Project Proposal

due October 11, 2021 by 11:59 PM

Mihika Rajvanshi, Bhavika Garg, Michelle Huang

10/11/2021

Load Packages

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

Load Data

```r
covid1=read_excel("/home/guest/R/the-statistical-power-puffs/data/COVIDiSTRESS_May30_First.xlsx", sheet=
covid2=read_excel("/home/guest/R/the-statistical-power-puffs/data/COVIDiSTRESS_May30_Second.xlsx", sheet
covid3=read_excel("/home/guest/R/the-statistical-power-puffs/data/COVIDiSTRESS_May 30_Third.xlsx", sheet
covid4=read_excel("/home/guest/R/the-statistical-power-puffs/data/COVIDiSTRESS_May 30_Fourth.xlsx", she
covid5=read_excel("/home/guest/R/the-statistical-power-puffs/data/COVIDiSTRESS_May 30_Fifth.xlsx", sheet
```

#Join Data

```r
covidstressdata <-covid1 %>%
  full_join(covid2) %>%
  full_join(covid3) %>%
  full_join(covid4) %>%
  full_join(covid5)
```

Our dataset had to be split in order to upload to github. Here the data is joined.

Introduction and Data, including Research Questions

After two on-campus suicides in the past month, students and administrators 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 and suicide to "a second pandemic" (Klessen, 2021). Mental health crises so close to home highlight how a pandemic like COVID-19 is not merely a medical phenomenon, but also show how institutional organizations have failed to destigmatize and target the socio-psychological consequences that arose from social and economic restrictions during this global health crisis (Javed et. al, 2020, Yamada et. al 2021). In particular, emerging evidence indicates that the mental health impacts will be long-lasting and greatest globally in under-served populations of lower and middle income countries, emphasizing a need for a balanced solution that identifies and addresses the social determinants of mental health during this pandemic (Kola, 2021). Because of the pertinence and importance of this critical 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 available in 47 languages across 179 countries. In total, the stored dataset represents 125,306 people with demographic characteristics for countries with over 200 responses who met the survey inclusion criteria (18 years of age and older and gave informed consent) and accessed the online survey link to provide their experiences over a period of 62 days. A variety of adults with varying personalities, loneliness levels, and stress levels 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, including general demographic data, psychological effects of COVID-19, levels of risk taking intention during the pandemic, personality characteristics, loneliness levels through the UCLA loneliness scale, stress levels from the Perceived Stress Scale, levels of interpersonal and institutional trust during the pandemic, and willingness to comply with pandemic regulations. More specific variables observed include experiences of distress and worry due to COVID-19 restrictions, psychological coping mechanisms during quarantine, and availability of social support using the Social Provisions Scale. To respond to the survey questions, participants typically supplied their answers on a 6-point Likert scale ranging from 'Strongly disagree' to 'Strongly agree,' and added text to explain some of their choices (Yamada, 2021)..

References Klessen, B. (2021). UNC Chapel Hill cancels classes Tuesday amid two suicide investigations. NBC News. https://www.nbcnews.com/news/us-news/unc-chapel-hill-cancels-classes-tuesday-amid-two-suicide-investigations-n1281229.

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.

Yamada, Y., Ćepulić, 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 19, but the full list can be found in the readme file.

```r
cleancovid <- subset (covidstressdata, select = -c(AD_gain, AD_loss)) #removed because of laTex incompa
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         <dttm> 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 research question is "How do age, gender, and the level of isolation affect mental health during COVID-19 globally?" In order to answer this, our predictor variables (x) are age, gender, and level of isolation, 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, and the levels of isolation. We will then consider if the relationship between our factors and the various psychological well-being scales is statistically significant in each of our three 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, and levels of isolation. We will also create sets of histograms and barplots that visualize the distribution of these scales in different age groups, gender, and levels of isolation. 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 a two-tailed, two-sample t-test. The two-tailed two-sample t-test will be used because the populations are independent from each other and we are measuring an overall difference. 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 for a series of nine paired t-tests overall.

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 will have a correlation 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).

```
cleancovid %>%
  filter(Dem_gender %in% c("Male", "Female")) %>%
  ggplot(aes(x = Dem_gender)) + geom_bar()
```
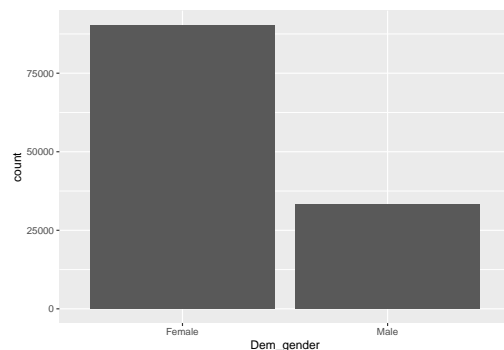
```
cleancovid %>%
  ggplot(aes(x = Dem_age)) + geom_histogram()
```

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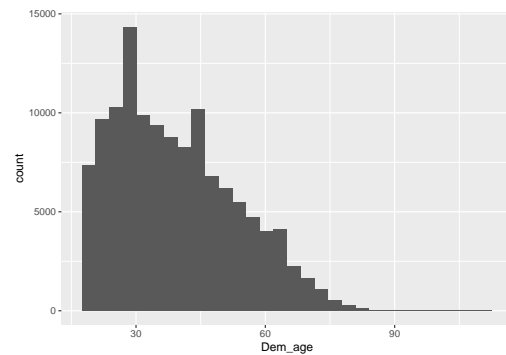

```
cleancovid %>%
  filter(Dem_isolation %in% c("Isolated", "Life carries on with minor changes", "Life carries on as usu
  ggplot(aes(x = Dem_isolation)) + geom_bar()
```