Final Report: The COVID-19 Pandemic, Public Health Restrictions, and Mental Health

due November 16, 2021 by 11:59 PM

Katherine Beltz, Matt Mohn, Melannie Nimocks

11/16/2021

Load Packages

Introduction and Data

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. 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 restriction type. Our project will make use of two data sets and merge the data sets based on country in order to ensure there are sufficient relationships to explore in the data. The first data set is of the frequency of different search topics related to mental health from January of 2019 through September of 2021 for a variety of countries (Google Trends, 2021). The data was collected from Google Trends and records the relative popularity of that search topic for any given week in a given country. This data set will be used to gauge how concern with certain mental health topics, including anxiety, depression, panic attacks, obsessive compulsive disorder, and insomnia, has changed over the course of the pandemic. The second data set tracks implementation of COVID-19 mitigation efforts in different countries (Zosimov, 2021). This data set was originally 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, such as masks, gatherings, businesses, schools, or travel. This data has daily measurements from January 2020 through September of 2020. The union of this data set with the previous will enable the exploration of any possible correlation between public health and social measures and mental health as measured by relative mental health search topic frequency.

Load Data

#Exploratory Data Analysis

Through a series of data tidying and wrangling steps, the data has been joined on both country and date. In order to do so, column names and data formats had to first be modified in order to ensure the column titles match across the data sets. Any duplicate columns that resulted from the merge were removed. 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. Shown below is a glimpse of the resulting data set.

```
## Rows: 522
## Columns: 17
## $ COUNTRY      <chr> "Brazil", "Brazil", "Brazil", "Brazil", "Brazil", "Brazil~
## $ DATE_START   <chr> "1/10/2021", "1/12/2020", "1/17/2021", "1/19/2020", "1/24~
## $ ISO3         <chr> "BRA", "BRA", "BRA", "BRA", "BRA", "BRA", "BRA", "BRA", "~
## $ WHO_REGION   <chr> "AMRO", "AMRO", "AMRO", "AMRO", "AMRO", "AMRO", "AMRO", "~
## $ MASKS        <dbl> 47, 0, 47, 0, 47, 0, 47, 47, 0, 47, 47, 47, 47, 47, 47, 4~
## $ TRAVEL       <dbl> 100, 0, 100, 0, 100, 0, 100, 100, 0, 100, 17, 17, 100, 17~
## $ GATHERINGS   <dbl> 30, 0, 30, 0, 30, 0, 30, 30, 0, 5, 5, 5, 5, 5, 5, 5, 30, ~
## $ SCHOOLS      <dbl> 25, 0, 25, 0, 25, 0, 25, 30, 0, 25, 25, 25, 25, 25, 25, 2~
## $ BUSINESSES   <dbl> 47, 0, 47, 0, 80, 0, 47, 80, 0, 80, 80, 80, 80, 80, 47, 4~
## $ MOVEMENTS    <dbl> 80, 0, 20, 0, 20, 0, 80, 20, 0, 80, 20, 20, 80, 80, 20, 8~
## $ GLOBAL_INDEX <dbl> 55, 0, 45, 0, 50, 0, 55, 51, 0, 56, 32, 32, 56, 42, 41, 5~
## $ depression   <dbl> 36, 41, 37, 46, 35, 40, 37, 33, 39, 40, 39, 37, 39, 39, 3~
## $ ocd          <dbl> 74, 66, 65, 59, 71, 61, 77, 77, 57, 51, 87, 66, 60, 70, 6~
## $ anxiety      <dbl> 87, 67, 86, 69, 86, 73, 86, 87, 67, 70, 74, 82, 71, 74, 7~
## $ insomnia     <dbl> 79, 87, 69, 74, 83, 83, 77, 68, 71, 55, 56, 53, 62, 64, 5~
## $ therapy      <dbl> 78, 83, 81, 85, 80, 84, 76, 74, 80, 86, 91, 88, 86, 80, 7~
## $ date         <date> 2021-01-10, 2020-01-12, 2021-01-17, 2020-01-19, 2021-01-~
```

Preliminary data analysis included summary statistics, namely, the mean and standard deviation for each mental health search term relative frequency grouped by country. For the purposes of making sure analysis was feasible, five countries of interest were selected for this report: Brazil, India, Mexico, The United States, and New Zealand. Such countries were chosen in order to provide a wide range of COVID conditions, as many of these countries approached COVID mitigation differently. This provided a baseline understanding of what mental illnesses may be more or less prevalent in each country, disregarding any public health social measures, over the period of study from January 2020 through September 2020.

