

# Final Project Draft Analysis

Brian

11/12/2020

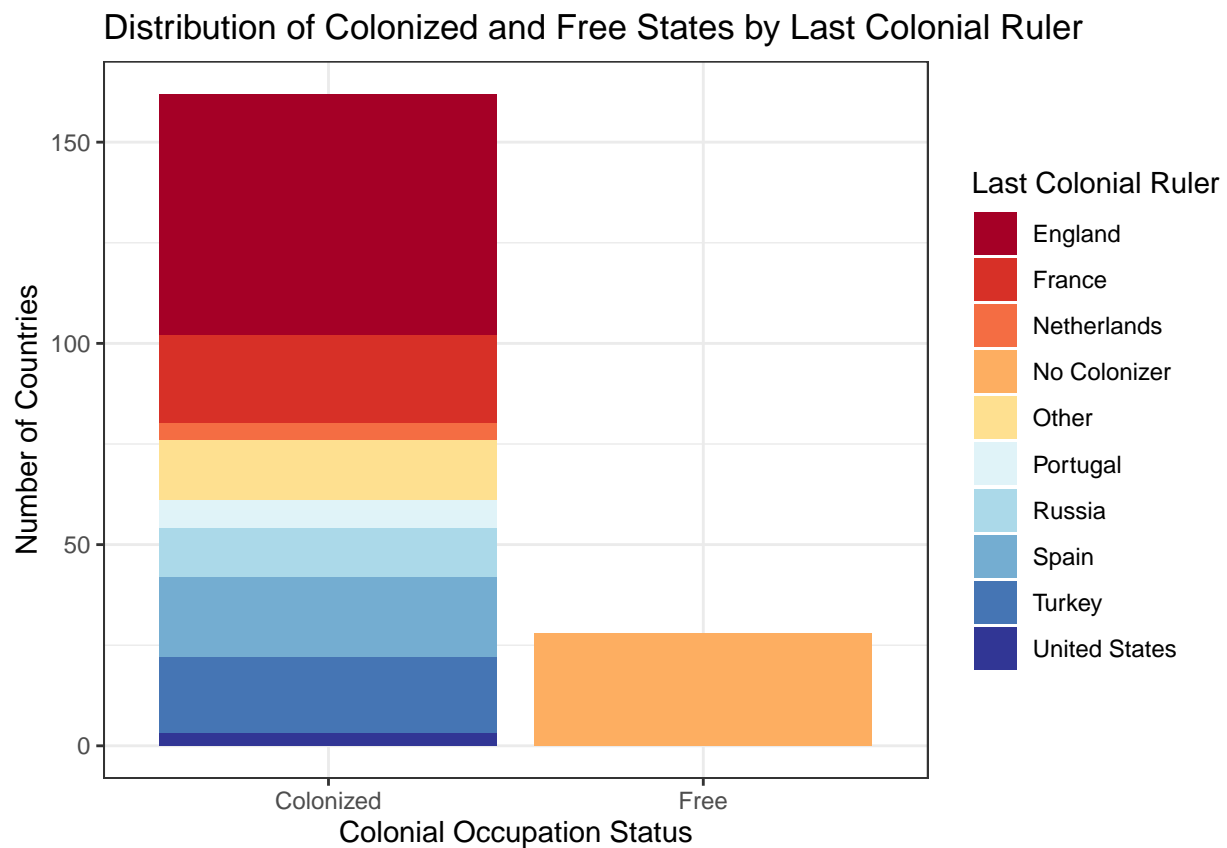
## Importing & Cleaning the Data

