Seminar Paper

Global Analysis of Mental Health Impacts during COVID-19: Findings from UMD Global CTIS

Christian Hobelsberger[1], Max Lang[2] and Elisa Shima[3]

[1]LMU Munich, Website: https://christian.hobelsberger-muc.de, Github: christian-hobelsberger.
[2]LMU Munich, E-mail: M.Lang@campus.lmu.de Github: MaxMLang.
[3]LMU Munich, E-mail: E.Shima@campus.lmu.de Github: elisashima.

Keywords: COVID-19, CTIS, R Shiny, Mental Health

Abstract
The impact of the COVID-19 pandemic on public mental health has been a widely discussed topic, particularly regarding the effectiveness of the various countermeasures implemented to address the crisis. Previous work suggests that the COVID-19 pandemic led to increased levels of mental distress worldwide. In contrast to other study studies that focused on a specific country or continent, we analysed the impact on the global mental health. We use the *UMD Global CTIS* survey, policy response information as well as Google trends data provided from *Our World in Data*. As part of our exploratory data analysis we built a Shiny App displaying important mental health metrics from the data. Our results show a trend in mental health, with older people being more resilient to the effects of the pandemic than young adults. Respondents with a lower educational level were more likely to report mental health problems than people with a higher educational level. Further, we found that worries about the financial situation, food security as well as concerns of becoming ill seem to negatively impact the mental health of respondents globally.

1. Introduction

The COVID-19 pandemic, caused by a novel coronavirus, infected more than 670 million people worldwide and caused millions of deaths by 30 January 2023 ("Covid-19 Dashboard", 2023). Governments imposed a series of non-pharmaceutical interventions to stop the spread of the virus. While maintaining low levels of daily new infections and deaths seemed to be the primary goal of interventions such as lockdowns and school closures, the mental health effects of these measures were mostly ambiguous. This is why we delve further into exploring mental health during the COVID-19 pandemic. Research found that 13.6% of respondents reported serious psychological distress in April 2020 compared with 2.9% of 2018 National Health Interview Survey respondents (McGinty et al., 2020). Another study compared the responses gathered between March to April 2020 to those obtained from the period of 2017 to 2018 and found that the incidence of depressive symptoms reported in the 2020 survey was over three times greater than that reported in the 2017-2018 data (Ettman et al., 2020). Godleski et al., 2022 conducted a analysis on the impact of relevant social vulnerabilities and subjective factors on mental health outcomes such as depression, anxiety, and isolation, with a specific focus on the influence of pregnancy on mental health during the pandemic. Our research utilizes the data obtained from the Global COVID-19 Trends and Impact Survey, which will be thoroughly discussed in the following section 2.1. Further, we will describe our methodology, analysis, and results. Specifically, we will outline the data sources and visualization techniques we used to develop our shiny application, describe our approach to analyzing and modeling the data, and present the most significant findings from our analysis efforts. This paper aims to provide a comprehensive and detailed analysis, with the goal of advancing our understanding of mental health during the COVID-19 pandemic.

2. Materials and visualization

2.1. Global COVID-19 Trends and Impact Survey (CTIS)

"The University of Maryland Social Data Science Center Global COVID-19 Trends and Impact Survey, in partnership with Facebook" (UMD Global CTIS)[1] serves as the data source for this study. During the survey period, a representative sample of Facebook users aged 18 and over were invited to report on symptoms, social distancing behavior, mental health issues, and financial constraints on a daily basis. To reduce non-response and coverage bias, Facebook provided weights. The United States were not included in the UMD Global CTIS (Kreuter et al., 2020).
The data from The Global COVID-19 Trends and Impact Survey are divided into micro- and macrodata. While the macrodata is publicly accessible, access to the microdata is restricted and can only be obtained by researchers who have been granted access. The microdata reflects daily individual survey responses, which in macrodata are aggregated on at least a weekly basis. The Global COVID-19 Trends and Impact Survey Open Data API and the **CTIS R** package allowed us to access and import the survey data for our analysis (Fan et al., 2020; Haensch & Xiong, 2021). Further information on the data, such as variable names and specifications, can be found in the sections 3.1 (microdata) and 3.2 (macrodata).

2.2. Our world in data - Policy responses

In our analysis we also used the policy response data set from Our World in Data (Mathieu et al., 2020), which simplifies government policy responses to the pandemic on a scale from 0 (no measures) to 3 (required) with 3 being the highest level of restrictions. As many countries have a federal or sub-regional governing structure on e.g passing school closures the authors mention:

> [. . .] that there may be sub-national or regional differences in policies on school closures. The policy categories shown may not apply at all sub-national levels. A country is coded as 'required closures' if at least some sub-national regions have required closures. (Mathieu et al., 2020)

This procedure was also used for the stay home requirements. Both of the data sets were used in our analysis.
Moreover, we used Google trends data provided by Our World in Data. Mathieu et al., 2020 used the data provided by Google from applications such as Google Maps. The data set measures the number of visitors to categorized locations (e.g. transit stations, pharmacies, workplaces) and compares the change relative to a baseline day, a median value over a five week period from January 3rd to February 6th 2020. This is important as the data does not take seasonal variation into account. For example the number of visitors to parks could be higher during summer months. Thus, rather than being entirely explained by changes brought on by the pandemic, the data may indicate some variations in seasonal changes. The raw data from Google was transformed by Mathieu et al., 2020 into seven-day rolling averages to reduce the amount of day-to-day variability.

2.3. Shiny Application

As part of our exploratory data analysis we built a R Shiny application for visualization of the data (Chang et al., 2023). The app contains observations from both the macro- and microdata and is structured into three geographical analysis levels (Kreuter et al., 2020):

- 2.3.1 Global

[1]Link to UMD Global CTIS: https://covidmap.umd.edu/

Figure 1. *Lading page of the CTIS mental health Shiny application.*

- 2.3.2 Continental
- 2.3.3 Country

In addition to the three geographical levels, there is a *More* tab containing additional information such as precise variable definitions and acknowledgements.
All levels are implemented as drop down menus with an *analysis* section. The "Global" and "Continental" menu also contain a *maps* section offering a spatial analysis of the specific data level. The "Country" menu does not contain a spatial analysis because we observed major differences between the starting e.g. anxiety values of different continents throughout the survey. Thus, the values of the respective mental health variables should always be set in context with neighbouring countries on the continent. Each level is described in more detail below.

2.3.1. Global menu

As mentioned previously the "Global" menu contains an *analysis* and *maps* section, while the latter functions also as the landing page of the Shiny application as seen in Figure 1.

Map section

The map section uses a sidebar in which the users can change the map characteristics. The inputs are the respective mental health variable and the survey date. On the right side, the map appears. The map is plotted with the **tmap** package, which offers Shiny output functions such that the created interactive maps can easily be rendered on the web app (Tennekes, 2018). Therefore, the user can freely zoom into the map, change the base layer by clicking on the layer symbol and hover over countries to see their name. Moreover, the user can get country-specific information by clicking on a country as depicted in Figure 2. The pop-up shows the name of the country, the value of the selected health variable as well as the level of stay-home requirements and school closures from the policy response data in the country. (Figure 2)

Analysis section

The analysis section uses a similar layout to the map section with a sidebar containing a co-variable input. In contrast to the "Global - Map" section, the "Global - Analysis" section uses microdata, such that the variable labels differ. In addition, there is a checkbox deciding whether or not the variable should be plotted over time, by default this box is unticked, resulting in a bar plot as shown in Figure 3. The bar plots as well as the line plots (triggered by the ticked checkbox) are plotted with the **plotly** package, which offers interactive plots, such that the user can hover over the bars or lines to see the exact value of the variable (Sievert, 2020).

