

Project Proposal

due October 11, 2021 by 11:59 PM

Mihika Rajvanshi, Bhavika Garg, Michelle Huang

10/11/2021

Load Packages

```
library(tidyverse)
library(infer)
library(readxl)
library(skimr)
library(dplyr)
```

Load Data

```
# covid1=read_excel("../data/COVIDiSTRESS_May30_First.xlsx", sheet= 1)
# covid2=read_excel("../data/COVIDiSTRESS_May30_Second.xlsx", sheet= 1)
# covid3=read_excel("../data/COVIDiSTRESS_May 30_Third.xlsx", sheet= 1)
# covid4=read_excel("../data/COVIDiSTRESS_May 30_Fourth.xlsx", sheet= 1)
# covid5=read_excel("../data/COVIDiSTRESS_May 30_Fifth.xlsx", sheet= 1)
```

Introduction and Data, including Research Questions

After two on-campus suicides in the past month, students at the University of North Carolina at Chapel Hill have spoken out about the mental health crisis that has arisen alongside the COVID-19 pandemic, comparing declining mental health to “a second pandemic”. Mental health crises so close to home highlight how COVID-19 is not merely a medical phenomenon, but also how institutional organizations failed to target socio-psychological consequences that arose from restrictions during this global health crisis (Javed et. al, 2020, Yamada et. al 2021). In particular, evidence indicates that the mental health impacts will be the greatest and longest-lasting globally in under-served populations of LMICs, emphasizing a need for a solution that identifies and addresses the social determinants of mental health during this pandemic (Kola, 2021). Because of the pertinence of this issue, our team has proposed the research question: How does age, gender, and level of isolation affect global mental health outcomes during COVID-19? To explore our question, we will reference data from the COVIDiStress Global Survey focused on the psychological and behavioural impact of the coronavirus crisis from the Open Science Framework database. The collection efforts for this data spanned from March to May of 2020, and used survey forms across 179 countries. In total, the data represents 125,306 people with demographic characteristics for countries with over 200 responses who met the survey inclusion criteria (over 18 years of age providing informed consent) and accessed the online survey link over a period of 62 days. A variety of adults with varying demographics provided self-reports about the proximate psychological effects of the COVID-19 pandemic, which were translated as indicators for mental health outcomes. A wide range of variables were collected during this study, demographic data, psychological effects of COVID-19, risk taking intention, personality characteristics, loneliness levels, stress levels, levels of trust, and willingness to comply with COVID-19 regulations. More specific variables observed include

experiences of distress from COVID-19 restrictions, coping mechanisms during quarantine, and availability of social support. To respond to the survey questions, participants typically supplied their answers on a 6-point scale ranging from ‘Strongly disagree’ to ‘Strongly agree,’ and added text to explain their choices (Yamada, 2021).

References

Javed, B., Sarwer, A., Soto, E. B., & Mashwani, Z. U. (2020). The coronavirus (COVID-19) pandemic’s impact on mental health. *The International journal of health planning and management*, 35(5), 993–996. <https://doi.org/10.1002/hpm.3008>

L. Kola, B.A. Kohrt, C. Hanlon, J.A. Naslund, S. Sikander, M. Balaji, C. Benjet, E.Y. Cheung, J. Eaton, P. Gonsalves, M. Hailemariam, N.P. Luitel, D.B. Machado, E. Misganow, O. Omigbodun, T. Roberts, T.T. Salisbury, R. Shidhaye, V. Su. (2021). COVID-19 mental health impact and responses in low-income and middle-income countries: Reimagining global mental health, *Lancet Psychiatry* (8) pp. 535-550, [https://doi.org/10.1016/S2215-0366\(21\)00025-0](https://doi.org/10.1016/S2215-0366(21)00025-0).

