

# Final Plan

---

Yuying Pan ID: 2850171

## Description of my proposed research:

Since I am provided with much information about different people's death causes and their personal information, I would like to dig into the relationship between those personal information and their causes of death. So that I can give them my suggestions to keep fit and avoid those causes.

## My work plan:

### 1. Data processing:

After downloading the data sets for year 2013 to year 2019, I found that the data sets on CDC have different forms. Some of them are .txt files, others are .DUSMCPUB files. So the first thing I plan to do is to convert these data sets into a uniform format that can be easily read in Rstudio. According to the recommended books of this course, I prefer to use .csv files for this project. As for the method of converting data sets, I haven't decided which one to be the most efficient one yet, however, I tend to learn python myself and use python to convert data sets into .csv files. Then I would prefer to integrate the .csv files for year 2013 to year 2019 into a data frame which contains all the information I need.

### 2. Hypothesis testing:

After data processing, I would consider my aim for this project. I prefer to figure out the relationship between month of death, sex, age, states, and causes of death.

### 3. Conclusion drawing:

According to the results I obtained in hypothesis testing, draw some conclusions and provide some suggestions for people in different states/ in different sex/ at different ages to pay attention to their health and thus may help them avoid their most likely cause of death.