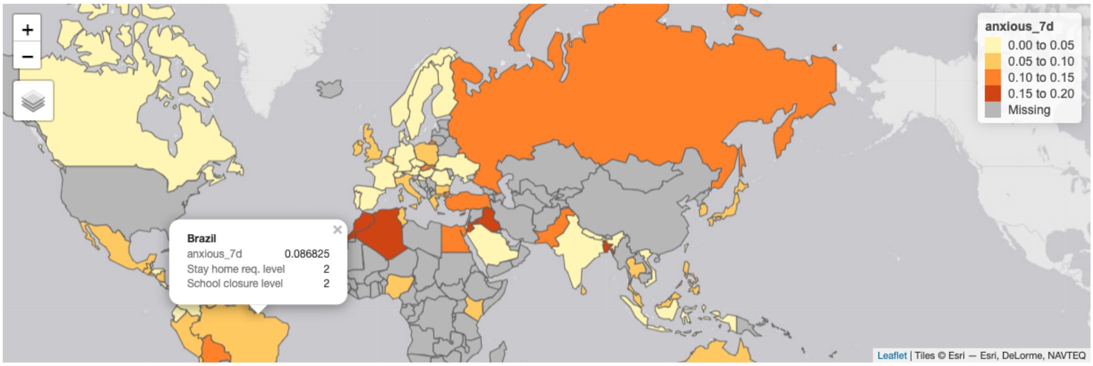

Figure 2. *Zoomed in version of the map output showing the country pop-up for Brazil.*

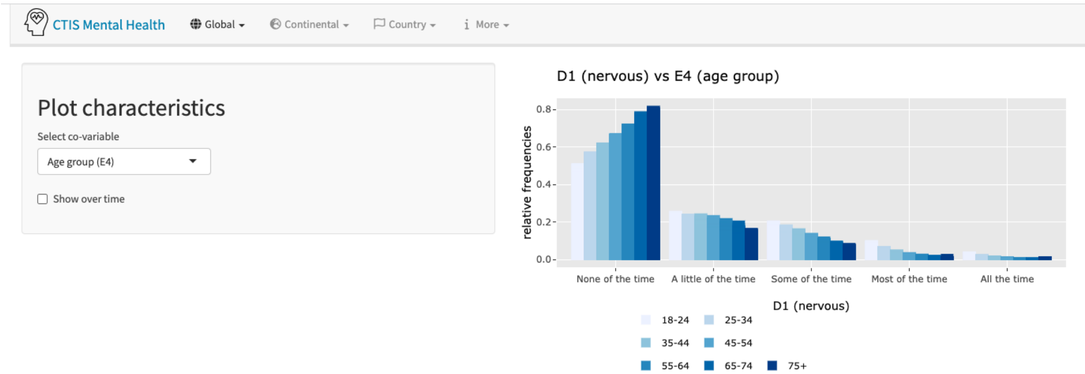

Figure 3. *The interface of the "Global - Analysis" section.*

2.3.2. Continental menu

Map section

The user interface of the continental map section is structured like the global map section, with an additional input for the specific continent. The continents were mapped to the country codes using the **countrycode** package and plotted with **tmap**. We want to emphasize that the sequential colour scale is free because of two characteristics of the data (Arel-Bundock et al., 2018).

First, the initial and ongoing values of the mental health variables from different continents differ drastically. For example, the aggregated anxiety values (`anxious_7d`) of African countries are always above European countries. Second, the data indicates variations in seasonal changes, policy responses or other events (e.g. wars, famines etc.) over time. We therefore suggest to adjust the scale for each continent and date instead of having a fixed scale for all continents and dates.

Analysis section

In the Analysis section the user can select a mental health variable out of the macrodata set and one or multiple continents. The line plot shows the continental average(s) of the variable of interest. (see Figure 4)

2.3.3. Country menu

Analysis section

As noted previously the country menu does only contain an "Analysis" section. The layout is similar to other "Analysis" sections. The user can select a mental health variable out of the macrodata set and

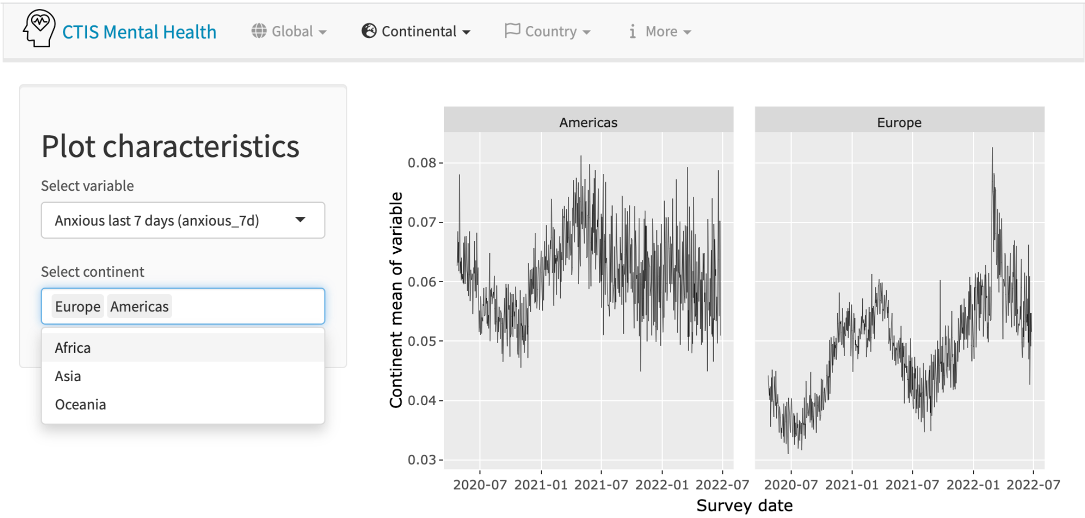

Figure 4. *Interactive line plot in the "Continent - Analysis" section showing the continental averages of Americas and Europe over time.*

one country. Countries with smaller populations often have missing data over time or a small number of responses such that the line plot sometimes might be unstable.

2.3.4. More menu

The "More" menu offers additional info about CTIS in the "About" section as well as a short "Data dictionary" of the used variables.

3. Methods and Results

As previously noted in section 2.1, the data collected by the UMD Global CTIS is divided into micro- and macrodata (Kreuter et al., 2020). In our analysis, we will further examine both data levels, and for this purpose, we will divide the following methods and results sections respectively. Since our focus is on the mental health variables, we will deal with the variables `anxious` and `depressed`, which we will introduce in more detail in the following sections. Since these two variables are highly correlated (Person correlation coefficient 0.8064), we will mainly focus on the variable `anxious`.

3.1. Microdata

In this section we will take a further look into the microdata collected by the Global COVID-19 Trends and Impact Survey which contains the individual responses (Kreuter et al., 2020). The central focus of the analysis is the variable generated by the question *During the past 7 days, how often have you felt so nervous that nothing could calm you down?*, hereafter referred to as `anxious` or `D1`.

3.1.1. Data structure and types

In our study, we analyzed data from 66,613,710 respondents worldwide, collected through the UMD Global CTIS between April 23, 2020, and June 15, 2022 (as seen in Table 1) (Fan et al., 2020; Kreuter et al., 2020)). The responses for the variable `anxious (D1)` were recorded on a Likert scale ranging from "None of the time" to "All of the time". `Age (E4)` was divided into the categories "18-24", "25-34", ..., "65-74", "75+". The gender of the respondents was categorized into: "Male", "Female", "Other", and "Prefer not to answer". Moreover, the educational level (`education (E8)`) ranges from

Table 1. *Summary statistics of categorical model variables (n = 66,613,710).*

Variables	N	%
During the past 7 days, how often did you feel so nervous that nothing could calm you down? (D1)	45318361	NA
None of the time	26878522	59.31
A little of the time	9548836	21.07
Some of the time	6122466	13.51
Most of the time	1972948	4.354
All the time	795589	1.756
What is your age? (E4)	54918689	NA
18-24	6749688	12.29
25-34	12502223	22.76
35-44	11752776	21.4
45-54	10183333	18.54
55-64	8084732	14.72
65-74	4484351	8.165
75+	1161586	2.115
What is your gender? (E3)	54700	532 NA
Male	27220165	49.76
Female	26893157	49.16
Other	135765	0.2482
Prefer not to answer	451445	0.8253
What is the highest level of education that you have completed? (E8)	23515043	NA
No formal schooling	92827	0.3948
Less than primary school	214264	0.9112
Primary school completed	908531	3.864
Secondary school complete	2987295	12.7
High school (or equivalent) completed	6265713	26.65
College/pre-university/ University completed	9797885	41.67
University post-graduate degree completed	3248528	13.81
Which of these best describes the area where you are currently staying? (E2)	53567701	NA
Town	12375973	23.1
City	32729968	61.1
Village or rural area	8461760	15.8

"No formal schooling" to "University post-graduate degree completed" with a total of 7 categories. Lastly, the `place of living (E2)` is simplified to "Town", "City" and "Village or rural area".