As seen in the Summary Statistic Tables below, anxiety as a search topic has the lowest mean frequency over the period of study. Therapy appears to have the highest mean frequency over the period of study. Interestingly, insomnia and depression as search topics have the greatest range of mean frequencies between each country. For insomnia, there is a range of about 53 points, with the minimum score from Italy and the maximum score from the United States. The United States also holds the highest score for depression search frequency, with a low from India, resulting in a range of 59.

It is important to note that the standard deviations for many of these measurements is fairly high. This is to be expected since the measurements are taken over a time period of rapidly changing social distancing policies and COVID severity across the globe. Further statistical analyses in this report determine how much this affects our understanding of this data.

Table 1: Summary Statistics for Depression and Anxiety

nation	depression_mean	depression_sd	anxiety_mean	anxiety_sd
Brazil	37.126	3.797	7.409	7.409
India	16.207	9.499	4.659	4.659
Italy	35.747	10.774	8.776	8.776
Mexico	55.851	8.648	8.439	8.439
New Zealand	57.356	13.813	11.478	11.478

nation	depression_mean	depression_sd	anxiety_mean	anxiety_sd
United States	75.759	6.457	4.451	4.451

Table 2: Summary Statistics for Therapy, Insomnia and OCD

nation	therapy_mean	therapy_sd	insomnia_mean	insomnia_sd	ocd_mean	ocd_sd
Brazil	83.402	5.645	70.851	11.968	69.563	9.717
India	77.874	7.219	57.816	13.727	67.506	10.580
Italy	56.644	10.005	26.069	9.629	56.667	11.998
Mexico	86.218	5.903	50.540	13.311	34.782	10.474
New Zealand	72.126	9.721	43.023	16.347	35.529	15.378
United States	84.345	4.742	79.609	7.727	71.184	5.858

As the next step in exploratory data analysis, the graph below was created to examine mental health search terms as a function of time, regardless of any COVID restrictions or of any specific country. This graph was made for all search terms, but it was found that the graph for OCD search topic frequency provided the most insight into how individuals' mental state shifted during the pandemic. By using geom_smooth, the line on the plot demonstrates a slight increase in frequency of searching for OCD during the first wave of COVID. From a qualitative standpoint, this increase would align with the notion that individuals became concerned with signs of OCD as much of the population became highly obsessive over cleanliness, which is often understood as OCD in the general population despite this definition being slightly inaccurate. It could also be due to individuals demonstrating signs of a compulsive need to follow certain rituals in order to feel safe from COVID.

Statistical analysis provided further in the report explores the significance, if any, of this increase.

OCD Searches Spiked In First COVID Wave
100 = 2019 mean
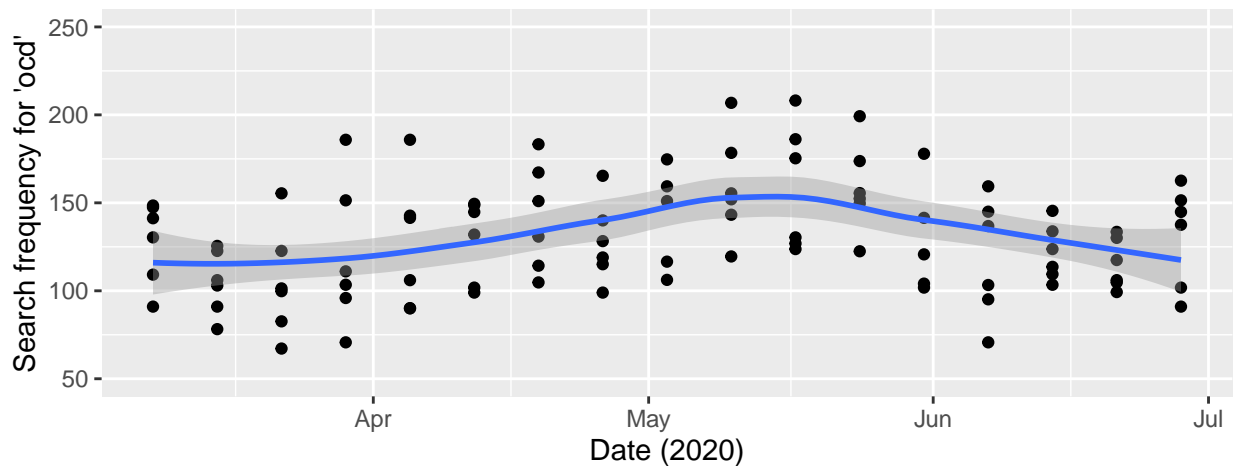
Methodology and Results

In order to conduct analysis, mental health search topic relative popularity was examined as explained by the index of public health measure severity. Mental health search topic popularity was used as an indicator for mental illness prevalence during a given time frame, and was analyzed alongside restriction intensity for the same time frame. This relationship was 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 to determine global averages for mental health search term popularity for each public health measure.

