Anxiety & Depression in the time of COVID

# description

There has been much concern about the impact of the COVID pandemic on our population’s mental health. And young people, in particular, whose education has been disrupted on a massive scale. But in researching this topic, one can find studies supporting both positions on this topic. I even found a study that claimed that teenagers haven’t been negatively impacted as they are getting more sleep and spending more time with their families. Our study aims to put that question to rest…..

# questions to answer

The fundamental questions this study will explore (and hopefully answer) are:

1. Has the COVID pandemic had a negative impact on the mental health of our population?
2. In particular, has it impacted the mental health of young adults (18-29)?
3. If so, in what ways has the pandemic impacted young adults?
   * By disrupting their college plans?
   * By putting them under financial strain (inability to pay rent, not enough food, etc.)?
   * By impacting their health, the health of their friends or family members?

# data sources

A preliminary glance at the data looks revealing:

* Anxiety/depression levels in 2019 were ~11%,
* In Jan. 2021 they were ~35%
  + This was even higher amongst young adults: 46%

The key datasets that will be used are the following:

1. Anxiety & depression levels prior to the pandemic – 2019 baseline data from CDC:
   * <https://www.cdc.gov/nchs/data/nhis/earlyrelease/ERmentalhealth-508.pdf>
2. Anxiety & depression levels during the pandemic – 9 months of Pulse survey results from CDC:
   * <https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm>
3. Johns Hopkins University COVID confirmed cases and death rates (2020-current):
   * <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-datasets>
     + time\_series\_covid\_19\_confirmed\_us.csv, time\_series\_covid\_19\_deaths\_us.csv
4. Key factors impacting young adults – snapshot (latest survey results) from Phase III of the CDC Household Pulse Survey
   * https://www.census.gov/programs-surveys/household-pulse-survey/data.html#phase3

# rough task breakdown

Key tasks that will need to be performed are outlined below.

## Data Exploration & Cleanup

1. The historical Pulse IAD .csv data and the JHU COVID .csv data should both be aggregated into weekly buckets using the same time intervals.
2. Key data elements to extract to Pandas dataframes for analysis:
   1. 9 month Pulse IAD df - columns:
      1. Date (use the Wednesday of each week?)
      2. State (All, California, New York)
      3. Symptoms of Anxiety or Depressive Disorder (percentage)
      4. Age (18-29, 30-39, etc.)
      5. Gender (M,F)
   2. JHU COVID Data df - columns:
      1. Date (use the Wednesday of each week?)
      2. Confirmed Cases
      3. Deaths
      4. Latitude & Longitude? (If we think we might want to geomap hotspots as part of our California & New York timelines we should probably grab them)
   3. Latest survey results from Phase III of the Pulse Household survey df:
      1. Education Table 6 df - columns:
         1. Date
         2. Age
         3. State
         4. Gender
         5. Plans to take classes this fall have not changed
         6. All plans to take classes this fall have been canceled
         7. Classes will be in different formats in the fall
         8. Fewer classes will be taken this fall
         9. More classes will be taken this fall
         10. Classes will be taken from a different institution
         11. Classes will be taken for a different kind of certificate or degree
         12. Did not report
         13. Total

## Key Analysis

1. Comparison of Indicators of Anxiety & Depression between 2019 baseline and most recent survey snapshot.
2. Relationship between anxiety/depression levels and COVID confirmed cases & deaths (2020-current)
3. Exploration of the potential contributing factors. Start with Education impacts for 18-29 year olds and add additional factors as appropriate (and time permits).

## Visualizations

1. Bar charts comparing baseline IAD percentage from 2019 to:
   1. First week of Pulse survey (April 28-2020)
   2. Most recent 2021 Pulse survey
2. Bar charts of IAD percentage from April 2020-current for:
   1. All
   2. 18-29 year olds
   3. California & New York
3. Bar chart of COVID confirmed cases & deaths for 2020-current
   1. All (US).
   2. California, New York
4. Scatterplot with regression line & correlation coefficient for:
   1. IAD % vs. Confirmed Cases
   2. IAD% vs. Deaths
   3. California & New York (if the bar charts in #2 & #3 look promising)
5. Pie charts & bar charts of Breakdown of Educational factors potentially impacting IAD levels of 18-29 year olds.
6. Explore producing Extra Credit Line chart: Combined IAD and COVID data
7. X-axis=date, on the left hand side of the chart the y-axis label would be IAD%, on the right-hand side of the chart the y-axis label would be Confirmed Cases or Deaths. We would plot both lines on the chart. (P.S.- I’m not sure if, how this can be done but would be powerful).
8. Explore producing Extra Credit heatmap of COVID hotspots for specific time periods if our analysis determines it would be relevant.