In total, 40.69% of respondents said that during the last 7 days they had felt so nervous that nothing could calm them down at least "a little of the time". 13.51% answered "Some of the time", 4.35% answered "Most of the time" and 1.76% answered "All the time" to this question about anxiety. Gender is evenly distributed with 49.76% men and 49.16% women and the majority of people surveyed hold some form of college degree and lives in a city.

3.1.2. Analysis

We will look at the `anxious` variable by itself and in combination with other variables in this subsection. The daily number of observations of the variable `anxious` has a bimodal distribution with an average of 57717.95 (Figure 5). The highest number of observations was recorded in mid-2020 and early 2021, with a decline over time to an average of only 23564.19 in 2022. Information on the number of observations for the variables `age (E4)` and `education (E8)` over time can be located in Appendix A.

When comparing the different levels of `anxious` over time, there is a tendency for respondents to be less nervous, particularly in the first half of the year 2022 (Figure 6). The number of responses to the 'Most of the time' and 'All of the time' levels appear unchanged. The relative frequencies of the 'Some of the time' and 'Most of the time' levels decrease, while the 'None of the time' level increases steadily. Starting from a value of around 0.55, the relative frequency of 'None of the time' has remained strictly above 0.6 since mid-2021, ending at 0.655. A decrease in anxiety from June to the end of 2021 is observed across all age groups.

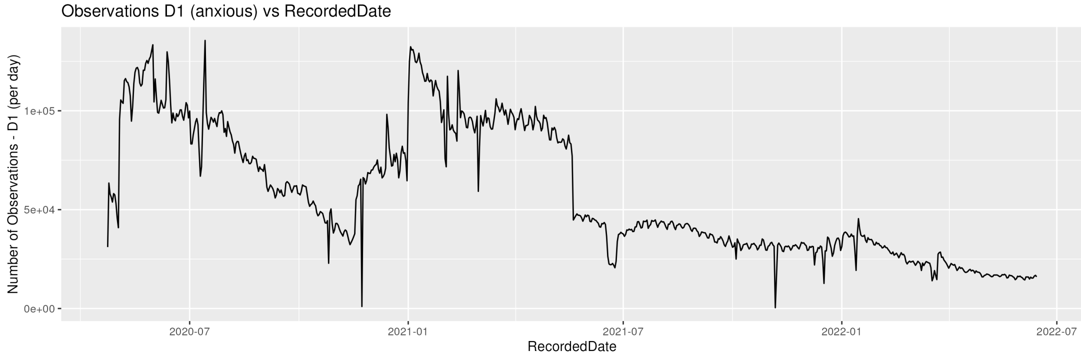

Figure 5. Number of observation of anxious variable over time. *Daily Aggregated Data: Observation Period April 23, 2020 to June 14, 2022.*

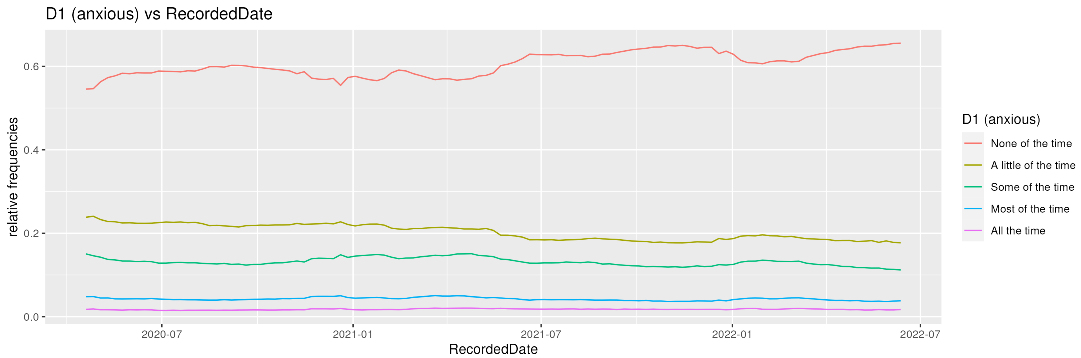

Figure 6. Relative frequencies of anxious variable over time. *Weekly Aggregated Data: Observation period 19 April 2020 to 12 June 2022.*

In the following we will analyse the relationship between `anxious` and variables, such as `age` and `education-level`. Looking at the relative frequencies of `anxious` by age group, there are clear structural differences between the age groups (Figure 7). For the level "None of the time", the higher the age group, the higher the relative frequency. The difference between the age group '18-24' with a value of 0.472 and '75+' with a value of 0.766 can be clearly seen here. This means that the older a person is, the more likely they are to never be nervous. The opposite is true for all other categorisations of the variable `anxious`. This means that the younger age groups tend to be more anxious.
The relative frequencies of the variable `anxious` over time, when separated by age group, reveal varying patterns (Figure 8). The contrasts in relative frequencies of the `anxious` variable over time are evident as mentioned before (Figure 6). While for the age group '75+' the relative frequency for 'None of the time' is strictly above 0.7 and the remaining values are all below 0.2, for the age group '18-24' the categories are much less segregated. For the age groups '18-24', '25-34' and '35-44', there is an increase in anxiety between the beginning and the middle of 2021 (especially for the 'Some of the time' category), which is not seen for other age groups. From mid to end of 2021, as shown in Figure 6, a relaxation is observed for all age groups.