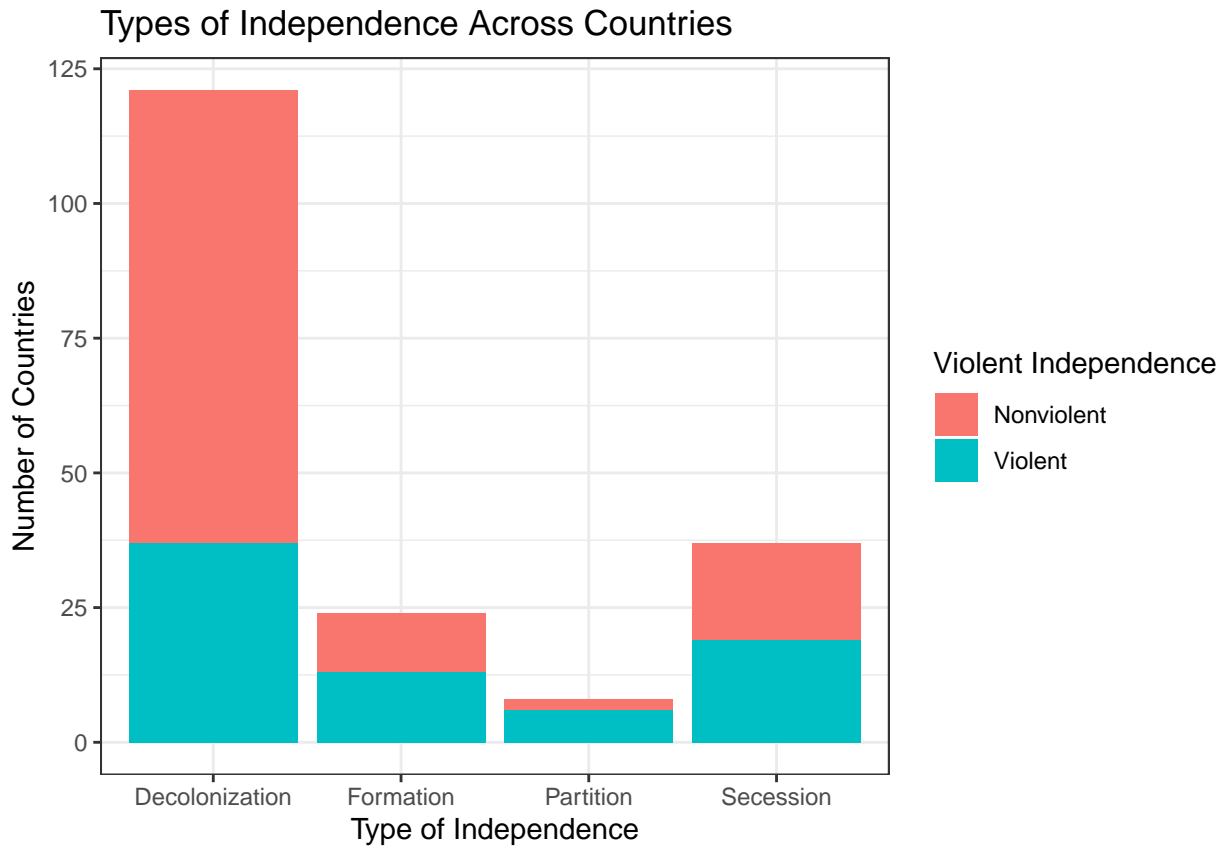
Merged datasets using naming.

```
## Warning: Missing column names filled in: 'X1' [1]
```

