Project Proposal: The COVID-19 Epidemic, Public Health Restrictions, and Mental Health

due October 18, 2021 by 11:59 PM

Matt Mohn, Melannie Nimocks, and Katherine Beltz

10/18/2021

Load Packages

```
library(tidyverse)
library(readxl)
library(lubridate)
library(tidymodels)
```

Load Data

```
setwd('../')
restrictions_worldwide <- readr::read_csv("data/phsm-severity-data-short.csv")
google_trends <- readr::read_csv("data/google_trends_data.csv")
```

```
# for each dataset, make the countries standardized
restrictions_worldwide$COUNTRY <- gsub("United States Of America", "United States", restrictions_worldw

restrictions_worldwide$COUNTRY <- gsub("United Kingdom Of Great Britain And Northern Ireland", "United


google_trends <-
  google_trends %>%
  mutate(COUNTRY = nation) %>%
  mutate(DATE_START = week)


match_dates<-merge(restrictions_worldwide, google_trends, by=c("COUNTRY","DATE_START"))

new_set <-
  match_dates %>%
  select(-c(week, nation))

new_set <- new_set %>%
  mutate(date = mdy(DATE_START))
```

Introduction and Data, including Research Questions

From the beginning of the COVID-19 pandemic until now, the global community has suffered social, economic, and medical burdens in unprecedented levels. Though the physical health of individuals has been of paramount concern due to the high infectivity of COVID-19, with 237.88 million cases and 4.85 million deaths in as of October 2021, another burden on individuals, governments, and health systems has manifested itself in the form of rapidly deteriorating mental health (Our World in Data, 2021). It has been widely accepted that as the COVID-19 pandemic has progressed, mental health has decreased (Centers for Disease Control and Prevention, 2021). However, there is a much less comprehensive body of data surrounding how certain mitigation efforts specifically have impacted mental health, and which mental health conditions each restriction affects the most. For example, the Centers for Disease Control and Prevention (2021) acknowledge that social distancing may increase loneliness, stress, and anxiety, but it is less understood if masking is more directly correlated to obsessive compulsive disorder than it is to depression. Therefore, there is a need to fully understand these intricate relationships in order to drive efforts towards creating more individualized mental health treatments, as well as being able to predict what kind of mental health treatment will be needed in response to an increase in any given public health restriction.

Our data analysis will answer the following research question: how do different COVID-19 mitigation efforts correlate to different types of mental illnesses? In doing so, our project will begin to uncover how certain restrictions may impact different mental illness depending on both the type of restriction and the type of mental illness. Our project will make use of two datasets and merge the datasets based on country in order to ensure there are sufficient relationships to explore in the data. The first dataset is of the frequency of different search terms related to mental health from January of 2019 through September of 2021 for a variety of countries. The data was collected from Google Trends and records the popularity of that search term for any given week in a given country. There is a new data value corresponding to each week, where the date collected is marked as the first of the week. This dataset will be used to gauge how concern with certain mental health topics, including anxiety, depression, obsessive compulsive disorder, therapists, and insomnia, has changed over the course of the pandemic. The second dataset tracks implementation of various mitigation efforts in different countries. This dataset was derived from the World Health Organization's tracking of public health and social measures, and indices were calculated on the raw data in order to quantify the intensity of the restriction, whether the restriction is on masks, gatherings, businesses, schools, or travel. This data has daily values from January 2020 through September of 2020. Through a series of data tidying and wrangling steps, the data has been joined on both country and date. Since the search terms are weekly observations whereas restrictions are daily observations, the search term frequency will be kept constant throughout the week, but each observation for restrictions will be observed for changes. This will enable our analysis to observe how daily changes may affect weekly averages without altering or extrapolating data.