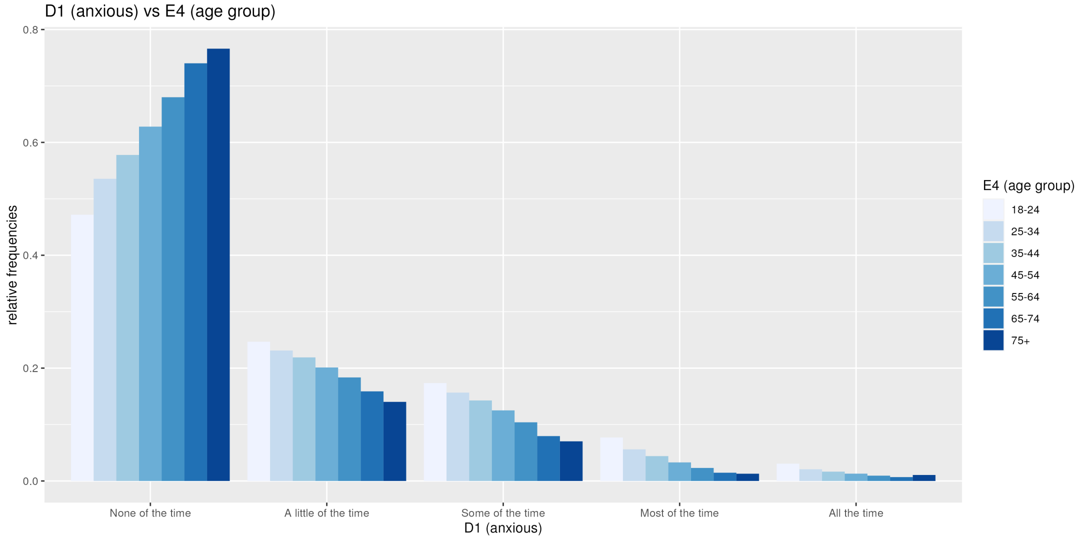

Figure 7. *Relative frequency of anxious variable by age group.*

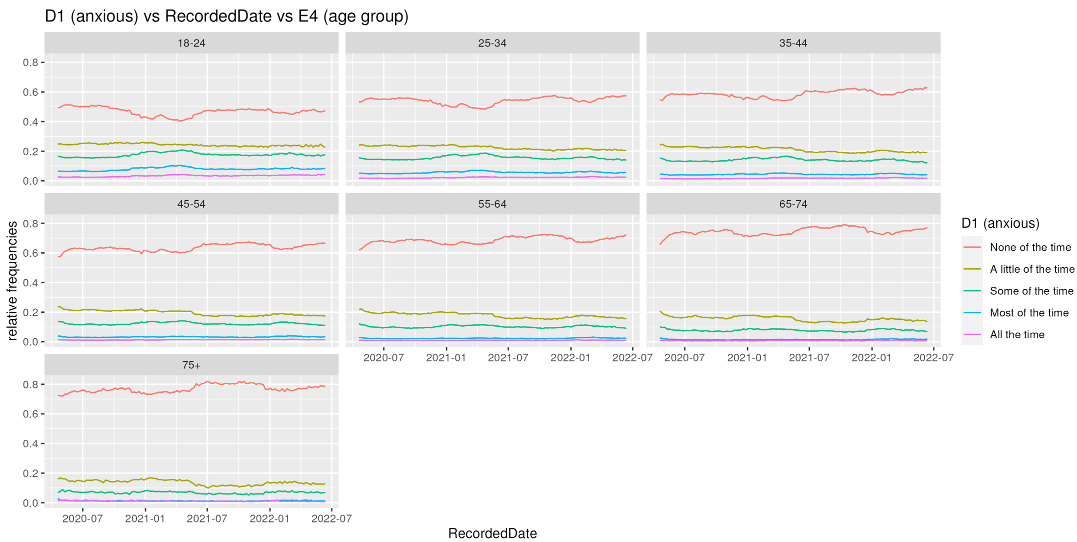

Figure 8. *Relative frequencies of the anxious variable over time by age group.*

Next, we will look at the relationship between the `anxious` and `education-level` variables. Different relative frequencies of the variable `anxious` are observed for different levels of education (Figure 9A). The highest relative frequencies in the 'None of the time' category are 0.66 and 0.653 for people who have completed primary and secondary education respectively. Our findings suggest that, in general, individuals with higher levels of education tend to have higher relative frequencies for the categories 'None of the time', 'A little of the time', and 'Some of the time'. In contrast to the aforementioned findings, a reverse relationship was observed between education level and relative frequencies of `anxious` for the categories 'most of the time' and 'all the time'. Specifically, individuals with higher levels of education tended to have lower relative frequencies in these categories. However, a striking observation is that the relative frequency of respondents with "No formal schooling" for the category "All the time" was found to be 0.0979, which is higher than the relative frequency of 0.0571 for the category "Most of the time" (see Figure 9).

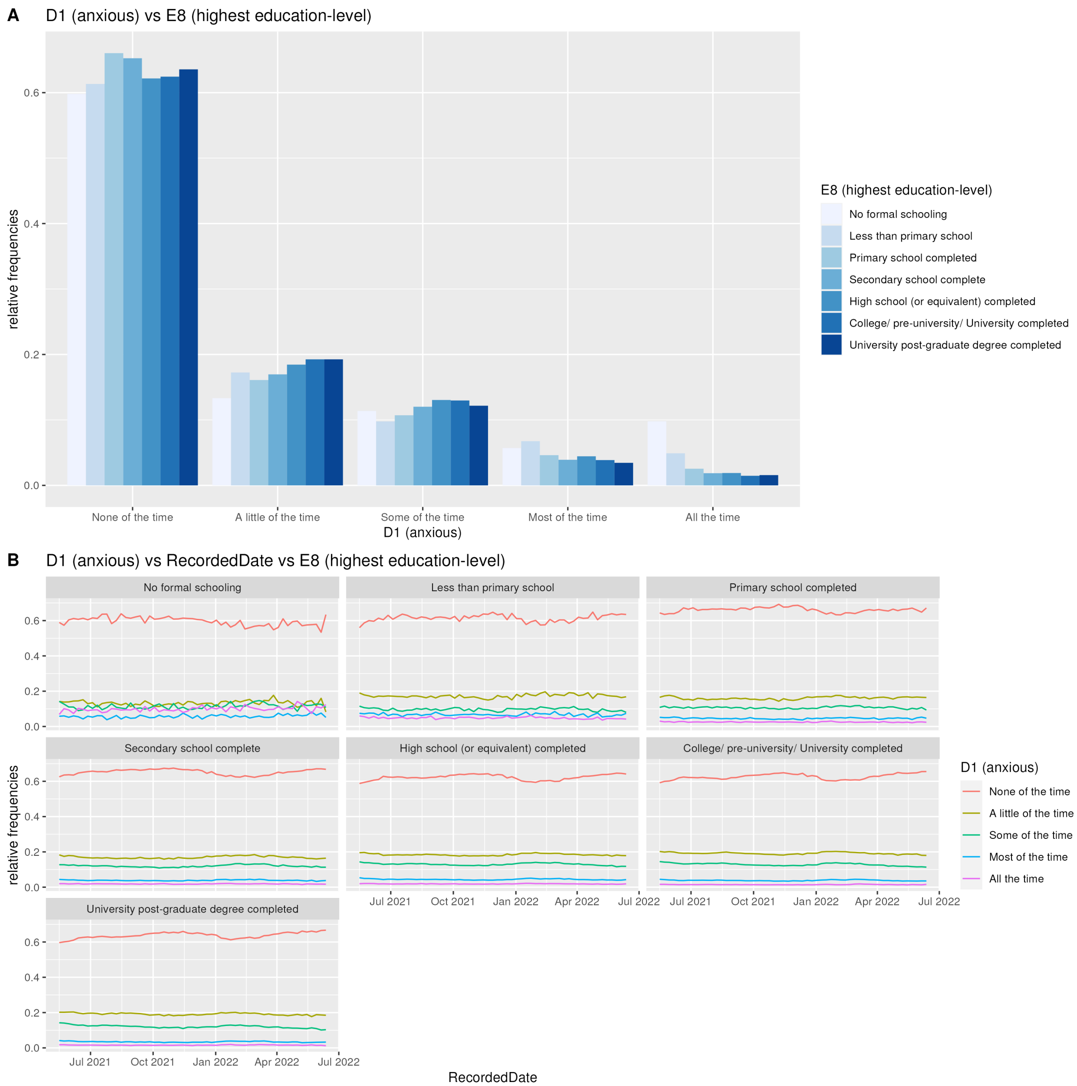

Figure 9. Relationship between the anxious and education variable. *Relative frequency of anxious variable by highest education-level (**A**). Shown over time on weekly aggregated data (**B**): Observation period 19 April 2020 to 12 June 2022.*

3.1.3. Model

Methods

We estimated a logistic regression model to examine the impact of the variables `age (E4)`, `gender (E3)`, `education (E8)` and `place of living (E2)` on our target variable `anxious (D1)`.

The linear predictor is therefore

$$\eta_i = \beta_0 + \beta_1 E2_i + \beta_2 E3_i + \beta_3 E4_i + \beta_4 E8_i$$

In our logistic regression model, the category "45-54" for the variable `age (E4)`, "Male" for the variable `gender (E3)`, "Secondary school complete" for the variable `education (E8)`, and "Town" for the variable `place of living (E2)` were used as the reference categories of the dummy encoding.

3.1.4. Results

Our findings suggest that younger respondents seem to report more anxiety (`anxious (D1)`) than older respondents on average (Figure 10). For instance, a person of the age group 18 - 24 is on average 2.387 times more likely to be anxious c.p (*ceteris paribus*) in comparison to a person of age between 45 - 54. In contrast, persons of age between 65 - 74 are on average 0.459 time less likely to be anxious in comparison to the reference category (age 45 - 54). Furthermore, we found that `gender (E3)` and `education (E8)` both show interesting results.

