PROJECT PROPOSAL FORMAT

For the first part of the project, you should identify the population(s) of interest, state the goal of your research, and describe the sampling procedure. Your proposal should be a text document with the following four sections:

Population(s). Give a precise description of the two populations you will compare or the one population which you will examine for detecting a relationship between two variables. Be as specific as possible. For example, "Trees" is far too vague, whereas "All trees 6 feet or taller growing within the borders of the city of Elk Grove" is both specific and precise.
Your main challenge here is to pick a population that is neither too big nor too small. The population should not be so small that you can actually survey it (for example, the population of your personal coffee mugs), but neither should it be so large that you are unable to obtain a representative sample (for example, all people currently living on Earth).

Goal. State which population parameters you are trying to compare, or else which variables you are testing for correlation or independence. The goals you can choose from are:

- To compare population means by measuring one quantitative variable in two independent samples.
- To compare population proportions by measuring one categorical (yes/no) variable in two independent samples.
- To test for linear correlation between two quantitative variables.
- To test for independence between two categorical variables.

Here are some sample study goals (please think of something different, ideally something that you are personally interested in).

- *Comparing means:* The purpose of this study is to detect a significant difference between the mean girth of an evergreen tree and the mean girth of a deciduous tree in Sacramento.
- *Comparing proportions:* We will attempt to show that the proportion of 2-story houses in East Sacramento is greater than the proportion of 2-story houses in West Sacramento.
- *Testing for correlation:* This is a proposal to test for linear correlation between the length and the width of cars parked in Alhambra Triangle neighborhood.
- *Testing for independence:* This is a proposal to test for independence between the color and the body type of cars parked in Alhambra Triangle neighborhood.

Sampling Method. Describe the sampling frame you will be using for taking your sample. In other words, describe in detail how exactly you will select individuals for measurement: will you be picking random items from a large list? And if so, then how will you make that list? Or will you be measuring convenient individuals at randomly selected times and locations? Recall our discussion of multistage sampling, and identify which sampling techniques you will be using.
Next, imagine yourself performing the data collection, and describe in minute detail what you will do and how. State the intended/expected sample size. The main idea behind this section is

to provide a description of a procedure so detailed, that anyone can repeat your experiment just the way you conducted it. Another reason to be thorough here is to lay bare all the limitations of your chosen sampling technique.

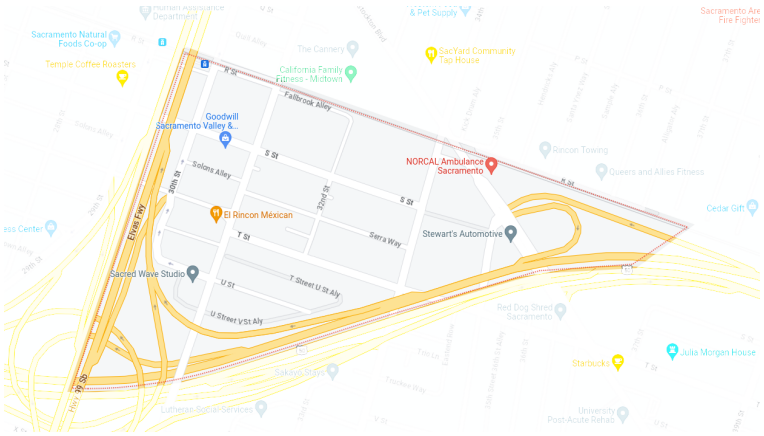
Discussion. Provide constructive criticism of your proposal. In other words, enumerate the flaws which are present by design, starting with possible sources of bias in your sampling procedure. There is no shame in finding flaws: almost every time you can blame them on either the lack of resources or an ethical dilemma or whatnot. There is, however, a great amount of shame in being willfully blind to the consequences of your own choices. Think of it this way: if you don't point out the built-in flaws now, someone will surely bring them up later in an attempt to discredit your work.

EXAMPLE: TEST FOR CORRELATION PROPOSAL

Population. Every 4-wheeled motorized vehicle parked on a public street in the Alhambra Triangle neighborhood of Sacramento will be considered for this study.

Goal. This is a proposal to test for linear correlation between the length and the width of cars parked in Alhambra Triangle neighborhood.

Sampling Method. Since we cannot predict which cars will be parked in the testing area during the sample collecting stage, we will have to use a multi-stage approach, taking a cluster sample of streets in Alhambra triangle, and then taking a systematic sample of parked vehicles on each selected street.
First, we will make a list of all streets in Alhambra Triangle (see the map image below), and choose a simple random sample of 4 streets.
We will walk down each street once, down the odd-numbered side, during a convenient time, and measure every 6th car parked on that side of the street using the systematic sample technique. To pick the first car to be measured, a fair six-sided die will be rolled.
We will measure the dimensions of a car by laying a measuring tape on the ground about 1 foot away from the car. Two measurements will be taken: the length from bumper to bumper and the width.
The sample size cannot be known in advance, but we expect to sample at least 5 cars on most streets, giving us a good chance of $n > 20$.



Map generated by Google

Discussion.

(a) Bias is expected due to choosing convenient times for sampling; a better approach would have been to randomize those times.
(b) Bias is expected due to using streets as clusters.
(c) Bias is possible, but unlikely due to the systematic stage.
(d) If all selected streets are very short, our sample size may end up being too small for a meaningful result.
(e) We expect to detect a significant positive correlation, since we expect larger cars to have greater width and length; the body style of a car (sedan, coupe, etc.) may well be the lurking variable in this case.