U.C. Berkeley School of Information

W200: Introduction to Data Science Programming

Professor: Dr. Gerald Benoit

Exploratory Data Analysis of DoD Serivce Members and Veterans Suicide Risk[1]

Authors: Denise Ferniza, Gauri Ganjoo, Elise Gonzalez

[1] Circulation: Internal, Project Partners, UC Berkeley I School

Introduction

Suicide continues to plague service members and veterans alike. The greatest threat of suicide exists during the transitional period between military service and becoming a veteran[2]. The Department of Defense continues to call for high-level solutions and digital solutions that can address these issues. There is a disparity between care providers, Command Teams, Veterans Affairs, and other pertinent services that inhibit full utilization of data which could provide critical and emergent care when needed.

This project aims to provide a small portion of a larger code which will provide actionable information for psychiatric and psychological practitioners within the Armed Forces and VA. This will allow transitional, and continued psychological care for Soldiers, Airmen, Sailors, and Marines that may or may not have preexisting conditions. The proposed specific condition of this code will be to enroll new users into the application, provide input of symptoms, link providers, provide emergent contact information, and file symptomatology. Once complete this project aims to compete for government funding. [3]

This project is in its infancy and this portion focuses on the creation of two objects intended to combine data and make it usable. One object is a data object which imports datasets from the DoD. The second object is a data visualization object which creates various graphs from the datasets and highlights an individual's 'placement' on the graph against known demographics for suicide risk such as age, race, and branch of service[4] [5] [6]. For example, a scatter plot might highlight the individual point which corresponds to an individual's datapoint. A bar chart might highlight the bar that the individual would be included in. The purpose of these visualizations is more for a healthcare provider to get a quick synopsis of where an individual relates to known risk factors. The main goal of this project is to create a rudimentary dashboard from an individual's and government datasets.

This code consist of the following classes:

- Exception Class (prototype)
    - Handling custom exceptions raised in the Data Handling and Servicemember

[2] Ravindran, Morley, Stephens, Stanley & Reger, 2020

[3]
- RFA-MH-21-111: "Service-Ready Tools for Identification, Prevention, and Treatment of Individuals at Risk for Suicide", closing date of June 2022.
- "Service-Ready Tools for Identification, Prevention, and Treatment of Individuals at Risk for Suicide" (R34 Clinical Trial Optional) June 2022
- Notice of Special Interest to: "Highlight High Priority Research Opportunities on Suicide Prevention Crisis Services" closing date of January 2022,
- Notice of Special Interest to: "Highlight Research Priorities for Risk Algorithms Applications in Healthcare Settings to Improve Suicide Prevention ", closing date of January 2022,
- Notice of Intent to Publish a Funding Opportunity Announcement for: "Service-Ready Tools for Identification, Prevention, and Treatment of Individuals at Risk for Suicide (R34 Clinical Trial Optional)" no closing date.

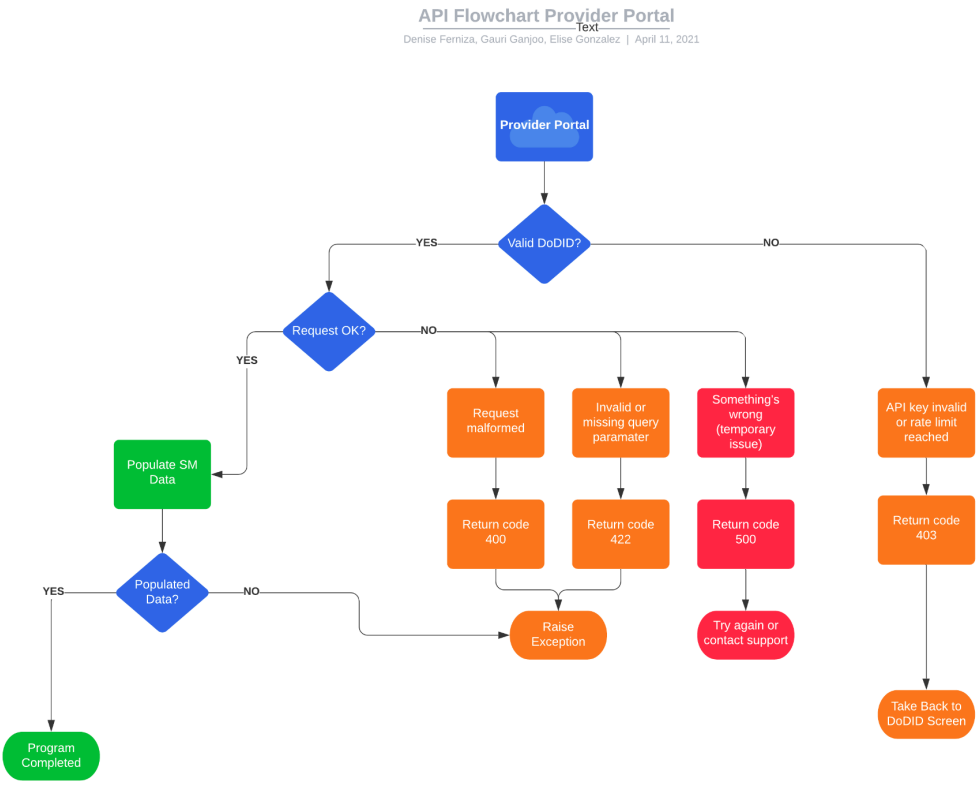
[4] DOD(2018)
[5] DOD(2019)
[6] DOD(2020)