(???)

Five countries in our data set were chosen as focus: Brazil, India, Mexico, The United Sates, and New Zealand. We first focused on quantifying the relationship between one mental health burden, as quantified by search topic frequency, and different types of restrictions. This was examined using multiple linear regression models where the response variable was relative mental illness search topic frequency conditioned on all of the different COVID restrictions. This was done for all search terms for each of the five countries. Once all the regression models were created, any model with 3 or more significant explanatory variable estimates (p-value <0.05) was selected for further evaluation.

Table 3: Nation: USA, Search: Therapy

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	93.4381	1.8105	51.6102	0.0000	89.8352	97.0411
MASKS	0.0706	0.0212	3.3349	0.0013	0.0285	0.1128
TRAVEL	0.0109	0.0481	0.2269	0.8211	-0.0848	0.1067
GATHERINGS	0.1804	0.0680	2.6523	0.0096	0.0450	0.3157
SCHOOLS	-0.2416	0.0739	-3.2691	0.0016	-0.3887	-0.0945
BUSINESSES	0.0994	0.0260	3.8175	0.0003	0.0476	0.1512
MOVEMENTS	0.0014	0.0200	0.0693	0.9449	-0.0385	0.0413

Table 4: Nation: USA, Search: Anxiety

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	101.6822	1.8172	55.9569	0.0000	98.0660	105.2985
MASKS	0.0706	0.0213	3.3207	0.0014	0.0283	0.1129
TRAVEL	0.0017	0.0483	0.0354	0.9718	-0.0944	0.0978
GATHERINGS	0.1011	0.0683	1.4803	0.1427	-0.0348	0.2369
SCHOOLS	-0.1738	0.0742	-2.3427	0.0216	-0.3214	-0.0262
BUSINESSES	0.0693	0.0261	2.6509	0.0097	0.0173	0.1213
MOVEMENTS	0.0428	0.0201	2.1284	0.0364	0.0028	0.0828

Table 5: Nation: Brazil, Search: OCD

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	101.6838	3.5204	28.8838	0.0000	94.6779	108.6897
MASKS	0.1732	0.0755	2.2947	0.0244	0.0230	0.3234
TRAVEL	-0.0516	0.0534	-0.9648	0.3376	-0.1579	0.0548
GATHERINGS	0.1819	0.1557	1.1681	0.2462	-0.1280	0.4917
SCHOOLS	-0.6105	0.1860	-3.2825	0.0015	-0.9806	-0.2404
BUSINESSES	0.0771	0.0624	1.2347	0.2206	-0.0472	0.2014
MOVEMENTS	0.1870	0.0694	2.6958	0.0086	0.0490	0.3251

Table 6: Nation: Brazil, Search: Therapy

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	108.1074	1.6076	67.2479	0.0000	104.9082	111.3066
MASKS	0.1889	0.0345	5.4795	0.0000	0.1203	0.2575
TRAVEL	-0.0760	0.0244	-3.1160	0.0025	-0.1246	-0.0275

term	estimate	std.error	statistic	p.value	conf.low	conf.high
GATHERINGS	-0.1496	0.0711	-2.1048	0.0385	-0.2911	-0.0082
SCHOOLS	0.1444	0.0849	1.7004	0.0929	-0.0246	0.3134
BUSINESSES	-0.0862	0.0285	-3.0224	0.0034	-0.1429	-0.0294
MOVEMENTS	-0.0322	0.0317	-1.0177	0.3119	-0.0953	0.0308

Table 7: Nation: Mexico, Search: Depression

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	103.9686	4.3731	23.7745	0.0000	95.2658	112.6714
MASKS	0.5112	0.1493	3.4236	0.0010	0.2141	0.8084
TRAVEL	-0.0224	0.0492	-0.4557	0.6499	-0.1204	0.0755
GATHERINGS	0.4850	0.1531	3.1672	0.0022	0.1803	0.7898
SCHOOLS	-1.0561	0.3770	-2.8010	0.0064	-1.8064	-0.3057
BUSINESSES	0.1051	0.0679	1.5477	0.1257	-0.0300	0.2402
MOVEMENTS	-0.6979	0.2007	-3.4766	0.0008	-1.0974	-0.2984