Glimpse

```
## This is the PHSM severity index data set.

glimpse(restrictions_worldwide)
```

```
## Rows: 142,506
## Columns: 11
## $ DATE_START   <chr> "8/20/2020", "9/4/2020", "3/13/2021", "10/18/2020", "4/18~
## $ COUNTRY      <chr> "Yemen", "Belarus", "Egypt", "Uzbekistan", "Finland", "Is~
## $ ISO3         <chr> "YEM", "BLR", "EGY", "UZB", "FIN", "IMN", "MLI", "MYS", "~
## $ WHO_REGION   <chr> "EMRO", "EURO", "EMRO", "EURO", "EURO", "EURO", "AFRO", "~
## $ MASKS        <dbl> 0, 67, 100, 100, 47, 0, 0, 67, 0, 0, 0, 67, 0, 100, 0, 10~
## $ TRAVEL       <dbl> 100, 0, 33, 100, 100, 0, 0, 67, 100, 100, 100, 33, 100, 1~
## $ GATHERINGS   <dbl> 30, 25, 50, 25, 5, 0, 0, 5, 0, 25, 0, 0, 0, 50, 0, 25, 25~
## $ SCHOOLS      <dbl> 25, 25, 25, 75, 25, 0, 50, 80, 25, 0, 25, 0, 25, 30, 25, ~
## $ BUSINESSES   <dbl> 13, 13, 67, 67, 67, 0, 0, 47, 0, 47, 33, 0, 0, 33, 0, 33,~
```

```
## $ MOVEMENTS    <dbl> 80, 40, 100, 40, 20, 0, 0, 20, 60, 60, 0, 0, 0, 40, 0, 80~
## $ GLOBAL_INDEX <dbl> 41, 28, 62, 68, 44, 0, 8, 48, 31, 39, 26, 17, 21, 59, 15,~
```

```
## This is the Google Search trend data set.

glimpse(google_trends)
```

```
## Rows: 3,146
## Columns: 9
## $ week       <chr> "1/3/2021", "1/31/2021", "9/5/2021", "4/14/2019", "11/17/20~
## $ depression <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2,~
## $ ocd        <chr> "0", "3", "4", "2", "0", "6", "6", "0", "5", "13", "14", "8~
## $ anxiety    <dbl> 7, 10, 29, 4, 0, 21, 16, 0, 9, 4, 14, 20, 17, 5, 6, 3, 8, 4~
## $ insomnia   <chr> "14", "10", "4", "0", "100", "3", "6", "6", "0", "4", "4", ~
## $ therapy    <dbl> 32, 36, 37, 17, 33, 27, 48, 55, 32, 31, 49, 36, 34, 15, 14,~
## $ nation     <chr> "Ukraine", "Ukraine", "Ukraine", "Iran", "Iran", "Argentina~
## $ COUNTRY    <chr> "Ukraine", "Ukraine", "Ukraine", "Iran", "Iran", "Argentina~
## $ DATE_START <chr> "1/3/2021", "1/31/2021", "9/5/2021", "4/14/2019", "11/17/20~
```

```
## This is the data set joining restrictions and search trends by both date and country.

glimpse(new_set)
```

```
## Rows: 1,740
## Columns: 17
## $ COUNTRY      <chr> "Argentina", "Argentina", "Argentina", "Argentina", "Arge~
## $ DATE_START   <chr> "1/10/2021", "1/12/2020", "1/17/2021", "1/19/2020", "1/24~
## $ ISO3         <chr> "ARG", "ARG", "ARG", "ARG", "ARG", "ARG", "ARG", "ARG", "~
## $ WHO_REGION   <chr> "AMRO", "AMRO", "AMRO", "AMRO", "AMRO", "AMRO", "AMRO", "~
## $ MASKS        <dbl> 100, 0, 100, 0, 100, 0, 100, 100, 0, 100, 100, 100, 100, ~
## $ TRAVEL       <dbl> 100, 0, 100, 0, 100, 0, 100, 100, 0, 100, 100, 100, 100, ~
## $ GATHERINGS   <dbl> 25, 0, 25, 0, 25, 0, 25, 25, 0, 25, 25, 25, 25, 25, 25, 2~
## $ SCHOOLS      <dbl> 25, 0, 25, 0, 25, 0, 25, 25, 0, 25, 25, 25, 25, 25, 25, 2~
## $ BUSINESSES   <dbl> 13, 0, 13, 0, 13, 0, 13, 13, 0, 13, 13, 13, 13, 13, 13, 1~
## $ MOVEMENTS    <dbl> 100, 0, 60, 0, 60, 0, 60, 60, 0, 60, 60, 60, 60, 60, 60, ~
## $ GLOBAL_INDEX <dbl> 60, 0, 54, 0, 54, 0, 54, 54, 0, 54, 54, 54, 54, 54, 54, 5~
## $ depression   <dbl> 13, 17, 10, 10, 20, 17, 9, 10, 21, 12, 9, 6, 29, 12, 24, ~
## $ ocd          <chr> "3", "7", "0", "0", "3", "3", "3", "3", "3", "3", "3", "6~
## $ anxiety      <dbl> 16, 34, 10, 17, 3, 14, 13, 23, 24, 12, 12, 24, 6, 9, 12, ~
## $ insomnia     <chr> "6", "3", "3", "3", "10", "3", "0", "13", "0", "3", "12",~
## $ therapy      <dbl> 45, 64, 35, 56, 42, 48, 35, 52, 52, 43, 27, 42, 26, 41, 4~
## $ date         <date> 2021-01-10, 2020-01-12, 2021-01-17, 2020-01-19, 2021-01-~
```

