**Capstone Final Project Summary**

**Name:** Catherine Yarovoi

**Project Title:** Pediatric Mental Health - 2020 SAMHSA Client-Level Data

**GitHub Link:** <https://github.com/catyar/Pediatric-Mental-Health-2020-SAMHSA-Client-Level-Data>

**Tableau Link:** <https://public.tableau.com/app/profile/catherine2385/viz/PediatricMentalHealth-2020SAMHSAClient-LevelData/PediatricMentalHealth2020-SAMHSAClient-LevelData>

**Data Collection Source:**

Data was downloaded from <https://www.samhsa.gov/data/data-we-collect/mh-cld-mental-health-client-level-data>

The data comes with a codebook, which was used to decipher the variables in the data: <https://www.datafiles.samhsa.gov/sites/default/files/field-uploads-protected/studies/MH-CLD-2020/MH-CLD-2020-datasets/MH-CLD-2020-DS0001/MH-CLD-2020-DS0001-info/MH-CLD-2020-DS0001-info-codebook.pdf>

One section of the analysis also looks at population information, and thus includes data from <https://cwoutcomes.acf.hhs.gov/cwodatasite/population/index>

The Data:

* Provided by Substance Abuse and Mental Health Services Administration (SAMHSA)
* Looks at Mental Health Client-Level Data (MH-CLD) system which provides information on mental health diagnoses and the mental health treatment services, outcomes, and demographic and substance use characteristics of individuals in mental health treatment facilities.

Time Frame:

* Data collected by SAMHSA for 2020 fiscal year.
  + Most states have a July 1 through June 30 fiscal year, but FY2020 may also be defined as an October through September fiscal year, a lagged fiscal year, or both for some states.

Limitations Of the Data:

* Dataset does not represent all national demand for mental health services.
  + Only includes data from facilities that operate under the authority of their state mental health administration.
* Some values populated as Missing/Unknown/Not collected/invalid across most columns; These were labeled as "Undefined" throughout the analysis.

**Project Summary:**

Project Objective

* Understand the use of treatment services, demographics, and characteristics of minors (children under 18 years of age) serviced through state mental health agencies.
* Assess the comorbidity of mental health diagnoses for this population.
* Provide state agencies with more information on the pediatric populations they are serving by understanding mental health needs in their state(s)

The audience may include mental health practitioners, especially those working with minors, who may want to better understand the landscape of minor mental health in the US. This dataset covers a large portion, but not all, mental health cases from FY2020.

Other audiences may include government officials or SAMHSA members for the purposes of allocating resources (training or staffing) to, or incentivizing distribution of resources to, areas and communities where assistance is most needed and utilized.

Methodology:

* This analysis was primarily descriptive. Data for this report was accessed through the SAMHSA website ([www.samhsa.gov/data](http://www.samhsa.gov/data)) and downloaded locally in CSV format. This data was then accessed and transformed via Jupyter Notebook and Big Query. Because of the use of Big Query and the size of the original CSV file, data was also temporarily stored in Google Cloud.
  + Some data was processed via Excel (see: Comorbidity dashboards). For this analysis, a matrix product was executed via Python and then color-coded and visualized via Excel.
* This analysis used a subset of the full dataset. This subset looks at pediatric clients only (children under 18 years of age), although the full dataset provides data on clients of all ages.
* The python libraries used for this analysis included: pandas, numpy, dask.dataframe, statistics, pandas\_gbq, os, google-cloud-storage, google-cloud-bigquery[pandas], and google-cloud.
* In the dataset, each row has a unique Case ID representing one record for each individual served. Therefore, a distinct count of Case ID was used as the primary metric throughout the analysis to indicate total number of cases seen across all state-operated mental health facilities.
* The dataset also used variables (rather than true values) created from the original variables submitted by the states. Therefore, all variables had to be mapped to their respective values for the data to be easily interpretable (e.g., Value "1" for Column dimension of "Race" had to be mapped to Label "American Indian/Alaska Native "). Some columns were left in their variable state in cases where the variables were binary and could be used as Boolean datatypes.

Additional Research or Data

When I began mapping case data to states, I realized I did not have a comparison benchmark. In other words, were the number of cases in each state proportionate to the population of children under 18 years old in that state?

I wanted to create an index to look at the number of children in each state as a percentage of the total US population of children under 18 versus the percentage of total pediatric cases reported by SAMHSA in that state. To do this, I pulled in an additional data source, which included population data for children under 18 years old by state in the year 2020: <https://cwoutcomes.acf.hhs.gov/cwodatasite/population/index>.

Key Takeaways and Recommendations for Further Investigation:

- Trauma or stressor related disorders are the most common diagnosis for clients under 15 years old, but ADHD is also among the most common diagnoses.

* ADHD cases decreased while Depression and Anxiety increased as the age range increased.
* It is important to further understand why certain diagnoses become more prevalent (or less prevalent) as children develop in order to properly treat, service, and staff for, client needs.

- While most clients in the SMHA system do not have a substance use problem or diagnosis, 17% of pediatric clients *do* report having substance use problems. This is a substantial percentage, especially considering the analysis is assessing a pediatric population.

* It is important to better understand how substance use intersects with mental health disorders.
* Special consideration must be given to clients reporting substance use problems, and SMHA staffing should reflect training in treating pediatric substance use cases.

- Almost a third of clients with a mental health disorder diagnosis have more than one diagnosis, meaning there is a high comorbidity rate among this population.

* A key to SMHA success will be ensuring service providers are equipped to treat clients with multiple diagnoses in a way that acknowledges the additional challenges such clients face.

- Patients skew white, male, and non-Hispanic or Latino.

* While this is representative of the dataset, it may not be representative of the true SMHA needs of the US pediatric population.
* There may be confounding factors leading to these skews, such as a lack of an understanding of differences in disorder presentation between different gender and culture groups.
* Future research and surveys should try to include additional demographic data such as income, family medical history, urban-rural classification, and sexual orientation, as well as distinguish between sex and gender. Such additional information might help provide context on psycho-social factors that affect patient care, as well as their (or their guardian(s)’s willingness to seek care and/or their access to care.

- The patient population is disproportionately large compared to the state's overall pediatric population in certain states.

* These disparities also differ by SMHA service type.
* Additional analyses to understand what factors are driving over indexing of needs in certain states would prove beneficial.

For Further Consideration:

- Additional analyses should look to understand how this data compares to the adult client population.

- Some states did not provide data. It may be beneficial to understand how reporting and legislative practices differ from state to state and, in turn, how they affect data quality and hence data interpretation.