Our model reveals that female respondents, relative to men, report more anxiety. Respondents that have a lower level of education seem to have poorer mental health, all other characteristics being held constant ($p < 1.16 \times 10^{-5}$)

Hence, the findings suggest three key factors:

- `age (E4)` ($p < 2 \times 10^{-16}$),
- `gender (E3)` ($p < 2 \times 10^{-16}$) and
- `education (E8)` ($p < 1.16 \times 10^{-5}$)

associated with mental health problems.

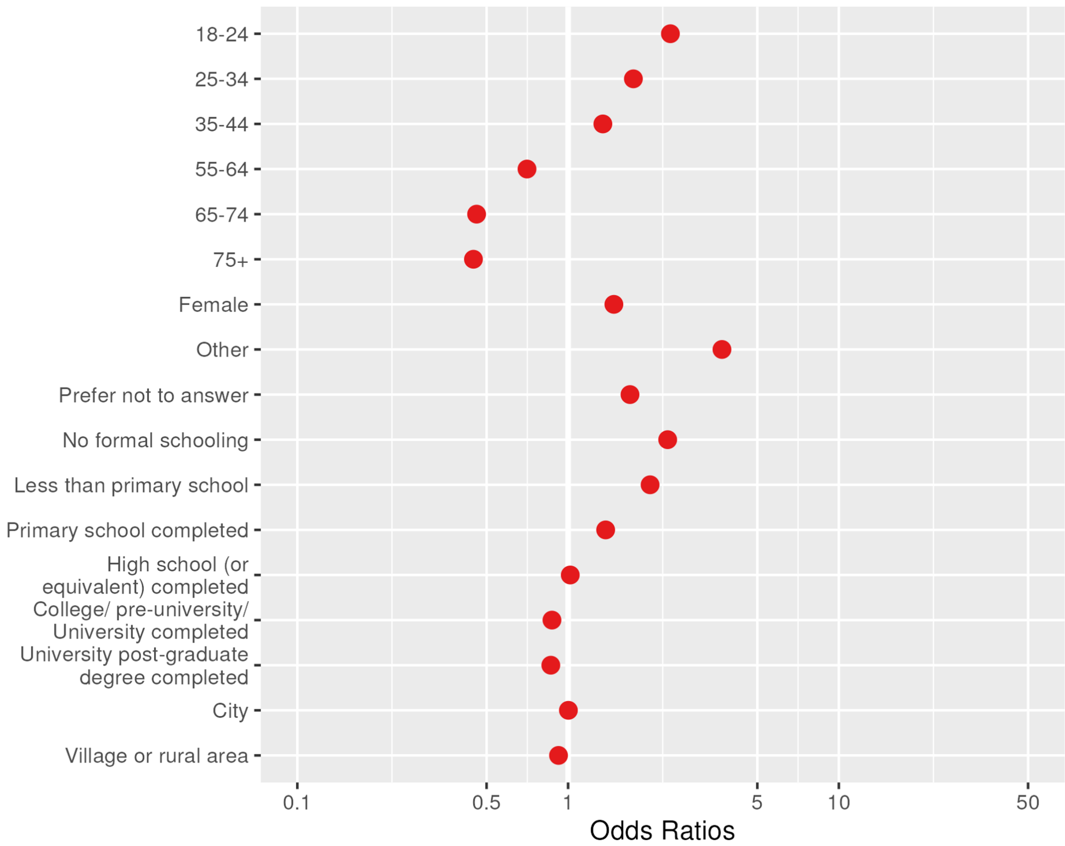

Figure 10. *Odds ratio of logistic regression model.*

3.2. Macrodata

In this section we will take a closer look into the macrodata collected by the Global COVID-19 Trends and Impact Survey which contains the aggregated data from the survey responses and into the Google Trends dataset. We observed that the Google trends data seem to better grasp the impact of the restrictions, rather than the policy data. Several countries showed no difference in mobility trends, besides higher level of restriction. Thus, we find the Google Trends the more suitable variable here. The analysis concentrates mostly on the variable `anxious_7d`, which refers to respondents who reported feeling nervous for most or all of the time over the past 7 days.

3.2.1. Data structure and types

The macrodata used in the analysis comprehend country and region-level statistics published daily via the opendata API and dashboards. The data was collected in 243 countries and territories, including 114 where Facebook provides survey weights. CTIS ran globally continuously from April 6, 2020 until June 26, 2022. Symptoms, behavior, economic, vaccine, mental health, testing, module and others indicator are included in the macrodata. In the analysis we focus on the economic (`finance`, `food_security`) and mental health (`anxious_7d`, `depressed_7d`, `worried_become_ill`) indicators. Note that these indicators are aggregated for each individual on a 7-day basis, i.e., anxiety is calculated as the average of the last 7 days. Respondents were asked if they felt the past 7 days for most or all of the time anxious, depressed or worried that they or someone in their immediate family might become seriously ill from

COVID-19. In the questions that covered the economic topic respondents were asked whether they where very worried or somewhat worried about themselves and their household's finances (`finance`) or whether they were very worried or somewhat worried about having enough to eat next week. In addition to the indicators listed above, in the data set are stored the country names as well, the ISO country code and the date of the survey for the aggregated observations. In the data set used for this analysis

Table 2. *Summary statistics of numerical model variables.*

Variables	Min	Mean	Max
During the past 7 days, how often did you feel so nervous that nothing could calm you down?	0	0.06357	0.2433
How worried are you about your household's finances in the next month?	0.05695	0.4682	0.8698
How worried are you about having enough to eat in the next week?	0.006407	0.262	0.7532
How worried are you that you or someone in your immediate family might become seriously ill from coronavirus (COVID-19)?	0.2075	0.6985	0.9805
Number of visitors to places of retail and recreation	-75	-24.32	107.9
Number of visitors to grocery and pharmacy stores	-56.57	3.471	132.9
Change in duration of time spent at home	-11.57	8.077	34.43
Number of visitors to transit stations	-71	-26.92	55.29
Number of visitors to parks and outdoor spaces	-72.57	-6.549	205.7
Number of visitors to workplaces	-69.43	-20.81	17.43

are included the `anxious_7d` variable, the economic indicator and the mobility trends. The economic indicator `finance` refers to respondents who reported to be feeling very worried or somewhat worried about themselves and their household's finances. The `food_security` variable refers to respondents who reported to be feeling very worried or somewhat worried about having enough to eat in the next week. The `worried_become_ill` variable alludes to respondents who reported feeling very or somewhat worried that they or someone in their immediate family might become seriously ill from COVID-19. The next variables are provided from the Google data set presented in COVID-19 Community Mobility Reports, where it is possible to see how people's movements have changed throughout the pandemic. These mobility trends are measured across six broad categories:

- (1) `retail and recreation`: places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters.
- (2) `grocery and pharmacy stores`: places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.
- (3) `residential`: places of residence.
- (4) `transit stations`: places like public transport hubs such as subway, bus, and train stations.
- (5) `parks`: places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens.
- (6) `workplaces`: places of work.

The `residential` category shows a change in duration, while the other categories measure a change in total visitors. The changes are compared to respective baseline days before the pandemic outbreak.

Baseline days are given as the median value over the five-week period from January 3rd to February 6th 2020.

3.2.2. Analysis

The following analysis is going to focus on the variable `anxious_7d` and its relationship with the other variables. The `anxious_7d` levels during the pandemic were quite different in every country. It turned out that the two countries with the highest `anxious_7d` levels in the past 7 days have been Turkey and Tunisia, while the ones with the lowest `anxious_7d` levels Denmark and Norway. Take into account that the availability of data varied between countries, with some having more data and others having less. Therefore to have a better overview of the `anxious_7d` levels in the different countries, we focused on the countries with observations on at least 600 days. From these, we selected a subset of countries with low, medium, and high levels of anxiety. These countries have very different levels of `anxious_7d` and we can see that in the box plots shown below (Figure 11). The median values for Egypt, South Africa and Turkey are around 0.10, while for Hungary, India and Mexico around 0.05. The two countries with lowest `anxious_7d` levels were Denmark and Norway. Considering that `anxious_7d` reached levels around 0.125, the `anxious_7d` for the first set of country was quite high.

