

Some Guidance for Learners New to the Specialization

If this is your first course from the specialization, and you have not yet chosen a data set to work with, you should do that first. Data set descriptions and codebooks can be found in Week 2 under Data Sets and Codebooks (<https://www.coursera.org/learn/regression-modeling-practice/supplement/sLhJw/course-codebooks-and-data>). You are also welcome to use your own data set or a data set from a different source if you're comfortable doing that.

You can check out the data sets codebooks to learn more about the studies, and identify a basic research question that interests you (for example, "Is there an association between depression and binge drinking in adolescents?"). Once you have a basic research question, you can think about other variables that might explain your association. For example, what else might explain why an adolescent has depression and/or binge drinks? I might want to see whether there are variables that ask about peer drinking, parent drinking, other substance use, and so on.

The assignment asks you to provide a general description of where the data came from and how it was collected (sample data and collection procedure). Once you've chosen a data set, you can find the information for this by looking at the study's home web site, which can be found by googling the name of the study (for example, Gapminder, NESARC, AddHealth, Outlook on Life Survey). The codebooks have information on the variables (what they measure, how they are coded, etc).

If you have not yet done any data management because you haven't taken one of the earlier courses in the specialization, you can describe what you intend to do with your variables. For example, do I want to look at how often adolescents binge drink, or do I want to create a binary variables (yes/no) for whether adolescents binge drank in the past month? Do I want depression to be a quantitative variable that aggregates multiple questions that ask about depression symptoms, or do I want to use a cutoff to create a categorical variable that groups adolescents into low, medium, or high levels of depression? These decisions should be described when you describe the variables you will be using.

As you continue in the course you may decide that you want to change how you measure a variable, or you might want to add more variables or even change your research question if you feel it's necessary. For new folks especially, the assignment can be thought of as a game plan to get you started, not a contract that requires you to do exactly what you said you were going to do in Assignment 1. However, you will want to personally document any changes you make because being able to explain your process is a necessary skill for anyone doing data analysis and you may also need to explain it at some point during the course.