class

- Handle data conversion, integration, and display exceptions
- Handle python default exceptions

- Display Class
  - Handles conversion of different data structures
  - Converts data to JSON files that can be utilized in the application
  - Handles construction of visualizations
- GUI Class(prototype)
  - GUI for interacting with the Data Handling, Service Member, and Exceptions class
- Roster class
  - Imports and converts SM data from CSV files for utilization

Application Programming Interface (Prototype)



Importing Data

This project imports data from two sources.

1. 2019 DoD Annual Suicide Report
2. 2020 National Veteran Suicide Prevention Annual Report

Since there were only 3 tables useful for this application in these reports, the relevant data were entered by hand into csv files. The csv files were then imported into a Jupyter notebook and restructured into json files. Three tables from Source 1 were used to make three json files containing information about current service members, and one table in Source 2 was used to create a json file with information about veterans. The data was compiled into json files so that it could be easily read into pandas dataframes later on. Information about how the individual files are compiled is below.

Compilation of Data

File 1: 2019_suicide_bycmpt.json
Stores data about the number of suicides per military component. Though our sample roster of soldiers contains only Army members, this data contains statistics about each military branch and component. The purpose of including other branches is to provide context to medical service providers regarding a soldier's relative risk of suicide given the branch and component in which they serve.
Data source: 2019 DoD Annual Suicide Report

Structure of json:

```
{ 'Army' : Percentage of total suicides ,
  'Marine Corps' : Percentage of total suicides,
 ...for each military component
 }
```

Components include: 'All Active', 'Army', 'Marine Corps', 'Navy', 'Air Force', 'All Reserve', 'Army Reserve', 'Marine Corps Reserve', 'Navy Reserve', 'Air Force Reserve', 'All National Guard', 'Army National Guard', 'Air National Guard'.

Note that statistics for 'All Active', 'All Reserve', and 'All National Guard' are a combination of those components across all branches.

File 2: 2019_suicide_bycmpt_bydemo.json
Stores data regarding the number of suicides per demographic group within each component, aggregated across all branches.
Data source: 2019 DoD Annual Suicide Report

Structure of json:

```
{ Military Component:
   { Demographic group : Percentage of total suicides,
    ...for each demographic group }
  ...for each component
  }
```

Demographic groups for soldiers are gender, ages, race, grade, and marital status.[7] Components included are: 'Active', 'Reserve', and 'National Guard'. Components are aggregated in this dataset; these three components reflect the same combinations of service branches as 'All Active', 'All Reserve', and 'All National Guard' in File 1 (i.e. “Reserve” reflects the percentage of total SM suicides completed by SM in the Army Reserve, Marine Corps Reserve, Navy Reserve, and Air Force Reserve combined).