Table 8: Nation: Mexico, Search: Insomnia

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	104.6186	7.6423	13.6894	0.0000	89.4100	119.8273
MASKS	-1.7312	0.2610	-6.6339	0.0000	-2.2505	-1.2118
TRAVEL	0.1912	0.0860	2.2221	0.0291	0.0200	0.3624
GATHERINGS	-0.4785	0.2676	-1.7881	0.0776	-1.0111	0.0541
SCHOOLS	3.9493	0.6589	5.9939	0.0000	2.6381	5.2606
BUSINESSES	0.0240	0.1187	0.2021	0.8403	-0.2122	0.2601
MOVEMENTS	1.6124	0.3508	4.5964	0.0000	0.9143	2.3106

Table 9: Nation: India, Search: Insomnia

| term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	118.1261	7.2126	16.3777	0.0000	103.7726	132.4797
MASKS	0.5406	0.2569	2.1043	0.0385	0.0293	1.0519
TRAVEL	-0.2096	0.1877	-1.1171	0.2673	-0.5831	0.1638
GATHERINGS	0.6567	0.2389	2.7490	0.0074	0.1813	1.1321
SCHOOLS	-2.2563	0.5041	-4.4758	0.0000	-3.2595	-1.2531
BUSINESSES	0.4375	0.2233	1.9595	0.0535	-0.0068	0.8819
MOVEMENTS	0.0890	0.1076	0.8265	0.4110	-0.1252	0.3031

The residual plots below were created to understand any concern over the linear regression models applied to the data. Residual plots were only created for models that showed significant findings, as these are the data sets the rest of the interpretation will primarily focus on.

None of the residual plots have a clear pattern. Although some of the residual plots are slightly concentrated on one side of the plot, there is still no discernible pattern in the variation, indicating that the models are appropriate for the data.

Next to each residual plot is the corresponding scatter plot with the regression line fitted to the data. The

