Lab 1: Let's collect some data!

This lab will guide you and your group through the process of designing and collecting some data on your fellow students at UMass. The main goals of this lab are to get you to think carefully about what the process of moving from a specific question, to data collection, to preparing the data for analysis.

Deliverable 1: (30 points) The lab writeup and any slides that you will present are due on Wednesday, September 12 at 5pm. The file(s) should be sent to the TA Bianca Doone via Piazza. Each group will give a 3-5 minute presentation (strict limit!) on Thursday September 13th summarizing their hypothesis, the data collection process, and qualitative results. Since we have not covered yet how to efficiently summarize or display your data, it is not required for this presentation. Although you should give us some basic quantitative information about your process: how long did you collect for? how much data did you collect? ... Tell us a complete (but brief) story about what data you collected and why. I encourage each team to use 1-2 slides to describe the study.

Deliverable 2: (20 points) Each student will write and submit a 1/2 page critique of the group project. This must be submitted by 5pm on Friday, September 14. Before submitting your files, create a folder on Google Drive with the name "[lastname]-[firstname]-pubhlth460" (e.g. "Reich-Nicholas-pubhlth460") and share it with the TA assigned to your table. To submit your file, just upload it into your shared Google Drive folder. This write-up should describe at least 2 key weaknesses of the experiment you conducted and address something that you would do differently if you were to do the same experiment again.

Questions to answer on your own (not for handing in)

Exercise 1 Look around the room at your classmates. What characteristics and features can you reliably and quickly identify about them?

Exercise 2 Pick three of these features, and sketch out below a table and collect some data on a few of your classmates. Note that some data that may seem "easy" to collect, may not be so straightforward. In particular, collecting data on race, ethnicity, or gender may appear at first glance to be something straightforward to identify, but in reality it could get you into situations that could potentially misrepresent someone's identify and possibly offend or disrespect them. The best way to avoid these pitfalls is to interview people directly using standardized survey methods designed for the purpose of eliciting this information. (E.g. for details on asking questions about gender identity, see, e.g. this presentation.) However, this exercise and the activities later on in this lab, have in the past mostly involved collecting data non-intrusively, i.e. just observing an environment and the people in it. This could make it challenging to assess an individual's racial or gender identity with certainty. If you or your group chooses to assess one of these characteristics, please do so in a way that conforms to accepted scientific classifications and that acknowledges biases or misclassifications that may arise from not directly interviewing individuals.

Exercise 3 We will talk in future classes about what "tidy data" means. In brief, it is a way of organizing data where each row represents a single observation (e.g. one row per student you observed) and each column represents a variable that you are making an observation on (e.g. hair color, eye color, etc...). Take a minute and reflect on the choices that you made when creating the table in the previous exercise. What defines a row and column in your small dataset? What types of variables were you collecting and how did you encode them in your table?

Questions to answer with your group (not for handing in)

Exercise 4 Give each member of the group 1-2 minutes to present the data they collected and how they organized their table. Discuss similarities and differences between how you each collected data. Can you reach a consensus on the "best" way (most clear, simplest, easiest to enter data into, most easily imported for analyses) to store data like this?

Exercise 5 As a group, brainstorm different settings on campus where you might expect the composition of students or groups of students (according to features that you think you can measure accurately) will be different. Write down a few below.

Exercise 6 Take the list you just wrote down and turn it into a series of hypotheses, with a proposed mechanism. For example, "we hypothesize that group A is different from group B because XX.".

Exercise 7 Plan out, as a group, how you will collect data to test one or two of the hypotheses you define above. This should include logistics (when and where will each member collect data? what technology [if any] will you use to collect data? what data will they collect?) and technical details (how will you create a dataset that you all can work from and enter data into?).

Things to hand in and present

Each group should prepare a brief written report (no more than 2 pages) describing their study. The report should include a description of the hypothesis and the data collection protocol. Figures or diagrams may be included for illustrative purposes. Additionally, include a description of the data collected as well as a proposal for how this data could be used to evaluate the hypothesis. Note that it is not required to actually do any formal statistical analyses here, just to discuss how you could do them.