**Summary Statistics:** measure center (mean/median) and spread (distribution - range, stdev)

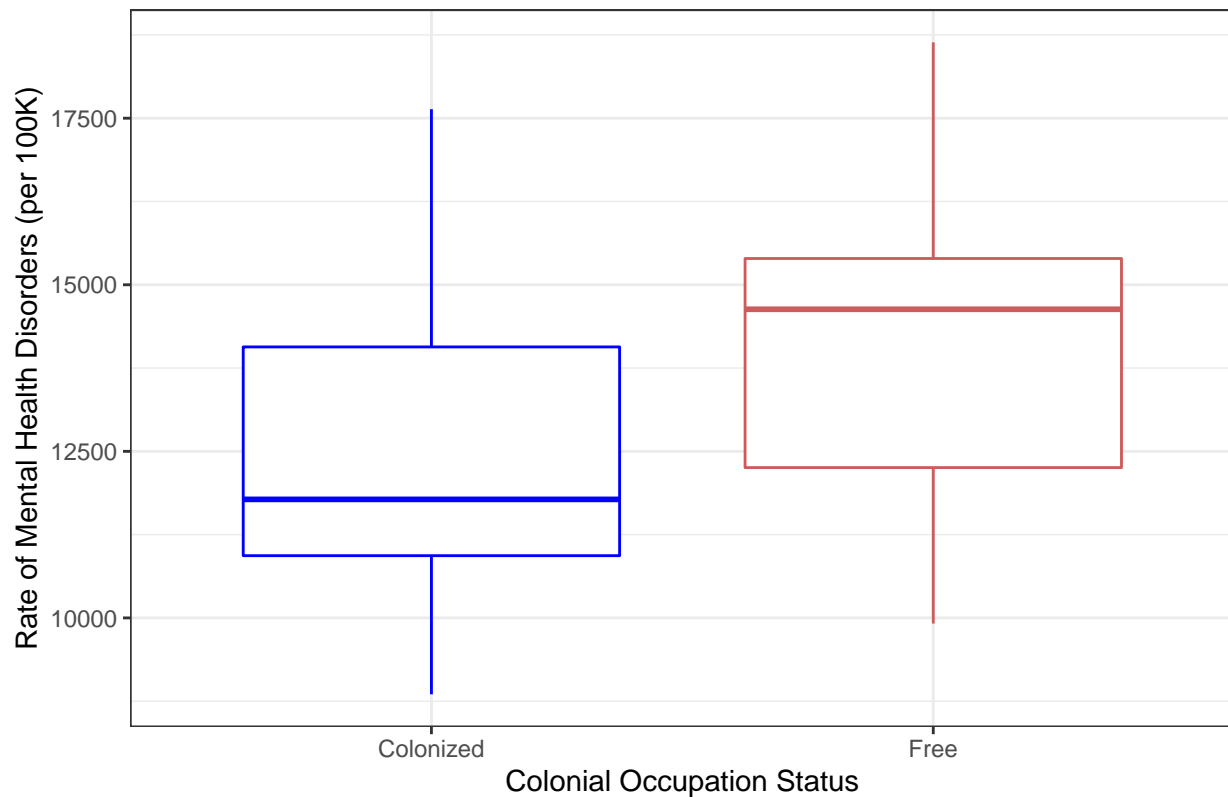
include bar chart for binary viables





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

## Distribution of Mental Health Burden in Countries With and Without Color



## Effect of Colonization on Rate of Mental Health Disorders

```
# 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 = "Mental Health Disorders Regressed by Multiple Variables",
  coef_rename = model,
  gof_omit = "AIC|BIC|Log.Lik.|F",
  stars = TRUE)
```