Data Analysis Plan

In order to conduct our analysis, we will examine various combinations of mental health search term popularity as explained by the index of public health measure severity. Mental health search term popularity will be used as an indicator for what mental illness is most prevalent during a given time frame, and will be analyzed alongside what restriction was most intense for the same time frame. This relationship will be explored across various countries in order to account for the differences in public health measures that each government enacted throughout the course of the pandemic, as well as determine global averages for mental health search term popularity for each public health measure.

In order to examine if there are some mental health issues that are impacted more strongly by certain public health measures, an analysis of variance (ANOVA) will be conducted on the data. This test will allow

for the comparison across multiple means, where each mean is the global averages of mental health search term popularity for each public health measure. We hypothesize that there will be a statistically significant difference among mental illness prevalence as a result of certain public health measures. In order to reject the null hypothesis that there is no difference among certain public health measures disproportionately affecting certain mental illnesses, our p-value for this ANOVA would need to be less that 0.05 for a confidence level of 95%.

```
summary_stat <-
  data.frame(depression_mean = aggregate(depression ~ COUNTRY, new_set, mean),
             depression_sd = aggregate(depression~COUNTRY, new_set, sd),
             anxiety_mean = aggregate( anxiety ~ COUNTRY, new_set, sd ),
             anxiety_sd = aggregate(anxiety~COUNTRY, new_set, sd),
             ocd_mean = aggregate( ocd ~ COUNTRY, new_set, mean ),
             ocd_sd = aggregate(ocd~COUNTRY, new_set, sd),
             insomnia_mean = aggregate( insomnia ~ COUNTRY, new_set, mean ),
             insomnia_sd = aggregate(insomnia~COUNTRY, new_set, sd),
             therapy_mean = aggregate( therapy ~ COUNTRY, new_set, mean ),
             therapy_sd = aggregate(therapy~COUNTRY, new_set, sd)
)
```

```
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.rm =
## na.rm): NAs introduced by coercion

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA

## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
```

```
summary_stat <- subset(summary_stat, select = -c(
                                                  depression_sd.COUNTRY,
                                                 anxiety_mean.COUNTRY,
                                              anxiety_sd.COUNTRY,ocd_mean.COUNTRY,
                        ocd_sd.COUNTRY,insomnia_mean.COUNTRY,insomnia_sd.COUNTRY,
                        therapy_mean.COUNTRY,therapy_sd.COUNTRY) )


(rename(summary_stat, COUNTRY=depression_mean.COUNTRY))
```

```
##           COUNTRY depression_mean.depression depression_sd.depression
## 1       Argentina                  13.873563                 6.671363
## 2          Brazil                  20.551724                 4.487084
## 3        Colombia                  14.241379                 8.594164
## 4          France                  73.574713                 8.769175
## 5         Germany                  74.195402                 7.597183
## 6           India                  17.126437                 9.449750
## 7       Indonesia                  15.563218                 3.244678
## 8           Italy                   8.793103                 2.724914
## 9        Malaysia                  33.206897                 5.951623
## 10         Mexico                  23.597701                 7.244019
## 11    New Zealand                  34.586207                 7.240955
## 12    Philippines                  46.620690                12.136090
## 13         Poland                  21.839080                 6.381065
## 14   South Africa                  56.678161                10.543922
## 15          Spain                  22.057471                 6.625104
## 16         Turkey                   3.873563                 1.327742
## 17        Ukraine                  12.482759                 6.484081
```