Yamada, Y., Čepulić, DB., Coll-Martín, T. et al. COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak. *Sci Data* 8, 3 (2021). <https://doi.org/10.1038/s41597-020-00784-9>

Glimpse

This is the condensed glimpse because there are 154 variables. We have selected the first 13, but the full list can be found in the readme file. The data was joined via full_join. Code not shown in output file.

```
cleancovid <- subset (covidstressdata, select = -c(AD_gain, AD_loss, RecordedDate, Duration..in.seconds
#removed because of LaTeX incompatible but will use this for analysis

filter(Dem_gender %in% c("Male", "Female"), answered_all == "Yes", Dem_isolation %in% c("Isolated", "Li

glimpsecovid <- subset (covidstressdata, select = -c(Dem_maritalstatus:Scale_UCLA_TRI_avg)) #condensed

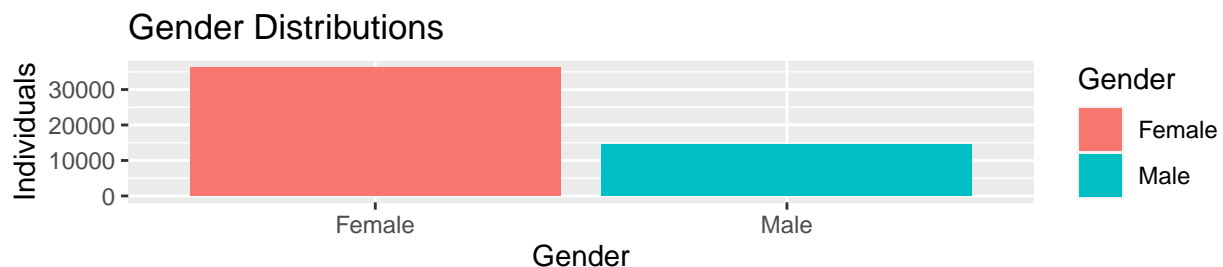
glimpse(glimpsecovid)
```

```
## Rows: 125,306
## Columns: 13
## $ ID <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1~
## $ answered_all <chr> "No", "No", "No", "No", "No", "Yes", "Yes", "Yes~
## $ Duration..in.seconds. <dbl> 180, 3100, 127, 1710, 2239, 1221, 1283, 1442, 19~
## $ RecordedDate <dtm> 2020-05-30 23:47:17, 2020-05-29 23:30:15, 2020-~
## $ UserLanguage <chr> "SAR", "UR", "SAR", "BG", "SAR", "IT", "SAR", "S~
## $ Dem_age <dbl> 29, 20, 47, 79, 61, 68, 29, 38, 35, 23, 42, 31, ~
## $ Dem_gender <chr> "Female", "Male", "Female", "Male", "Female", "M~
## $ Dem_edu <chr> "College degree, bachelor, master", "College deg~
## $ Dem_edu_mom <chr> "Some College or equivalent", "None", "Some Coll~
## $ Dem_employment <chr> "Not employed", "Student", "Self-employed", "Not~
## $ Country <chr> "Argentina", "Pakistan", "Argentina", "Bulgaria"~
## $ Dem_Expat <chr> "yes", "yes", "no", "no", "no", "no", "no", "no"~
## $ Dem_state <chr> "Tucum·n", "<U+0622><U+0632><U+0627><U+062F> <U+~
```

