The project is an opportunity for you to select a health-related problem (broadly defined) and a dataset and to apply your knowledge in the form of a scientific paper.

**Introduction**

Formulate a health-related question and describe its components (population, intervention (or exposure), comparison group, and outcome).

Give some background about what is currently known about this question, as well as some gaps in knowledge.

Motivate the selection of this question with respect to the “triple aims” to justify why it is important.

**Methods**

Name and describe the data set you are planning to use to answer the question.

Note: If you want to look at a question whose data is derived from different datasets (e.g., percentage of people with insurance in a state and mental health in a state) but can be joined together (e.g., using the “state” variable), then describe both datasets.

Where can it be downloaded or accessed?

What is its primary or intended use case?

What are some strengths and weaknesses of this data set with respect to its ability to answer your research question? Consider both the way in which the data elements were collected (i.e., the study design used in this study) and the data elements themselves.

What is the exposure/intervention variable you plan to look at? What is the outcome variable? Describe both in detail.

Describe your hypothesis. What do you expect to find and why?

Describe two potential confounders that are present in the data set (variables that you think may be associated with both the exposure and the outcome). Why do you think these variables could potentially be confounders?

**Results**

Provide a table containing key characteristics (akin to a Table 1 you would find in a clinical journal). If your unit of measurement is patients, then focus on demographic and relevant medical characteristics. Otherwise, focus on relevant characteristics of your choosing.

**Conclusions**

How would you answer your specific health-related question, taking into account the unstratified and stratified analyses (i.e., confounders)?