**Graduate Project Proposal**

Kathryn Clark

Some datasets that I find interesting are from the Austin Animal Center, specifically their outcomes and intake data. The data is available from data.austintexas.gov. The Austin Animal Center is the largest no-kill animal shelter in the US and these datasets provides information on the animals they care for; the date, reason, and status of each animal upon intake; and the ultimate outcome (e.g. adopted, transferred, missing, etc) of each animal and the date when that happened. Animals are tagged with a unique ID that allows the two datasets to be joined together. Animal adoption from shelters is important to not overburden shelters (so they can afford care for the animals they have and so they have space for additional animals as needed). However, while some animals are adopted out immediately, some stay at shelters for a long time. For example, it is theorized that a phenomenon called “black dog syndrome” exists. In this phenomenon, black dogs are passed over at shelters, with lighter dogs chosen instead.

I want to use these datasets to better understand the mix of animals at a shelter (with data exploration and visualization) and then create machine learning models to predict ultimate outcome and, if an animal was adopted, the length of time from intake to adoption based on the characteristics of the animal. The ultimate goal of this is to understand what characteristics help/hurt an animal’s chance of adoption.

The datasets themselves don’t include a large number of features, so this will require some feature engineering. Additionally, I don’t yet have a sense of what the best models will be for this data and this problem, so I intend to try several and find the best.