significance of these relationships will be further explored numerically from the results of regression analysis. (???)
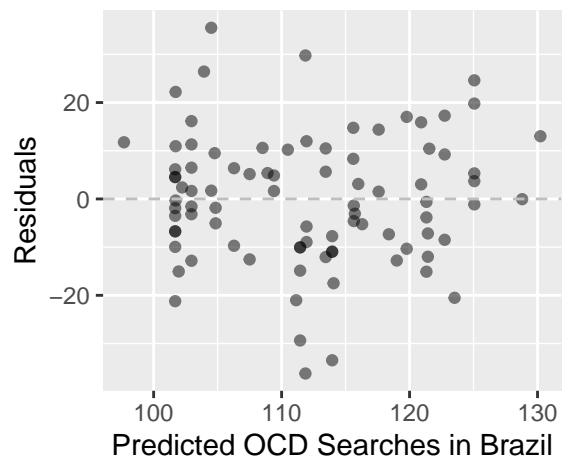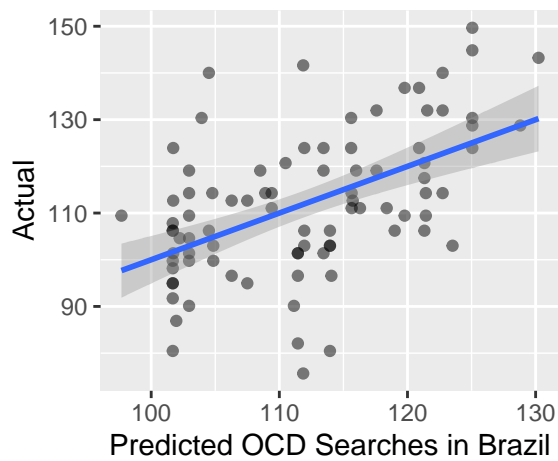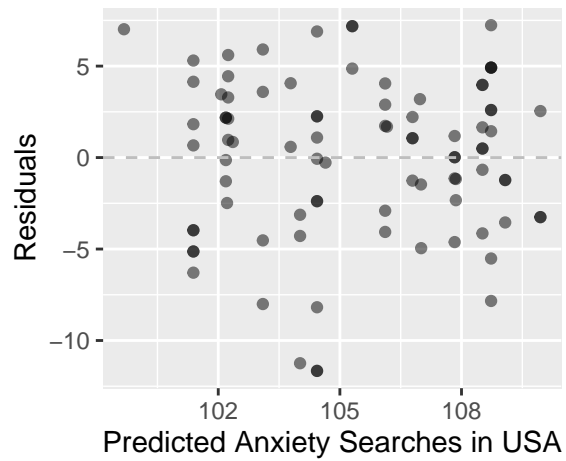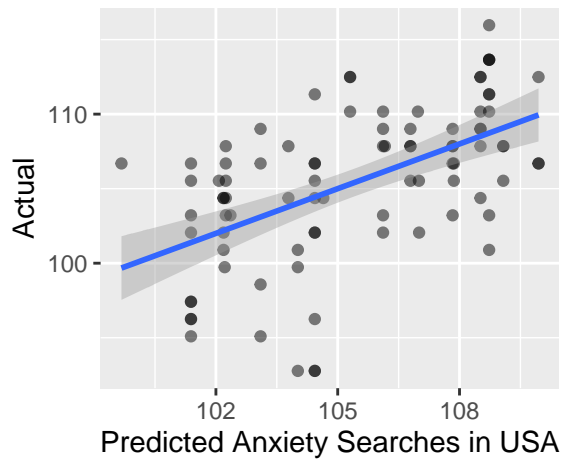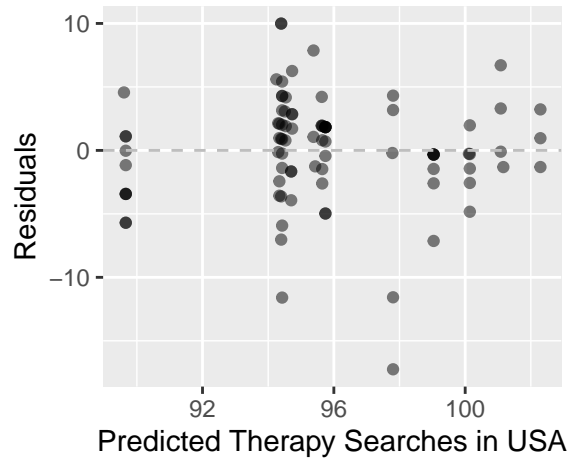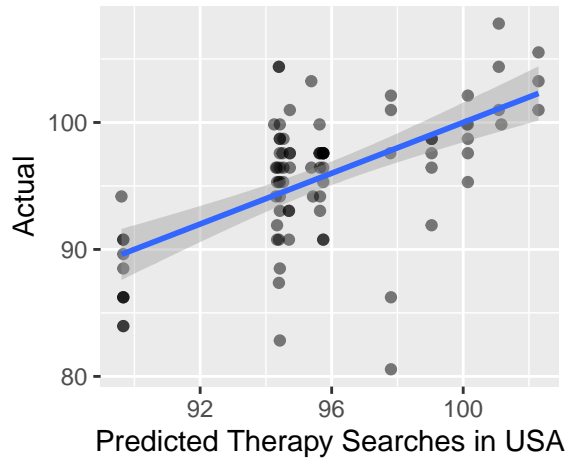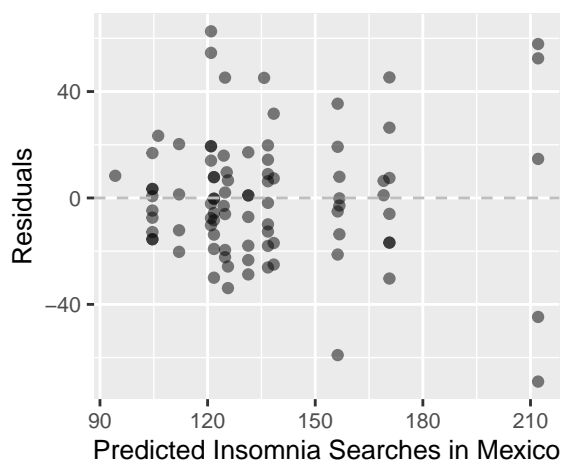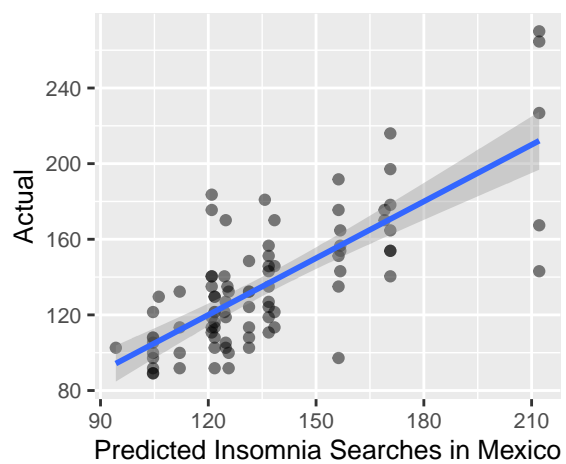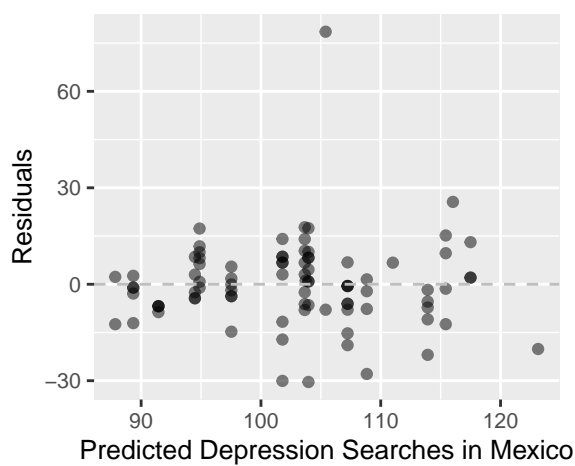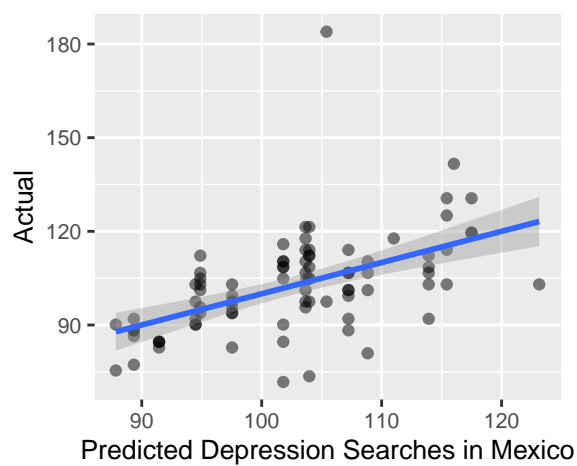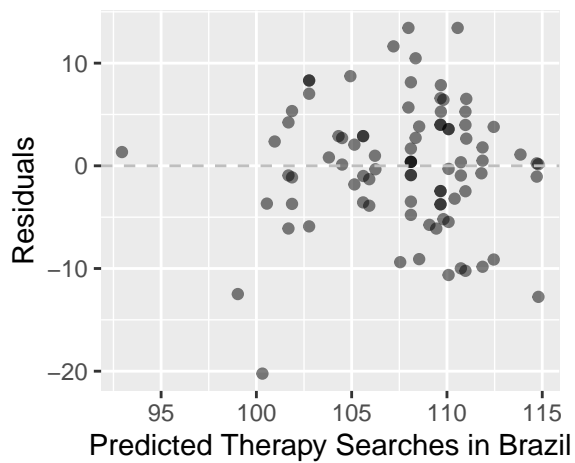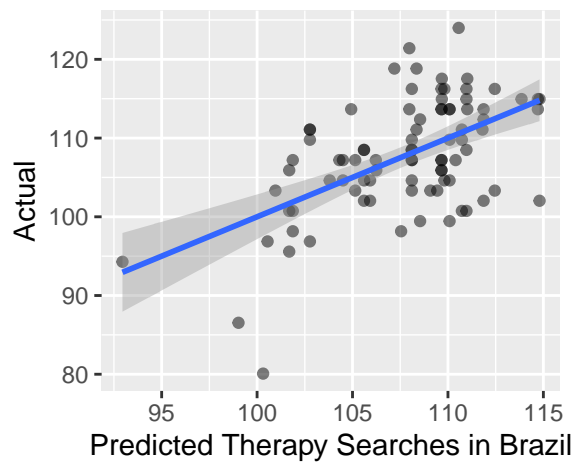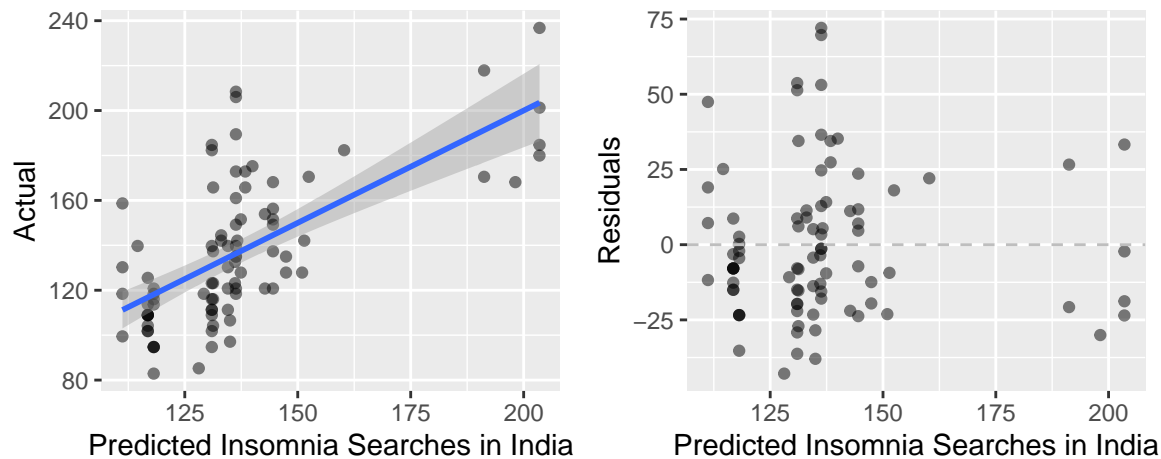
The confidence intervals and statistical significance of each variable were then examined. Multiple linear regression models were chosen for analysis in order to evaluate how the different restrictions, measured through a global index, affect the relative search popularity of mental health topic. This form of analysis was used for the purpose of determining which restrictions have a significant effect in determining the relative search popularity for a specific mental health term. 5 multiple regression models–one for each mental health search topic being evaluated for this study–were fit for each of the five countries chosen for analysis. Mental health search topic relative frequency was set as the response variable and the different restrictions were set as explanatory variables. As previously explained, for the purposes of concision, only models with 3 or more significant explanatory variable estimates (p-value <0.05) was selected for further evaluation. These were evaluated as follows.