File 3: 2019_suicidemethod_bycmpt.json
Stores data regarding the use of particular methods of suicide within each component, aggregated across all branches.[8]
Data source: 2019 DoD Annual Suicide Report

Structure of json:

```
{ Military Component :
      { Method of suicide: Percentage of total suicides,
        ...for each method }
 ...for each component[9]
 }
```

File 4: 2018_vetsuicide_bydemo.json
Stores data about the age, race, and sex of veterans who committed suicide in 2018.
Data source: 2020 National Veteran Suicide Prevention Annual Report[10]

[7] 'Male', 'Female', '17-19', '20-24', '25-29', '30-34', '35-39', '40-44', '45-49', '50-54', '55-59', 'White', 'Black/African American', 'American Indian/Alaska Native', 'Asian/Pacific Islander', 'Other/Unknown', 'All Enlisted', 'E1-E4', 'E5-E9', 'O (Commissioned Officer)', 'W (Warrant Officer)', 'Cadet', 'Never Married', 'Married','Divorced', 'Unknown', 'Total'

[8] Methods include: 'Firearm', 'Hanging/Asphyxiation', 'Drugs/Alcohol', 'Sharp/Blunt Object', 'Poisoning', 'Falling/Jumping', 'Other', 'Pending/Unknown', 'Total'

[9] Components include: 'Active', 'Reserve', 'National Guard', each combining all branches in which that particular component exists.

[10] There is no information included about the component in which these veterans served.

Structure of json:

{ Demographic Group:
{ Age Group : Percentage of total suicides,
...for each age group },
...for each demographic group[11]
}

In its current stage, the program reads files 1 and 2 into separate pandas dataframes, which are then manipulated to form graphs relevant to the particular soldier a provider is looking up.

Charts and Figures

For any DODID corresponding to a soldier or veteran that a provider enters into the system, four figures will appear. These figures are:

1. % of Service Member Suicides by Specific Component, 2019
2. % of [Active Duty, Reserve, or National Guard] Suicides by Age, 2019
3. % of [Active Duty, Reserve, or National Guard] Suicides by Sex, 2019
4. % of [Active Duty, Reserve, or National Guard] Suicides by Grade, 2019

These figures are customized to show which groups the relevant soldier is a part of and, as such, their relative suicide risk. The groups the relevant soldier is a part of is highlighted in red. Figure 1 shows what percentage of suicides were committed by a soldier from each branch, with the components being those available in File 1.

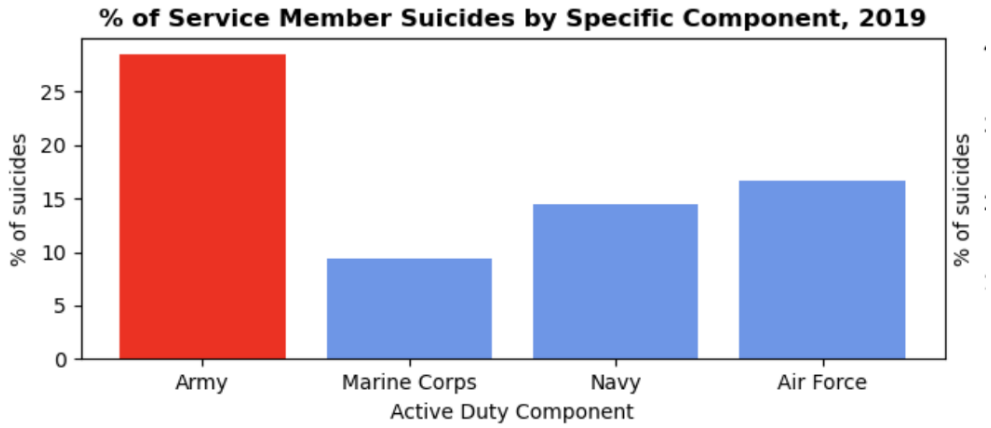


Figure 1 Active Duty Suicides Branch

The data represented in figures 2, 3, and 4 are customizable depending on if the soldier is Active Duty, in the Reserve, or in the National Guard. For example, if the DODID entered by the provider belongs to a soldier in the National Guard, then figures 2, 3, and 4 would present data regarding the National Guard specifically. Beyond that, the data is stratified by age, sex, and grade, respectively. Again, the groups to which the relevant soldier belongs are highlighted in red so the provider can see their level of risk relative to other demographics or

[11] Demographic groups: Male, Female, 18-34, 35-54, 55-74, 75+, White, Black/African American, Multiple Race, Asian/Native Hawaiian/Pacific Islander, American Indian/Alaskan Native

grades within their component. Figure 1 indicates that the soldier, who will be called soldier A, is in the Army and has a higher risk of suicide compared to other soldiers of the same Active Duty status in different miltary branches. Figure 2 shows that soldier A is between 30 and 34 years old and at a much lower risk of suicide compared to other soldiers of the same Active Duty status in their 20s. Figure 3 shows that soldier A has a pay grade between O1 and O9 and is less at risk for suicide than other soldiers of the same Active Duty status with lower pay grades. Figure 4 shows that soldier A is male and has a higher risk of suicide compared to female soldiers with the same Active Duty status.