Table 2: Mental Health 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

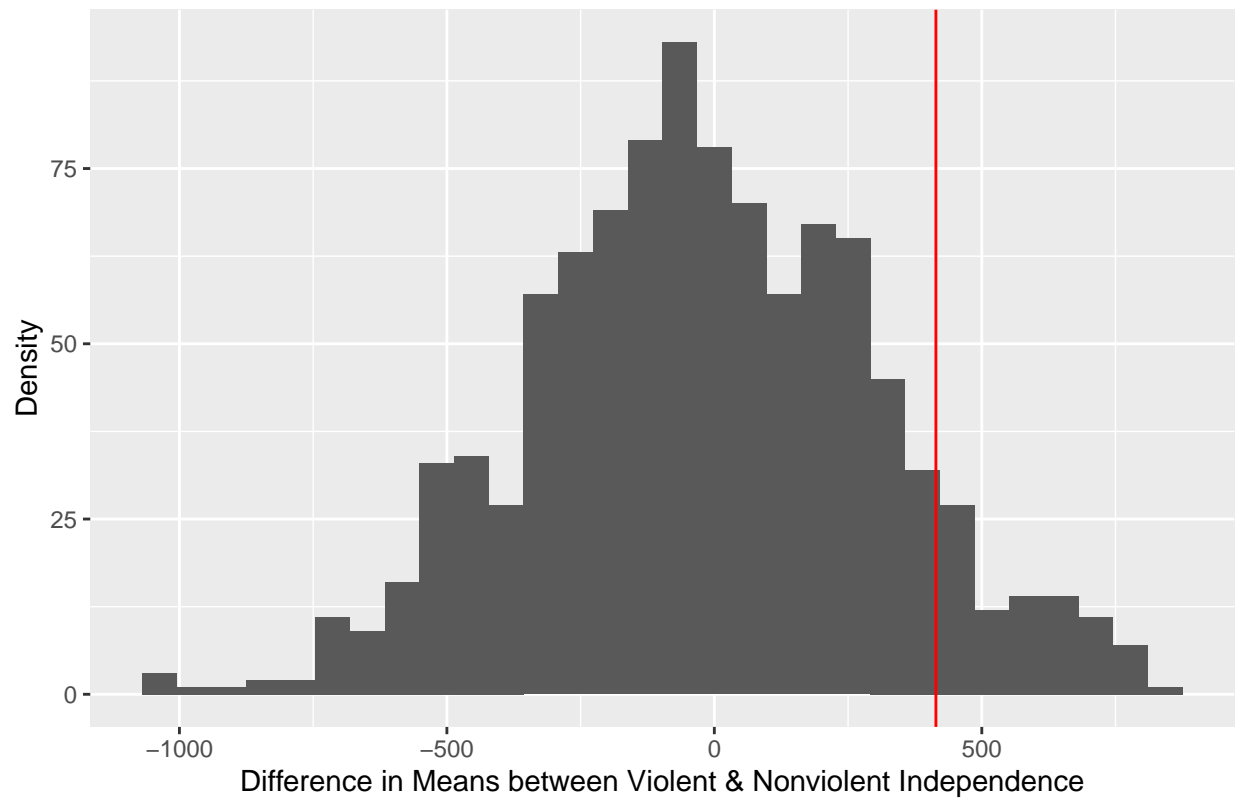
\* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