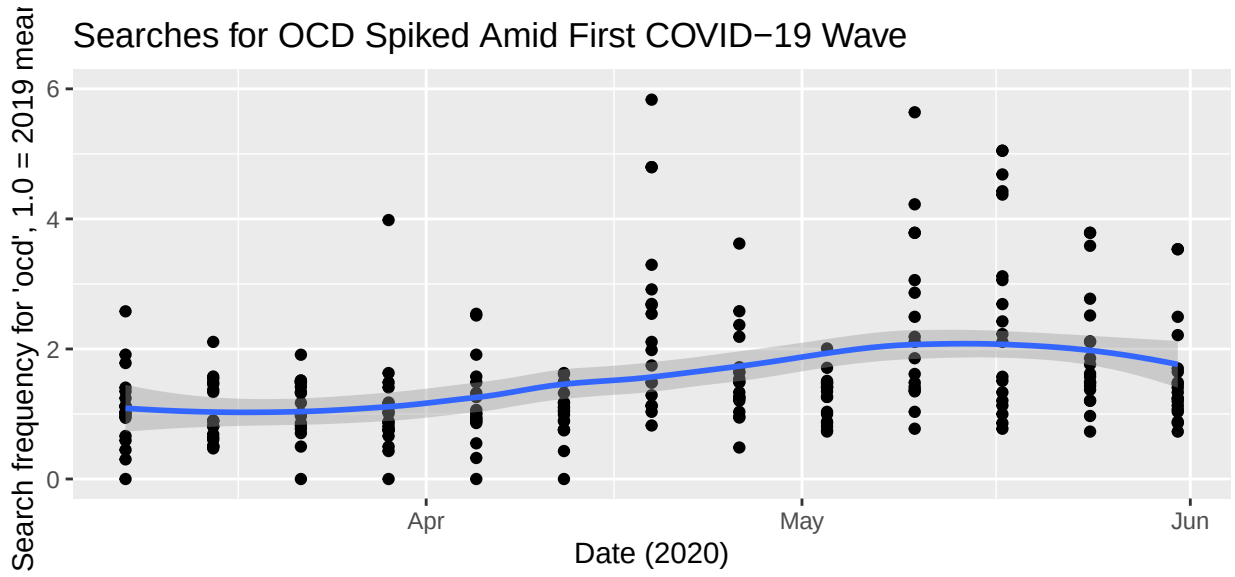
```
## 18 United Kingdom                  47.367816                5.335521
## 19  United States                  37.459770                5.410594
##    anxiety_mean.anxiety anxiety_sd.anxiety ocd_mean.ocd ocd_sd.ocd
## 1              6.375533           6.375533           NA  8.2149029
## 2              4.881022           4.881022           NA  3.9957208
## 3              7.439464           7.439464           NA  7.7526069
## 4              1.612650           1.612650           NA  2.0879486
## 5              2.440525           2.440525           NA  2.1172526
## 6              2.482389           2.482389           NA  0.7175205
## 7              8.550587           8.550587           NA  5.6939685
## 8              2.490938           2.490938           NA  3.3926292
## 9              8.565018           8.565018           NA  3.3542912
## 10             7.406208           7.406208           NA  9.2088106
## 11             8.411258           8.411258           NA  2.8528933
## 12            10.201433          10.201433           NA  2.1307188
## 13             8.195160           8.195160           NA  7.0696122
## 14             7.782443           7.782443           NA  2.4595090
## 15             6.233194           6.233194           NA  9.2667557
## 16             1.371123           1.371123           NA         NA
## 17             6.628049           6.628049           NA  3.6876873
## 18             6.659552           6.659552           NA  1.5624743
## 19             2.586592           2.586592           NA  0.5236967
##    insomnia_mean.insomnia insomnia_sd.insomnia therapy_mean.therapy
## 1                      NA             4.216888              44.20690
## 2                      NA             3.743836              70.54023
## 3                      NA             4.970751              47.57471
## 4                      NA             1.434855              21.48276
## 5                      NA             1.252210              16.39080
## 6                      NA             0.911039              20.96552
## 7                      NA            13.287257              45.98851
## 8                      NA            10.621658              33.52874
## 9                      NA             3.578874              41.90805
## 10                     NA             3.824155              56.54023
## 11                     NA             2.621442              53.10345
## 12                     NA             4.316966              37.63218
## 13                     NA             8.514199              67.72414
## 14                     NA             4.455968              50.96552
## 15                     NA             4.743910              60.68966
## 16                     NA             1.306943              23.54023
## 17                     NA             4.865911              48.95402
## 18                     NA             1.654459              68.80460
## 19                     NA             1.168279              85.06897
##    therapy_sd.therapy
## 1           14.436821
## 2           11.229948
## 3           18.865270
## 4            4.331770
## 5            2.446432
## 6            4.621608
## 7            6.975862
## 8            6.313284
## 9            6.714497
## 10          13.107609
## 11          13.790825
```

```
## 12           4.516417
## 13          13.461996
## 14           8.106053
## 15          13.484108
## 16          17.100119
## 17          12.866857
## 18           8.750866
## 19           9.360721
```