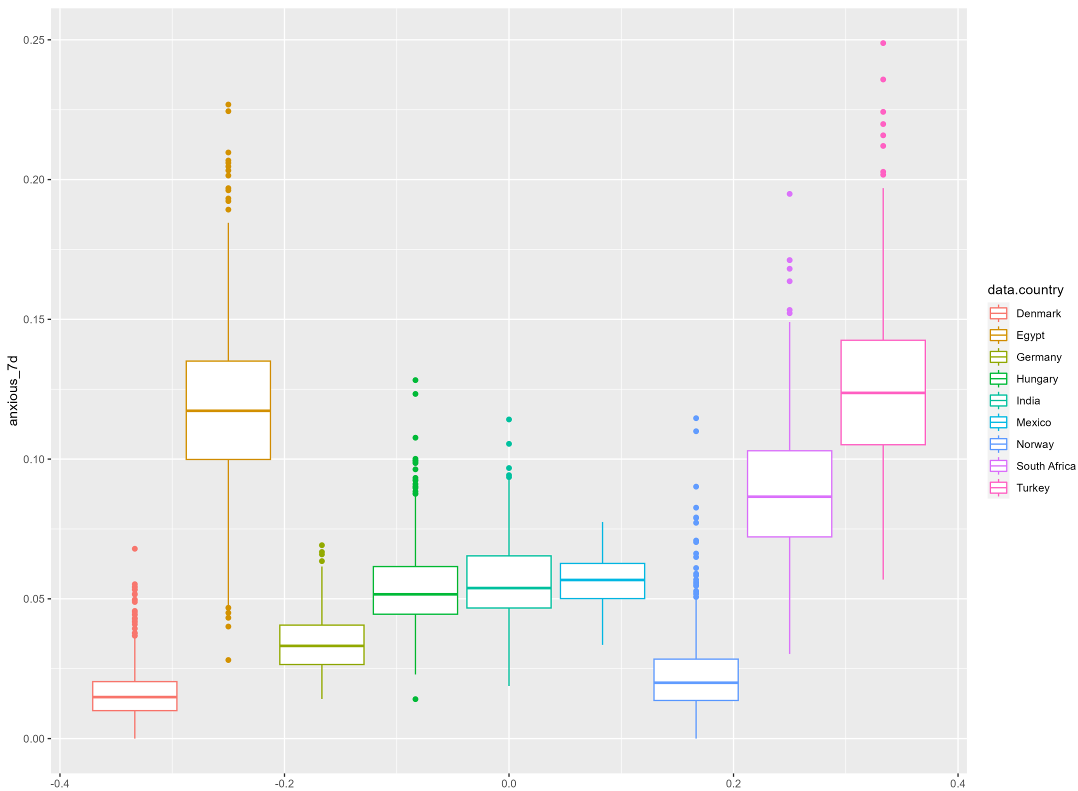

Figure 11. *Distribution of the anxious variable for a selection of countries with observations on at least 600 days.*

3.2.3. Model

Methods

The data has been randomly split into a training data set (80% of the data) and test data set (20% of the data). We tried to fit three different models on the training data:

- (1) Model without regularization
- (2) Model with Ridge-regularization
- (3) Model with Lasso-regularization

Given that the Lasso model turned out to have the best fit for on the test data with respect to the mean squared error (MSE), we are going to further examine it.

We use the Lasso method, that applies a penalty to shrink the parameter estimates, to select the best predictors for `anxious_7d`, using the other variables, which are `finance`, `food_security`, `retail_and_recreation`, `worried_become_ill`, `grocery_and_pharmacy`, `residential`, `transit_stations`, `parks` and `workplaces`. With the help of the **glmnet** package in R we fit the linear regression with Lasso and define the design matrix of the model, X, and the outcome, Y. The first step is to find by cross validation the amount of penalty, λ, that gives the minimum MSE. In the image below (Figure 12) we can see the mean squared error for different lambdas on the test data. The plot has the mean square error on the y-axis and the natural log of λ on the x-axis. Across the top is the number of variables included during different points. With a small λ 9 variables are included. As λ increases, the mean square error increases and variables are dropped, since the penalty for inclusion of variables is weighted more heavily. Two special values along the λ sequence are indicated by the vertical dotted lines. λ_{min} is the value of λ that gives minimum mean cross-validated error, while λ_{1se} is the value of λ that gives the most regularized model such that the cross-validated error is within one standard error of the minimum (Trevor Hastie, 2021). Once the optimum value of the penalty parameter λ is found we obtain the lasso coefficients that have been fit.

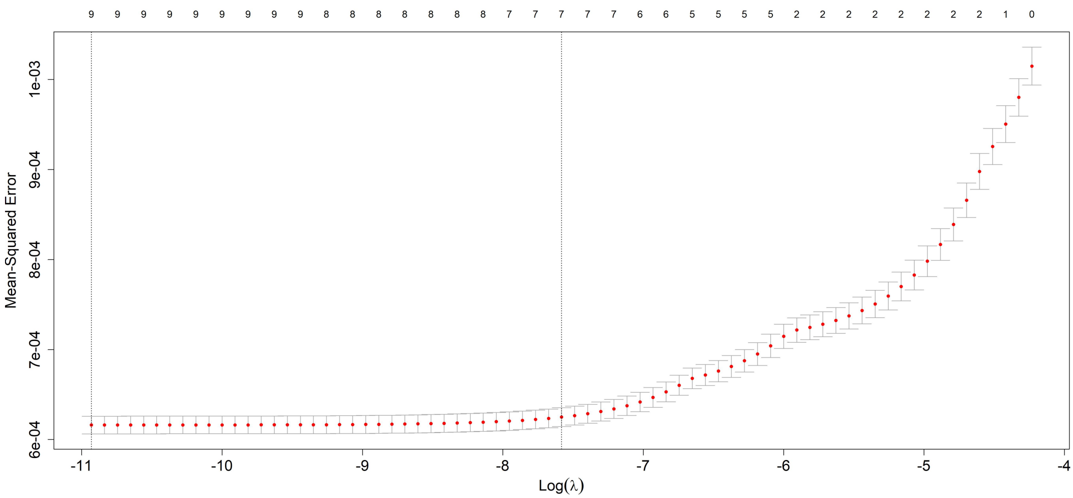

Figure 12. *Mean-Squared-Error for different lambdas.*

3.2.4. Results

The Lasso model is with an MSE of 0.0006167289 for the testing data and an MSE of 0.0006144904 for the training data the model that produces the best predictions for the data. The minimum MSE is achieved when $\lambda = 1.790\,134 \times 10^{-5}$. In the plot below we can also see the impact of different λs in the estimated coefficients (Figure 13). At certain critical values of the tuning parameter λ the coefficients jump out of the model completing dropping to zero. The estimated lasso coefficients get smaller with a higher λ and they all shrink to 0 when $\log \lambda$ is equal to -4. When λ is large only a few coefficients are still in the model. The `food_security` variable starts going upward quite quickly before its effect declines toward 0. The `finance` variable is forced towards zero when λ is quite high and it has a really significant large weight on the model when the other coefficients disappear from the model. The `finance` variable is valuable to the predictions and it takes a large λ value to say that it's not relevant. The Lasso model penalizes all the coefficients. However all the coefficients are chosen in the model and none of them are shrunk exactly to zero.