(???) overall, need to include language about rejecting null hypothesis in both intro and in conclusions/analyses.

USA-THERAPY: This regression model evaluates the effects of different restrictions on searches for therapy in the United States. Restrictions with significant p-values (< 0.05) in this model are masks, gatherings, schools, and businesses. Masking restrictions have an estimate of 0.0706 (95% CI = (0.0285,0.1128)), gatherings restrictions have an estimate of 0.1804 (95% CI = (0.0450,0.3157)), and business restrictions have an estimate of 0.0994 (95% CI = (0.0476, 0.1512)). The positive estimates from these data demonstrate that an increase in restrictions in masking, gatherings and businesses leads to an increase in therapy searches in the US. School restrictions have an estimate of -0.2416 (95% CI= (-0.3887, -0.0945)). This negative estimate demonstrates that an increase in school restrictions leads to a decrease in therapy searches in the US. (make sure this is explained in conclusion) (??? i don't think we can use this specific language).

USA-ANXIETY: This regression model evaluates the effects of different restrictions on searches for anxiety in the United States. Restrictions with significant p-values (< 0.05) in this model are masks, schools, and businesses. Masking restrictions have an estimate of 0.0706 (95% CI = (0.0283,0.1129) and business restrictions have an estimate of 0.0693 (95% CI = (0.0173, 0.1213)). These positive estimates demonstrate that an increase in restrictions in masking and businesses leads to an increase in anxiety searches in the US. School restrictions have an estimate of -0.1738 (95% CI= (-0.3214, -0.0262)). This demonstrates that an increase in school restrictions leads to a decrease in anxiety searches in the US.

BRAZIL-OCD: This regression model evaluates the effects of different restrictions on searches for OCD in Brazil. Restrictions with significant p-values (< 0.05) in this model are masks, schools, and movements. Masking restrictions have an estimate of 0.1732 (95% CI = (0.0230,0.3234)), movements restrictions have an estimate of 0.1870 (95% CI = (0.0490,0.3251)). These positive estimates demonstrate that an increase in restrictions in masking and movements leads to an increase in therapy searches in the US. School restrictions have an estimate of -0.6105 (95% CI= (-0.9806, -0.2404)). This negative estimate demonstrates that an increase in school restrictions leads to a decrease in OCD searches in Brazil.

BRAZIL-THERAPY: This regression model evaluates the effects of different restrictions on searches for therapy in Brazil. Restrictions with significant p-values (< 0.05) in this model are masks, gatherings, travel,

and businesses. Masking restrictions have an estimate of 0.1889 (95% CI = (0.1203,0.2575), gatherings restrictions have an estimate of -0.1496 (95% CI = -0.2911,-0.0082), travel restrictions have an estimate of -0.0760 (95% CI = (-0.1246, -0.0275)), and business restrictions have an estimate of -0.0862 (95% CI = (-0.1429, -0.0294)). These data demonstrate that an increase in restrictions in masking leads to an increase in therapy searches in Brazil, while an increase in restrictions in gatherings, travel, and businesses leads to a decrease in therapy searches in Brazil.