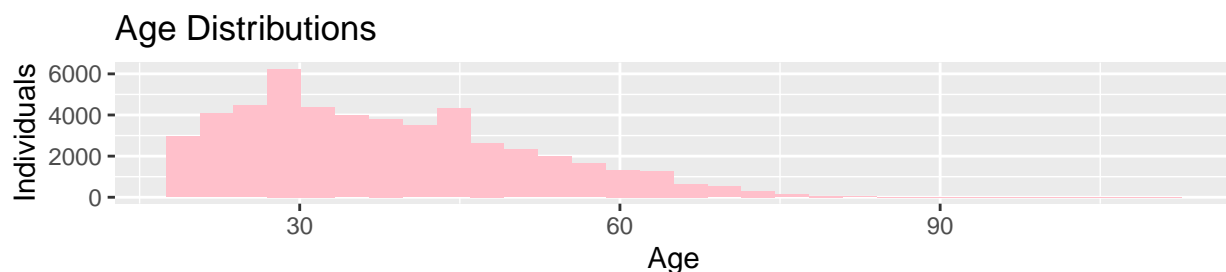
Data Analysis Plan

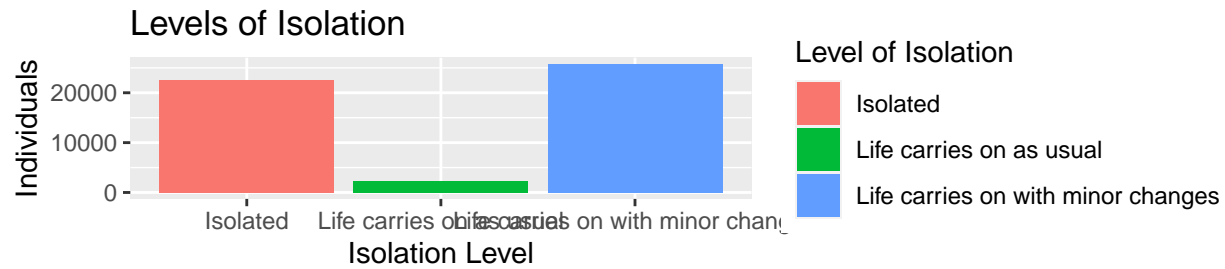
Our main change to the proposal resubmission was removing the t-tests and adding the more appropriate ANOVA tests for our data analysis plan. We also added more predictor variables to get a more holistic view of the data. These changes are incorporated within the following section. In order to answer our research

question, our predictor variables (x) are age, gender, level of isolation, marital status, dependents, employment status, while the outcome variables (y) are the results of several measures of psychological well-being during COVID-19. We will pick the three most relevant measures in the study: PSS10, SPS-15, and SLON-3 scales. The PSS scale measures perceived stress in an individual. SPS measures an individual's social connectedness to others. Finally, the SLON-3 scale measures loneliness in individuals. Together, we take these three scales to represent a quantitative measure of psychological well-being. The comparison groups that will be used will be two groupings of ages, the two genders, the levels of isolation, married or single, number of dependents, and unemployed or employed. We will then consider if the relationship between our factors and the various psychological well-being scales is statistically significant in each of our cases. Because this dataset is extremely large with over 125,000 observations, we will first examine the dataset by understanding the demographics of the data (the visualizations are below). Then we will calculate the mean levels of the PSS-10, SPS-15, and SLON-3 scales in different age groups, gender, levels of isolation, marital status, dependents, employment status. We will also create sets of histograms and barplots that visualize the distribution of these scales. Finally, we will create a series of visualizations just to see the overall geographic distributions with the spatial mapping method. The statistical methods that will be useful in answering our research question would be the ANOVA test. The ANOVA test will be used because we want to find out which one of these factors has a significant effect on the measures of depression we chose. We will primarily focus on age, gender, and levels of isolation as we are most interested in those three and are the factors we made visualizations for. For the age analysis, the age groups will be collapsed into children and adults. For the gender analysis, only female and male individuals will be used while those that answered NA will be discarded. Finally, for the level of isolation, we will collapse the four values of isolation into two groups of isolated or not isolated. We will compare the mean values of our three scales of psychological well-being via ANOVA tests to see which factor is most significantly correlated. Our null hypothesis is that no factor has a statistically significant correlation on any scale of psychological well-being in a global population. Our alternate hypothesis for each of our tests is that age, gender, or level of isolation (or other factors) will have a more significant correlation than other factors with our various scales of psychological well-being, PSS-10, SPS-15, and SLON-3. In order for our hypothesized answer to be correct, the mean levels of the well-being scales must be statistically different (p-value <0.05) when comparing mean values of the PSS-10, SPS-15, or SLON-3 scales across each of our two different scenarios of three explanatory variables (age, gender, or isolation).