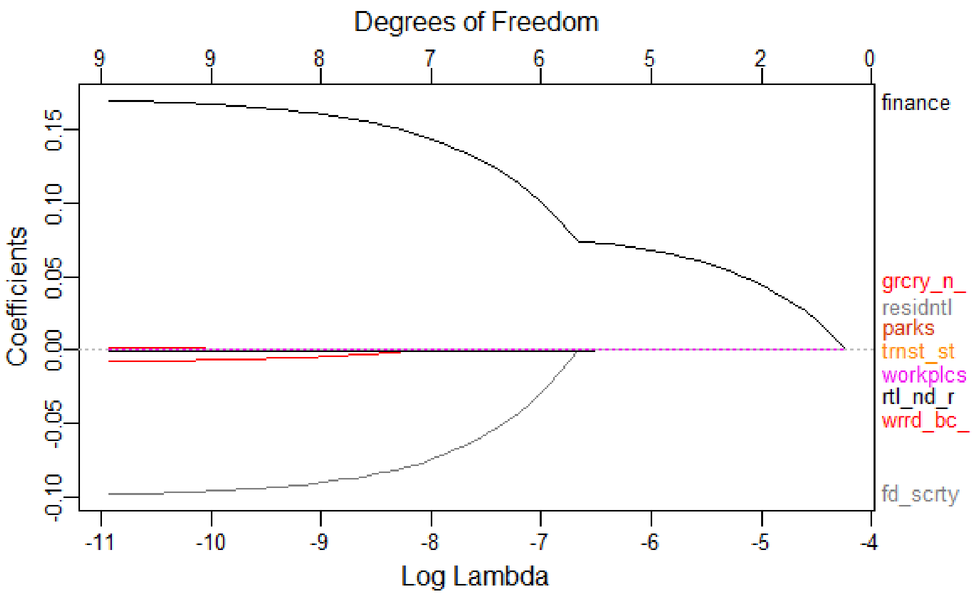

Figure 13. *Lasso path coefficients.*

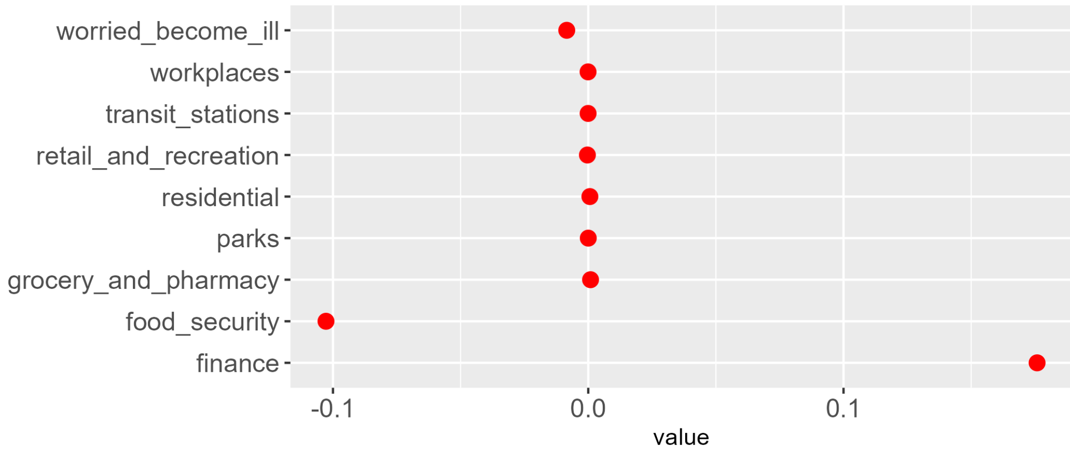

Figure 14. *Coefficients of Lasso model.*

4. Discussion

In general the results of the micro data analysis suggests three major factors: the age, education level and gender. While we found that women appear to be more vulnerable to mental health problems during the pandemic, it should be noted that men are not as open about mental health issues, so some of the results may be biased. Especially further research found men are less likely than women to seek help for mental health concerns and often do not admit mental distress (Sagar-Ouriaghli et al., 2019). The age variable seems to have an tremendous impact on the mental health. We observed younger people to report a higher level of mental health problems than seniors. These findings are also backed by other general mental health studies that propose older age groups are more resilient than younger ones (Thomas et al., 2016). Our findings show that this paradox trend for mental health did continue during COVID-19 and older people experienced less mental problems than young adults. We want to stress that the younger generation seems to have had major mental health problems during the pandemic and therefore suggest targeted mental health care for this group. Our analysis of macrodata revealed varying levels of anxiety across countries and continents over time. African countries recorded the highest levels of anxiety in the past 7 days, while those in Oceania had the lowest levels with some high peaks starting from January, 2022. The trends in America and Asia were relatively stable, while a spike in anxiety levels was observed in Europe during February 2022. The results of the Lasso model shed new light on the relationship between the anxious variable and the other covariates, with all covariates being included in the final model.

Even though our model results meet the general criteria for a confident interpretation, we want to discuss two limitations of our research.

First and most obvious we want to point out that the CTIS study only retrieved data from Facebook users. Countries were Facebook is blocked/censored, such as e.g. China, are not included in the responses. Therefore, the continental Analysis for e.g. Asia should be interpreted with caution. However, as the Facebook user base is quite stable and has a wide rang of users, the transitions of the data to for example government responses are still useful for analysis, even if the correct value might be different.

Secondly, we want to emphasize that the Google trends data from the policy data set used might not entirely explain changes brought on by the pandemic as the resulting value is compared to the median over a five week period from January 3rd to February 6th 2020. Seasonal variation was therefore not taken into account.

In addition to the variables already analyzed, further research could benefit from considering other factors available in the UMD Global CTIS data such as pregnancy, employment status, and field of work (Kreuter et al., 2020). Examining the number of individuals who slept in the same place as the respondent the previous night could also provide insight into the impact of isolation on mental health during the pandemic. Additionally, extending the observation period to include mental health data prior to April 2020 could offer a clearer understanding of the differences in mental health before and during the COVID-19 pandemic. Research could also investigate the effect of different policy responses, such as lockdowns and economic support measures, on mental health outcomes. Finally, studying the long-term effects of the pandemic on mental health is crucial to determine the persistence of its impact and to identify the need for continued support and intervention for those affected. This research highlights the importance of monitoring and understanding the psychological effects of pandemics, which could aid in better preparedness and support in the event of future public health crises.

References

Arel-Bundock, V., Enevoldsen, N., & Yetman, C. (2018). Countrycode: An r package to convert country names and country codes. *Journal of Open Source Software*, *3*(28), 848. https://doi.org/10.21105/joss.00848

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). *Shiny: Web application framework for r* [R package version 1.7.4.9001]. https://shiny.rstudio.com/

Covid-19 Dashboard. (2023). Retrieved January 30, 2023, from https://coronavirus.jhu.edu/map.html

Ettman, C. K., Abdalla, S. M., Cohen, G. H., Sampson, L., Vivier, P. M., & Galea, S. (2020). Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. *JAMA Network Open*, *3*(9), e2019686–e2019686. https://doi.org/10.1001/jamanetworkopen.2020.19686

Fan, J., Li, Y., Stewart, K., Kommareddy, A. R., Garcia, A., O'Brien, J., Bradford, A., Deng, X., Chiu, S., Kreuter, F., Barkay, N., Bilinski, A., Kim, B., Galili, T., Haimovich, D., LaRocca, S., Presser, S., Morris, K., Salomon, J. A., . . . Vannette, D. (2020). The university of maryland social data science center global covid-19 trends and impact survey, in partnership with facebook. https://covidmap.umd.edu/api.html

Godleski, S. A., Harris, C. T., Fitzpatrick, K. M., & Kothari, A. (2022). Social and behavioral vulnerability, pregnancy, and negative mental health outcomes in the U.S. during the Covid-19 pandemic. *AIMS Public Health*, 331–341. https://doi.org/10.3934/publichealth.2022023

Haensch, A.-C., & Xiong, Y. (2021). *The CTIS R package*. https://github.com/CaroHaensch/CTIS

Kreuter, F., Barkay, N., Bilinski, A., Bradford, A., Chiu, S., Eliat, R., Fan, J., Galili, T., Haimovich, D., Kim, B., LaRocca, S., Li, Y., Morris, K., Presser, S., Sarig, T., Salomon, J. A., Stewart, K., Stuart, E. A., & Tibshirani, R. (2020). Partnering with a global platform to inform research and public policy making: What needs to be in place to make a global covid-19 survey work? *Survey Research Methods*, *14*(2), 159–163. https://doi.org/10.18148/srm/2020.v14i2.7761

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., & Roser, M. (2020). Coronavirus pandemic (covid-19) [https://ourworldindata.org/coronavirus]. *Our World in Data*.