Once the data for the visualizations is read into the program, the highlight is applied to individuals from the constructed csv. The csv is read-in through a roster class as a dictionary of dictionaries. The development team debated creating a separate patient class for an individual, but decided against it because in the real world, individuals would be added to the database via weekly data pulls, not by individual entries. Dictionaries are also more flexible and will be easier to change later as the project is brought up to scale and more attributes are added to an individual patient's record.

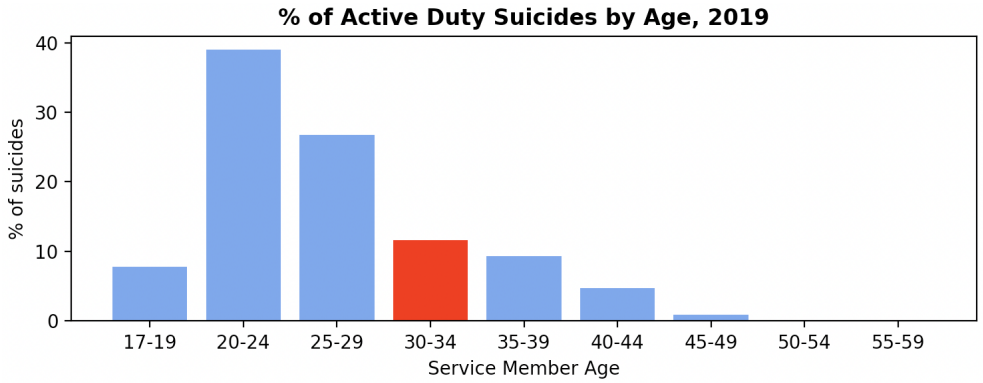


Figure 2 Active Duty Suicides by Age

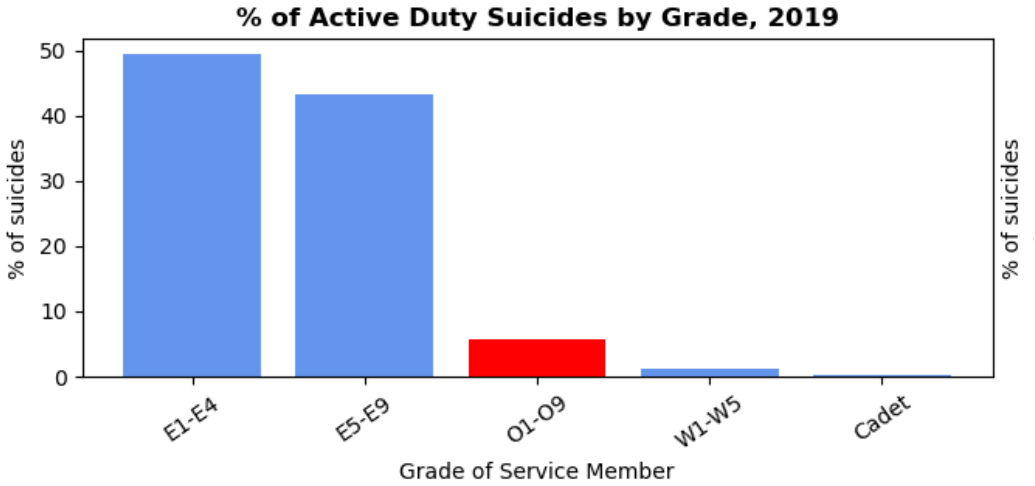


Figure 3 Active Duty Suicides by Grade

Next, a display class was created that laid out four subplots that saved an individual soldier's information. All of the plots are bar plots for readability and allowed for simple communication of risk via highlights. Next, a solution to how to highlight an individual on the bars was needed to distinguish between multiple bars. There was not a readily available method to do this, so the color parameter in the plt.bar method was used to accomplish this. This was done by looping through the xaxis indices for a given graph while creating a list object with a list of two colors repeated. If the soldier is a part of the index group, "red" would be added to the list. If not, then "cornflower blue" would be added to the list. This list object would then be fed in as the color parameter when creating the bar plot. Some issues that we troubleshooted when adding this feature included: axis names changing depending on which data the plot was created from, calculating age with datetime, checking the datetime gave the correct year (ex: 1976 not 2076), and having an index cutoff in the json files of at least 0.001 so that each index would have a visible bar.

