppp_mean =
"Out of Pocket Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)")

modelsummary::modelsummary(
  fit_col_multiple,
  title = "Relationship Between Colonization and Mental Disorders Regressed by Multiple Variables",
  coef_rename = model,
  gof_omit = "AIC|BIC|Log.Lik.|F",
  stars = TRUE)


# Fit multiple variable regression model.
```

```r
fit_multiple <- lm(val ~ years_indep + IndViol + the_per_gdp_mean +
                       dah_total_ppp_mean + ghes_total_ppp_mean +
                       oop_total_ppp_mean, colonized)

model = c(years_indep = "Years of Independence",
         IndViol = "Violent Independence",
         the_per_gdp_mean =
           "Total Health Spending per GDP (Expressed as Percent)",
         dah_total_ppp_mean =
"Development Assistance for Health (Per 1 Billion USD - 2019 Purchasing Power Parity)",
         ghes_total_ppp_mean =
"Government Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)",
         oop_total_ppp_mean =
"Out of Pocket Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)")

modelsummary::modelsummary(
  fit_multiple,
  title =
"Mental Disorders in Colonized Countries Regressed by Years of Independence and Other Variables",
  coef_rename = model,
  gof_omit = "AIC|BIC|Log.Lik.|F",
  stars = TRUE)


# Plot years since independence to rate of mental health disorders
ggplot(colonized, aes(years_indep, val, label = country)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = lm, color = "red") +
  theme_bw() +
  labs(x = "Years Since Independence",
       y = "Rate of Mental Health Disorders (per 100K)",
       title = "Figure 2: Mental Health Burden Based on Independence Year")
  # geom_label(nudge_y = 50, size = 2)
```