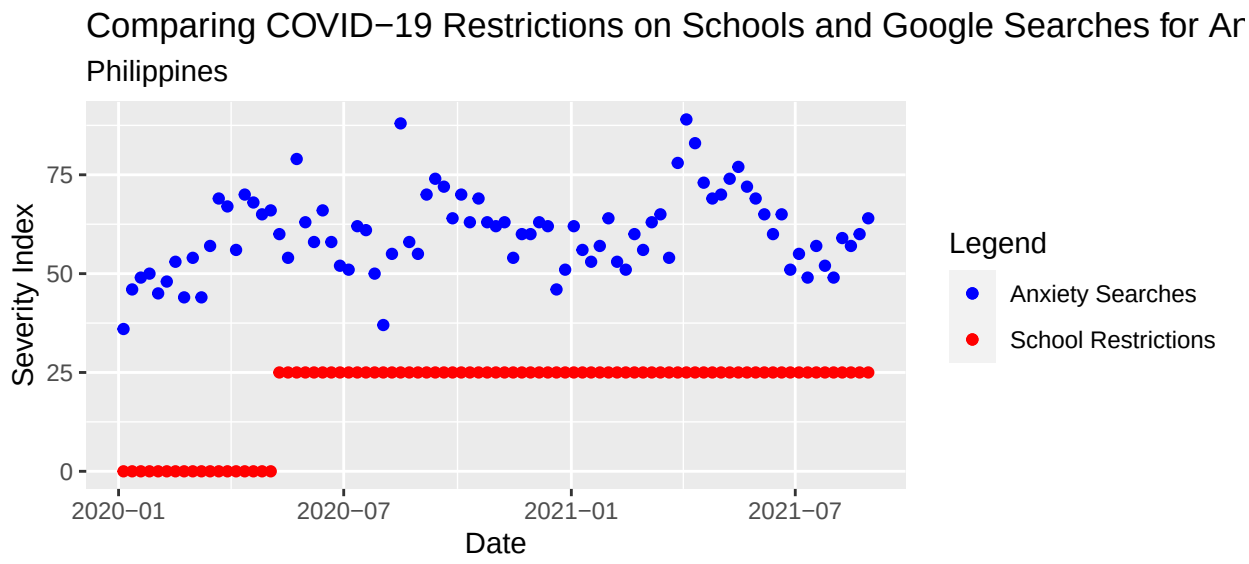
```
natl_avg <- google_trends %>%
  mutate(tidytime = mdy(DATE_START), ocd2 = as.integer(ocd), insomnia2 = as.integer(insomnia)) %>%
  filter(tidytime < mdy("1/1/2020")) %>%
  group_by(COUNTRY) %>%
  summarise(Ndepression = mean(depression), Nocd2 = mean(ocd2), Nanxiety = mean(anxiety), Ninsomnia2 = m

lim_set <- merge(new_set, natl_avg, by=c("COUNTRY")) %>% mutate(relative_ocd = as.integer(ocd) / Nocd2)

ggplot(data = lim_set, mapping = aes(x = mdy(DATE_START), y = relative_ocd)) + geom_point() + geom_smoo
```




```
new_set %>%
  filter(COUNTRY == "Philippines") %>%
  ggplot() + geom_point(mapping = aes(y = SCHOOLS, x = date, color = "Red")) + geom_point(mapping = aes
```