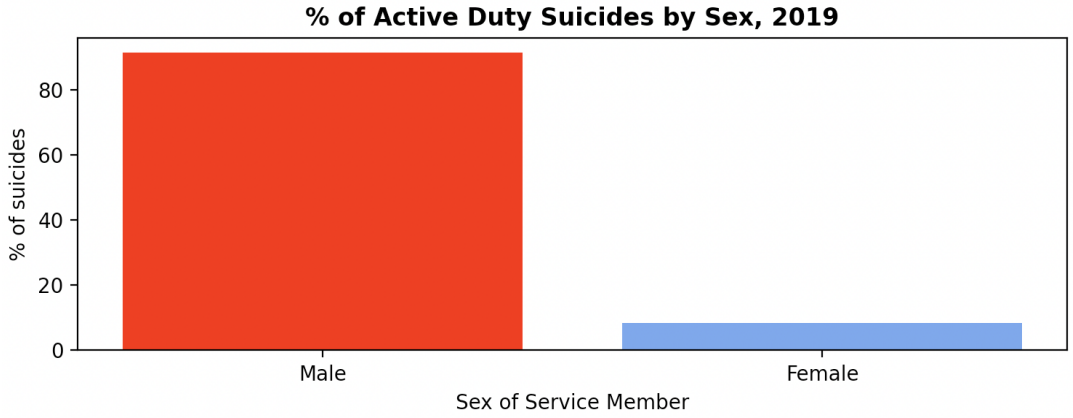


Figure 4 Active Duty Suicides by Sex

Assessment of Work Completed

Significant work was completed in assessing various data sources and how they can be utilized for simply capturing necessary data that can assist providers in making quick assessments. This data will continue to be developed in complexity and usefulness. The prototyping of the GUI allows for continued work in providing a provider board that is easy to navigate, and allows for simple addressing of patient risks.

Priorities of work became an issue with a significantly shorter and less committed timeline then would normally be seen in a project of this scale. Creating both meaningful data along with an interface that could provide insight allowed for a 'divide and conquer' approach to this project.

Successes

This application successfully populates plots using different data depending on the military component of a given servicemember. It offers service providers two distinct views of the

servicemembers demographic information: written in a list, and represented in the context of relative suicide risk. When this project was started the team did not know risk could be simply displayed. The solution to highlight the soldier's data on bar graphs proved to be very effective.

Each member of the development team brought a unique perspective to the project which allowed for strengths to be exploited and weaknesses minimized.

Improvements

This application uses data that was entered by hand into csv files. As this project grows, it will be necessary to import larger quantities of data to inform service providers about the risks their patients face. While the approach to creating a visual display of that data and highlighting where the patient fits into it can be reused, the process for importing and cleaning data cannot be reused.

This application also was able to successfully import information from PDFs, and CSV files which in turn was parsed and distributed throughout the program as needed. This process will need to be refined and further understood to appropriately utilize data through a knowledge framework to override redundancy that is found in many DOD datasets. It will also become imperative to clearly define and articulate data as branches and components are added.

Further Work

This application is designed specifically for a roster of soldiers that does not include veterans. The data is prepared to create a display similar to this one with relevant risk factors for veterans; however, the creation of that display has not yet been implemented.

Further work on this application will include compiling JSON totally into the interface, creating dataframes utilizing multiple branch data, applying more variables into the dataset, and implementing a knowledge management tool such as an ontology to manage data relation.

References

DOD. (2018). Annual Suicide Report. https://www.dspo.mil/ASR/

DOD. (2019). Annual Suicide Report. https://www.dspo.mil/ASR/

DOD. (2020). Annual Suicide Report. https://www.dspo.mil/ASR/

Ravindran, C., Morley, S., Stephens, B., Stanley, I., Reger, M. (2020). Original Investigation| PsychiatryAssociation of Suicide Risk With Transition to Civilian LifeAmong US Military Service Members AMA Network Open.2020;3(9):e2016261. doi:10.1001/jamanetworkopen.2020.16261 Retrieved from file:///tmp/mozilla_denise0/ravindran_2020_oi_200607_1599159695.28426.pdf