McGinty, E. E., Presskreischer, R., Han, H., & Barry, C. L. (2020). Psychological Distress and Loneliness Reported by US Adults in 2018 and April 2020. *JAMA*, *324*(1), 93–94. https://doi.org/10.1001/jama.2020.9740

Sagar-Ouriaghli, I., Godfrey, E., Bridge, L., Meade, L., & Brown, J. S. (2019). Improving mental health service utilization among men: A systematic review and synthesis of behavior change techniques within interventions targeting help-seeking. *American Journal of Men's Health*, *13*(3), 155798831985700. https://doi.org/10.1177/1557988319857009

Sievert, C. (2020). *Interactive web-based data visualization with r, plotly, and shiny*. Chapman; Hall/CRC. https://plotly-r.com

Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software*, *84*(6), 1–39. https://doi.org/10.18637/jss.v084.i06

Thomas, M. L., Kaufmann, C. N., Palmer, B. W., Depp, C. A., Martin, A. S., Glorioso, D. K., Thompson, W. K., & Jeste, D. V. (2016). Paradoxical trend for improvement in mental health with aging. *The Journal of Clinical Psychiatry*, *77*(08). https://doi.org/10.4088/jcp.16m10671

Trevor Hastie, K. T., Junyang Qian. (2021). *An introduction to glmnet*. https://glmnet.stanford.edu/articles/glmnet.html

Acknowledgments. We are grateful for the assistance of Prof. Dr. Frauke Kreuter and Dr. Anna-Carolina Haensch.

Data Availability Statement. The code of this paper can be accessed on Github. The macrodata of the UMD Global CTIS study is publicly accessable, the microdata is under restricted access (see UMD Website).

A. Appendix

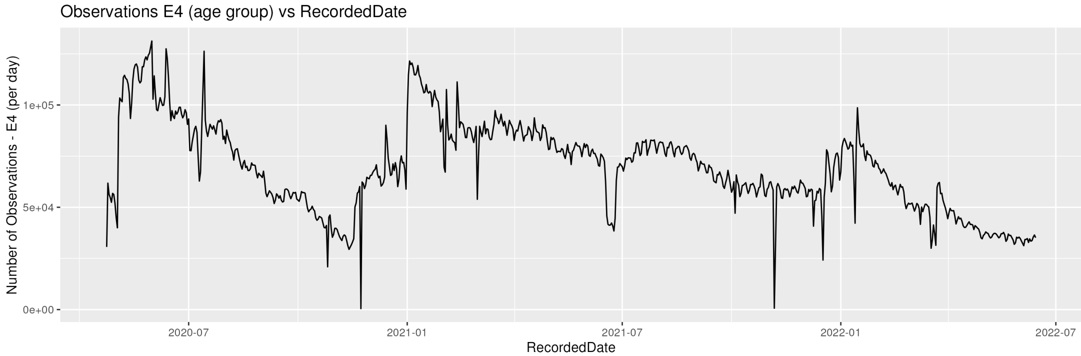

Figure 15. Number of observation of age group variable over time. *Daily Aggregated Data: Observation Period April 23, 2020 to June 14, 2022.*

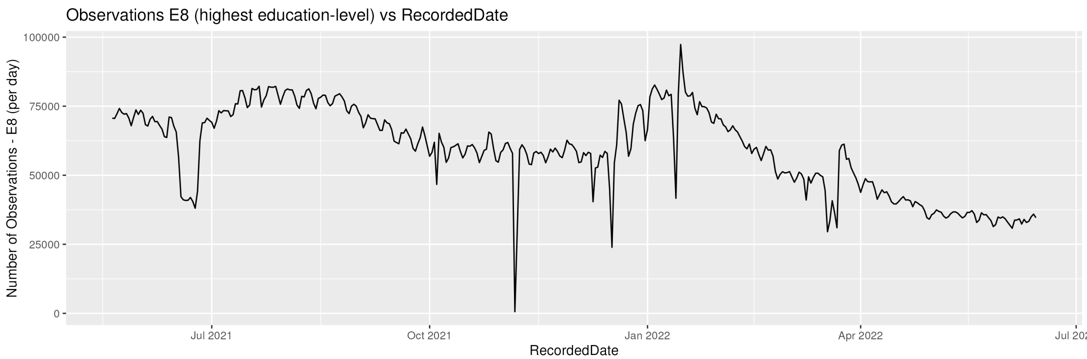

Figure 16. Number of observation of education variable over time. *Daily Aggregated Data: Observation Period May 20, 2021 to June 14, 2022.*

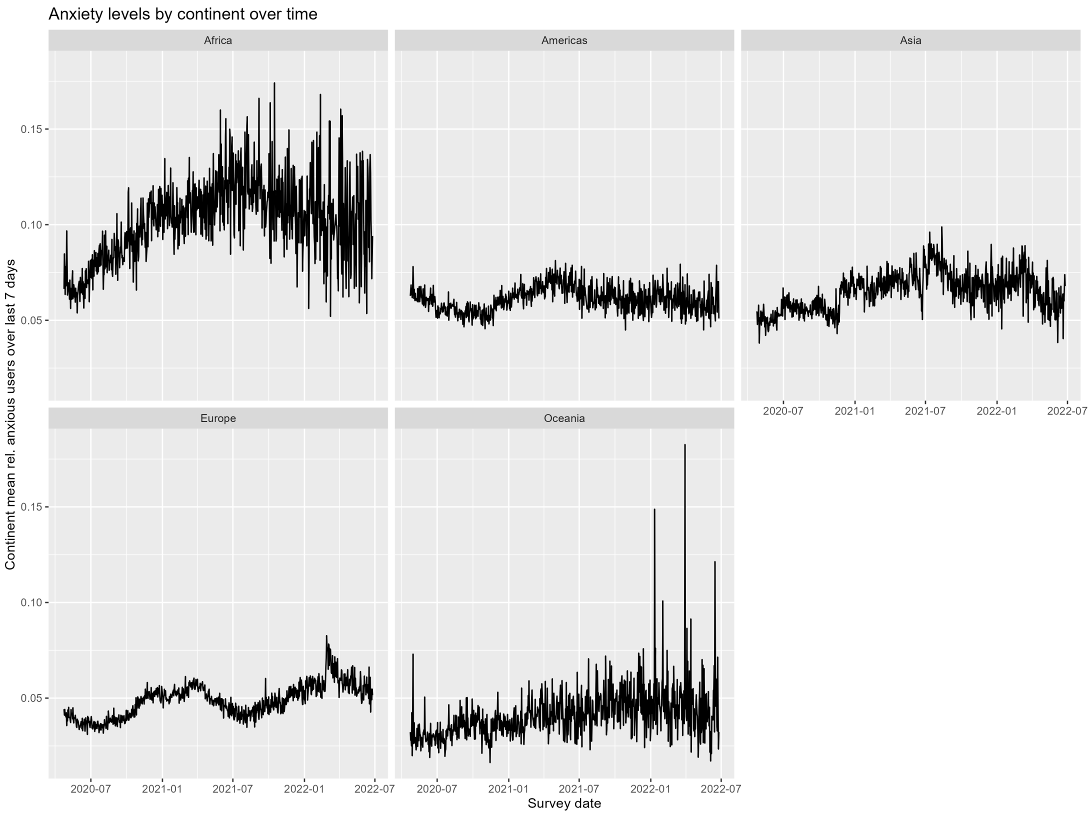

Figure 17. Mean anxiety levels over the past 7 days of every continent over time. *Daily Aggregated Data: Observation Period April 23, 2020 to June 25, 2022.*