MEXICO-DEPRESSION: This regression model evaluates the effects of different restrictions on searches for depression in Mexico. Restrictions with significant p-values (< 0.05) in this model are masks, gatherings, schools, and movements. Masking restrictions have an estimate of 0.5112 (95% CI = (0.2141,0.8084)), movement restrictions have an estimate of -0.6979 (95% CI = (-1.094,0.3251)), gathering restrictions have an estimate of 0.4850 (95% CI = (0.1803,0.7898)), and school restrictions have an estimate of -1.0561 (95% CI = (-1.8064,-0.3057)). These data demonstrate that in Mexico an increase in restrictions in masking and gatherings leads to an increase in depression searches, while an increase in restrictions in schools and movements leads to a decrease in depression searches.

MEXICO-INSOMNIA: This regression model evaluates the effects of different restrictions on searches for insomnia in Mexico. Restrictions with significant p-values (< 0.05) in this model are masks, travel, schools, and movements. Masking restrictions have an estimate of -1.7312 (95% CI = (-2.2505,-1.2118)), movement restrictions have an estimate of 1.6124 (95% CI = 0.9143,2.3106), travel restrictions have an estimate of 0.1912 (95% CI = (0.0200,0.3624)), and school restrictions have an estimate of 3.9493 (95% CI = (-2.6381,5.2606)). These data demonstrate that in Mexico an increase in restrictions in movement, travel, and school leads to an increase in insomnia searches, while an increase in masking leads to a decrease in insomnia searches.

INDIA-INSOMNIA: This regression model evaluates the effects of different restrictions on searches for insomnia in India. Restrictions with significant p-values (< 0.05) in this model are masks, gatherings, and school. Masking restrictions have an estimate of 0.5406 (95% CI = (0.0293,1.0519)), gathering restrictions have an estimate of 0.6567 (95% CI = 0.1813,1.1321)), and school restrictions have an estimate of -2.2563 (95% CI = (-3.2595,-1.2531)). These data demonstrate that in Mexico an increase in restrictions in masking and gatherings leads to an increase in insomnia searches, while an increase in school restrictions leads to a decrease in insomnia searches.

These regression models helped us better understand the relationship between certain restrictions' severity and the number of searches for a mental health term. Out of the 20 models that had data for maskings, 12 showed significant p-values (< 0.05) for mask restrictions, and 11 of those showed a positive estimate. Out of the 25 models that had data for travel, 6 showed significant p-values, and 3 of those showed a positive estimate. Out of the 25 models that had data for gatherings, 7 showed significant p-values, and 3 of those showed a positive estimate. Out of the 25 models that had data for schools, 13 showed significant p-values, and 4 of those showed a positive estimate. Out of the 25 models that had data for businesses, 6 showed significant p-values, and 3 of those showed a positive estimate. Out of the 25 models that had data for movements, 4 showed significant p-values, and 3 of those showed a positive estimate. The only restrictions where a majority of the regression models showed significance are masking and school. Furthermore, 11 out of 12 significant masking values were positive, demonstrating that an increase in masking generally demonstrated an increase in searches for mental health terms. For the school restrictions, 9 out of 13 significant school values were negative, demonstrating that an increase in school restrictions generally demonstrated a decrease in searches for mental health terms. (??? needs to be more specific. talk about what this means for restrictions and mental health overall, not just saying it has positive estimates)

There are various limitations to our analysis that must be considered in order to fully understand how the conclusions apply. Some of these limitations come from considering how data was collected, particularly from Google Trends. Firstly, Google Trends measures searches relative to all other searches, and is not a nominal measure of the amount of searches being done. This adds noise to our model and may therefore skew our multiple regression models. For example, search terms for depression could have doubled but the relative score of its frequency would not change significantly if other search terms increased as well. Additionally, though the point of this report was to examine mental health concern as a function of search terms, they may not be fully indicative of actual mental health, especially considering that the search terms had to be

broadened into topics in order to account for language differences. In terms of the restrictions data set, the indices are compressed into one score for an entire country, which does not account for local differences.

In general, modeling never accounts for every single factor that could contribute to explaining a response variable. There may be some factors in our model that are impacted by confounding variables or extra noise in the data set, as explained above. This is why they are models and not static measurements. Though this report has outlined interesting possibilities in terms of the relationship between mental health concern and COVID restrictions, it is important to understand the limitations in both data collection and modeling for a relationship as complex as this one.

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.

Google Trends. Google Trends, Google. (2021). Retrieved November 5, 2021, from https://trends.google.com/trends/?geo=US.

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.

Zosimov, R. (2021, September). COVID-19: Public Health Social Measures, Version 1. Retrieved October 11, 2021, from https://www.kaggle.com/hijest/covid19-public-health -social-measures/metadata.