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





```
covidforanalysis <- subset (cleancovid, select = c(13:22)) %>%
  filter(Scale_PSS10_UCLA_1 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_2 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_3 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_4 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_5 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_6 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_7 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_8 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_9 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_PSS10_UCLA_10 %in% c(1, 2, 3, 4, 5))
```

```
covidforanalysis$Scale_PSS10_UCLA_1 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_1))
covidforanalysis$Scale_PSS10_UCLA_2 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_2))
covidforanalysis$Scale_PSS10_UCLA_3 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_3))
covidforanalysis$Scale_PSS10_UCLA_4 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_4))
covidforanalysis$Scale_PSS10_UCLA_5 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_5))
covidforanalysis$Scale_PSS10_UCLA_6 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_6))
covidforanalysis$Scale_PSS10_UCLA_7 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_7))
covidforanalysis$Scale_PSS10_UCLA_8 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_8))
covidforanalysis$Scale_PSS10_UCLA_9 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_9))
covidforanalysis$Scale_PSS10_UCLA_10 <- as.numeric(as.character(covidforanalysis$Scale_PSS10_UCLA_10))
```

```
covidforanalysis <- covidforanalysis %>%
  select(Scale_PSS10_UCLA_1:Scale_PSS10_UCLA_10) %>%
  mutate(PSSmeans = rowMeans(across(where(is.numeric))))
```

```
covidforanalysis2 <- subset (cleancovid, select = c(23:25)) %>%
  filter(Scale_SLON_1 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_SLON_2 %in% c(1, 2, 3, 4, 5))%>%
  filter(Scale_SLON_3 %in% c(1, 2, 3, 4, 5))
```

```
covidforanalysis2$Scale_SLON_1 <- as.numeric(as.character(covidforanalysis2$Scale_SLON_1))
covidforanalysis2$Scale_SLON_2 <- as.numeric(as.character(covidforanalysis2$Scale_SLON_2))
covidforanalysis2$Scale_SLON_3 <- as.numeric(as.character(covidforanalysis2$Scale_SLON_3))
```

```

covidforanalysis2 <- covidforanalysis2 %>%
  select(Scale_SLON_1: Scale_SLON_3) %>%
  mutate(SLONmeans = rowMeans(across(where(is.numeric))))

covidforanalysis3 <- subset (cleancovid, select = c(101:110)) %>%
filter(SPS_1 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_2 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_3 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_4 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_5 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_6 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_7 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_8 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_9 %in% c(1, 2, 3, 4, 5, 6))%>%
filter(SPS_10 %in% c(1, 2, 3, 4, 5, 6))

covidforanalysis3$SPS_1 <- as.numeric(as.character(covidforanalysis3$SPS_1))
covidforanalysis3$SPS_2 <- as.numeric(as.character(covidforanalysis3$SPS_2))
covidforanalysis3$SPS_3 <- as.numeric(as.character(covidforanalysis3$SPS_3))
covidforanalysis3$SPS_4 <- as.numeric(as.character(covidforanalysis3$SPS_4))
covidforanalysis3$SPS_5 <- as.numeric(as.character(covidforanalysis3$SPS_5))
covidforanalysis3$SPS_6 <- as.numeric(as.character(covidforanalysis3$SPS_6))
covidforanalysis3$SPS_7 <- as.numeric(as.character(covidforanalysis3$SPS_7))
covidforanalysis3$SPS_8 <- as.numeric(as.character(covidforanalysis3$SPS_8))
covidforanalysis3$SPS_9 <- as.numeric(as.character(covidforanalysis3$SPS_9))
covidforanalysis3$SPS_10 <- as.numeric(as.character(covidforanalysis3$SPS_10))

covidforanalysis3 <- covidforanalysis3 %>%
  select(SPS_1:SPS_10) %>%
  mutate(SPSmeans = rowMeans(across(where(is.numeric))))

```