Linear Regression

```
#new_set <- new_set %>%
#  filter(MASKS != 0, TRAVEL != 0, GATHERINGS != 0, SCHOOL != 0, BUSINESSES != 0, MOVEMENTS != 0)
```

```
US_data <- new_set %>%
  filter(COUNTRY == "United States")


US_ocd_reg <- linear_reg() %>%
  set_engine("lm") %>%
  fit(ocd ~ MASKS + TRAVEL + GATHERINGS + SCHOOLS + BUSINESSES + MOVEMENTS, data = US_data)

US_ocd_reg %>%
  tidy()
```

```
## # A tibble: 7 x 5
##   term         estimate std.error statistic  p.value
##   <chr>           <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)  5.74       0.213      27.0   6.20e-42
## 2 MASKS       -0.00277    0.00249    -1.11  2.69e- 1
## 3 TRAVEL       0.00164    0.00565     0.290 7.73e- 1
## 4 GATHERINGS  -0.00269    0.00798    -0.336 7.37e- 1
## 5 SCHOOLS      0.0158     0.00868     1.82  7.29e- 2
## 6 BUSINESSES  -0.00267    0.00306    -0.872 3.86e- 1
## 7 MOVEMENTS    0.000607   0.00235     0.258 7.97e- 1
```

```
US_dep_reg <- linear_reg() %>%
  set_engine("lm") %>%
  fit(depression ~ MASKS + TRAVEL + GATHERINGS + SCHOOLS + BUSINESSES + MOVEMENTS, data = US_data)

US_dep_reg %>%
  tidy()
```

```
## # A tibble: 7 x 5
```

```
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept) 44.1        1.43      30.8   3.65e-46
## 2 MASKS       -0.0136     0.0167    -0.811 4.20e- 1
## 3 TRAVEL       0.0800     0.0380     2.10  3.85e- 2
## 4 GATHERINGS  -0.271      0.0538    -5.04  2.87e- 6
## 5 SCHOOLS     -0.283      0.0584    -4.85  6.06e- 6
## 6 BUSINESSES  -0.00677    0.0206    -0.329 7.43e- 1
## 7 MOVEMENTS   -0.00404    0.0158    -0.255 7.99e- 1
```

References

Centers for Disease Control and Prevention. (2021). Coping with stress. Centers for Disease Control and Prevention. Retrieved October 11, 2021, from (https://www.cdc.gov/mentalhealth/stress-coping/cope-with-stress/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fdaily-life-coping%2Fmanaging-stress-anxiety.html).

Our World in Data. Cumulative confirmed covid-19 cases and deaths. Our World in Data. (n.d.). Retrieved October 11, 2021, from (https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid-19).