

# Project Stage 1

*Isaac Slagel, Jack Welsh, Va Yang*

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.)

What factors influence how long an animal will stay in the humane shelter? Are certain breeds euthanized at a higher rate than others? Are certain animals adopted more during certain times of the year (ex. are more rabbits adopted around easter?). What differences do we see between the conditions of dogs and cats in the humane society? This project is of interest to us because we love animals and are intrigued about how different animals are handled in the humane shelter system. The dataset we found to look into this made us even more motivated to look into these questions as it is detailed and clean!

The variables `animal_breed`, `animal_origin`, `animal_type`, and `chip_status` are all important variables to our observational units. We may end up subtyping by any of these variables. We can glean insights on how animals fare in the animal shelters by the variables `intake_condition`, `intake_date`, `intake_subtype`, `intake_type`, `outcome_condition`, `outcome_date`, `outcome_subtype`, `outcome_type`. We will use these variables to create different response metrics that we are interested in (like total shelter days or euthanized).

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.

1. Lepper, M., Kass, P. H., & Hart, L. A. (2002). Prediction of adoption versus euthanasia among dogs and cats in a California animal shelter. *Journal of Applied Animal Welfare Science*, 5(1), 29-42.
2. Posage, J. M., Bartlett, P. C., & Thomas, D. K. (1998). Determining factors for successful adoption of dogs from an animal shelter. *Journal of the American Veterinary Medical Association*, 213(4), 478-482.
3. Lampe, R., & Witte, T. H. (2015). Speed of dog adoption: Impact of online photo traits. *Journal of applied animal welfare science*, 18(4), 343-354.

The first paper (Lepper et al) had a panel of individuals record features of a young or old Labrador retriever’s photo at an animal shelter. The goal of this study was to pilot research into animal advertising and adoption rates to help shelters effectively improve how animals are taken out of the system. Not surprisingly, they found that dogs with poor images got adopted at a lower rate. However, this phenomenon was more pronounced for older dogs. While we have no information on how these dogs are marketed, our research would look into which animals have the lowest odds of adoptions. Thus shelters could use their resources to market these dogs more and get them out of the shelter.

The second study, Posage et al, much like the first one, looked into what characteristics influence a dog's chance of being adopted using data from a local humane society. They discovered certain characteristics of dogs which are associated with higher rates of adoption. The characteristics include things such as coat color (gold) and breed (Terrier). Our data has information on breeds, so this may be something we look at as well.

The third study, Lampe and Whitte, looked into animal shelter data from Sacramento County California. The authors are interested in what characteristics of dogs and cats influenced those animals' rates of adoption. They were able to conclude on 5 variables that influenced a cat's odds of adoption and 7 variables that influenced a dog's odds of adoption using logistic regression. While we are not totally confident that the authors of this paper employed good statistical practices, they went about answering some similar questions to what we are planning to do in our research.

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:

The observational units in this data are individual animals which have ended up in an animal shelter. There may be some multilevel data with correlation between observations at each shelter (we still need to do some work with this variable).

Variable Name	Variable Role	Variable Type	Range of Values	Units of Measurement
animal_breed	explanatory	categorical	296 unique breeds	NA
animal_origin	explanatory	categorical	4 sources of shelter animals	NA
animal_type	explanatory	categorical	5 species of animal	NA
chip_status	explanatory	binary	(0,1)	NA
intake_type	potential confounder	categorical	how animal came to be at the shelter	NA
outcome_type	response	categorical	how animals was removed from shelter	NA
intake_condition	potential confounder	categorical	keyword description of animal status at intake	NA
outcome_condition	potential confounder	categorical	keyword description of animal status at outcome	NA
intake_date	response	date	(2017-10-01, 2019-04-03)	y-m-d
outcome_date	response	date	(2017-10-01, 2019-04-03)	y-m-d

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).

We plan to use Poisson regression to investigate the factors that relate to how long an animal will at the humane shelter. We will use logistic or binomial regression to investigate what factors influence the euthanization rates. To look at how the rates of adoption differ throughout the year, we plan on using poisson regression with a categorical variable to look at the month. To look at the differences between cats and dogs, we plan to do poisson regression on just cats and just dogs. Then we will compare the coefficients between the two. We think that there may be some problems with the independence condition for our models because we believe that the outcome for an animal in the shelter may depend on what

other animals are at the shelter. We are unsure what type of mixed affects model would be most appropriate to account for this kind of correlation.

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. Our work is shared on Github.