Project Stage 1

1. Identify the important research question(s) which will guide your project (e.g. What factors are associated with lower BMI?) – and describe why your chosen project is interesting to you. Provide rationale for each variable included in your proposed data set (e.g. Does taking a PE class help? We may subgroup by gender. We need to control for diet.)

2. Find references for at least two articles in the refereed literature that are relevant to your question of interest. You should avoid articles that are too technical to be relevant to the project or to be informative for the non-specialist (e.g. "Beta 2-adrenergic receptor polymorphisms and haplotypes are associated with airways hyperresponsiveness among nonsmoking men," Chest, 2004). Articles that appear in the popular press (such as The New York Times or Washington Post) or news articles in journals are not acceptable as refereed references, although they may help motivate ideas for your project. Be sure you obtain the entire paper and not just an abstract! You will eventually use these references in the introduction of your paper. Pay close attention to the figures, tables, and methods sections of the papers you select as they can give you an idea of what I'll be expecting from you in your final write up.
   For the project proposal, include the following information:

   - Give the citation for each reference (in APA format or similar) and a link, if appropriate.
   - In 1-2 paragraphs, summarize the primary findings and how they relate to your proposal.

3. Complete a variable chart (similar to the one that follows) for your anticipated variables. A typical list will include 6-10 variables. List the variable name, variable role if known (response, explanatory, potential confounder), an indication of whether or not the variable is quantitative or categorical, the range of values for each variable, and the units of measurement for each variable (if appropriate). For any variable whose definition is unclear, provide a short definition. Also mention your observational units; if your data has multilevel structure, provide the level for each variable as well. As an example, if body mass index (BMI) were the response variable, attending a physical exercise class daily is the explanatory variable, and age is a potential confounder, the first few lines of the chart might read:

4. Outline how you plan to address your research question(s) with the data you have listed in (3) (e.g. We plan to run multiple linear regression models with BMI as response and daily attendance and age as explanatory variables, possibly examining interactions between the two).

5. Describe how you obtained your data, providing a link if appropriate. Store your data, properly labeled, as a .csv file in the Project folder of our class folder on the RStudio server.

We accessed our data from the Dallas Open Data website. This website contains many public datasets, including animal shelter records, which anyone can request an access key to and then download. We installed the data, selected the variables of interest to us, and created a csv file in the CreateDataset.Rmd file. Our dataset is saved as adoptions.csv.