## Average Treatment Effect for Violent Independence vs. Non-violent Independence

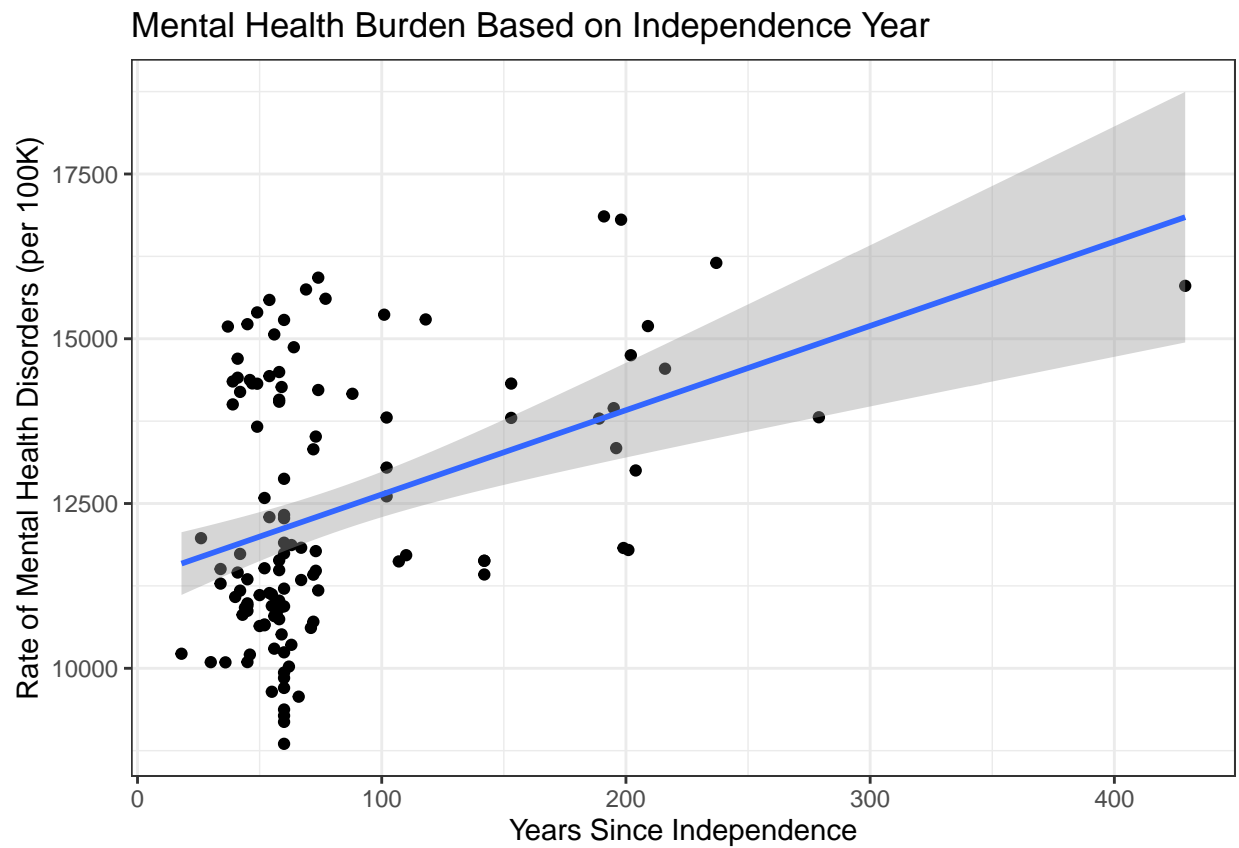
### Permutation Test

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

### Permutation Test for Violent Independence



## Data Analysis



There appears to be a positive relationship between the number of years since independence and the rate of mental health disorders in a country. This is surprising and contradicts my initial hypothesis, considering my theory that having more time since independence creates more room for the traumas and violence of colonization to heal, and thus results in less mental health problems over time.

## Regression Table

Table 3: Mental Health Burden Based on Independence Year and Violent Independence

	<i>Dependent variable:</i>	
	Rate of Mental Health Disorders	
	(1)	(2)
Years Since Independence	12.789*** (2.722)	14.745*** (3.130)
Violent Independence		-527.982 (419.828)
Constant	11,358.330*** (277.621)	11,353.890*** (276.956)
Observations	119	119
R <sup>2</sup>	0.159	0.170
Adjusted R <sup>2</sup>	0.152	0.156
Residual Std. Error	1,815.128 (df = 117)	1,810.633 (df = 116)
F Statistic	22.072*** (df = 1; 117)	11.882*** (df = 2; 116)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

For each year since independence, the rate of mental health disorders per 100K population goes up by about 12-15 cases. The adjusted-R squared values indicate that the linear model is able to explain about 18% of the data.

Interestingly, countries with a violent independence had a coefficient of -690, indicating that having a violent independence is correlated with a rate of 690 per 100K less mental health disorders today. This may potentially validate Frantz Fanon's theory about violence by the oppressor against the colonizer has the potential of being cathartic and restoring the humanity of the colonized, thus positively affecting mental health in the future.

## Additional Ideas & Further Directions

- Analyze the role of violence and whether or not that is correlated with less mental health burden (maybe go back to earlier years)
- Subset the data based on different types of independence (decolonization in particular)
- Cross check mental health rates for countries that were colonizers (or not colonized) – maybe use that as a control, and then calculate the ATE.
- Use data for schizophrenia and particular mental health disorders
- Determine confidence intervals? How does that work for a line graph/ regression??
- Check interpretation of regression tables.

What are possible confounders? Additional variables to include? - Global health financing/ spending - GDP of a country/ income level - Development assistance

Table 4: Mental Disorders in Colonized Countries Regressed by Years of Independence and Other Variables

	Model 1
(Intercept)	11056.814*** (472.328)
Years of Independence	10.830*** (3.174)
Violent Independence	-367.067 (399.701)
Total Health Spending per GDP (Expressed as Percent)	13312.987* (6860.045)
Development Assistance for Health (Per 1 Billion USD - 2019 Purchasing Power Parity)	-981.736*** (273.273)
Government Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)	-4.023* (2.310)
Out of Pocket Spending (Per 1 Billion USD - 2019 Purchasing Power Parity)	20.171** (8.884)
Num.Obs.	115
R2	0.298
R2 Adj.	0.259

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

## Code

```
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_Prevalence_IHME-GBD_2019.csv") %>%
  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_2017.csv") %>%
  select(location_name, location_id, year, the_total_ppp_mean, ghes_total_ppp_mean, oop_total_ppp_mean, dah_total_ppp_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) %>%

# 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") %>%
  filter(IndType == 2)

# Barplot for colonized countries.
ggplot(newdata, aes(x = colonized, fill = ColRuler)) +
  geom_bar() +
  labs(title = "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 = "Types of Independence Across Countries",
         x = "Type of Independence",
         y = "Number of Countries") +
    guides(fill = guide_legend(title = "Violent Independence")) +
    theme_bw()

# 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, col.names = c("Type of Independence", "Number of Countries", "Average Rate of M

# 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 = "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 = "Mental Health Disorders Regressed by Multiple Variables",
  coef_rename = model,
  gof_omit = "AIC|BIC|Log.Lik.|F",

```

```

stars = TRUE)

# Calculate the observed average treatment effect for violence.
viol_ate <- newdata %>%
  group_by(IndViol) %>%
  summarize(mean = mean(val)) %>%
  pivot_wider(names_from = IndViol, values_from = mean) %>%
  mutate(ate = '1' - '0') %>%
  pull(ate)

# Create empty vector for simulation.
nsims <- 1000
viol_test <- rep(NA, nsims)

# Create simulation for permutation test
for (i in 1:nsims) {

  # Randomly reshuffle the treatment indicator (IndViol).
  viol <- newdata %>%
    mutate(shuf_treat = sample(IndViol))

  # Calculate average treatment effect.
  control <- viol %>% filter(shuf_treat == 0)
  treated <- viol %>% filter(shuf_treat == 1)

  # Store this particular ATE in a vector
  viol_test[i] <- mean(treated$val, na.rm = TRUE) -
    mean(control$val, na.rm = TRUE)
}

ggplot(mapping = aes(viol_test)) +
  geom_histogram() +
  labs(title = "Permutation Test for Violent Independence",
       x = "Difference in Means between Violent & Nonviolent Independence",
       y = "Density") +
  geom_vline(xintercept = viol_ate, color = "red")

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

# Fit linear regression models.
fit_indep_date <- lm(val ~ years_indep, colonized)
fit_indep_viol <- lm(val ~ years_indep + IndViol, colonized)

```

```

stargazer::stargazer(fit_indep_date, fit_indep_viol, title =
  "Mental Health Burden Based on Independence Year and Violent Independence",
  covariate.labels =
    c("Years Since Independence", "Violent Independence"),
  dep.var.labels = "Rate of Mental Health Disorders",
  header = FALSE)

# Fit multiple variable regression model.
fit_multiple <- lm(val ~ years_indep + IndViol + the_per_gdp_mean + dah_total_ppp_mean + ghes_total_ppp_mean + oop_total_ppp_mean)

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)

```