Research Proposal

Dataset

The data used is the animal shelter outcome data for City of Austin Animal Center from October 1, 2013 to December 7, 2017. The data is from data.world. The dataset contains 13 variables: animal's ID, name, birth date, type, sex age, breed and color; outcome's type, sub type and date.

Research Question

The research question is what variables affect the adoption chance of a sheltered animal, and what is the coefficient of these variables. In particular, I have the hypotheses that adoption chance is affected by animals' age, sex, type, as well as adoption date (people are more likely to adopt animals in more modern days). It is an interesting research question because animals are cute. Also animal adoption have important ethical implications, and can affect multiple disciplines such as sociology, psychology and public health.

The method will be a logistic model through Bayesian regression using R and stan. A detailed draft model is appended after the EDA.

Data Cleaning

- There are animals that has been intaken multiple times. 63388 animals have been intaken uniquely 1 time, 4584 animals have been intaken 2 times, 760 animals have been intaken 3 times, 175 animals have been intaken 4 times, 92 animals have been intaken 5 times or more. For the multiple intakes animals, only the first intake is kept.
- This report will focus on cats and dogs, as they are the more common pets for households, the dataset is then filtered to be only cats and dog.
- Animals' sex will be a candidate predictor of the model, and rows with unknown or null sex are dropped.
- These leaves us with 61778 rows of data.
- There are 2101 unique breed values and 519 unique color values, but many are correlated or have some level of subjectivity. For example breed "Pit Bull" and "Pit Bull Mix", or color "Black/Brown" and "Black/Tan". Due to the sizes, correlations and potential interpretation difficulties, I decided to not include these two variables.
- This means that even when two animals are both dog, their expected outcome may differ due to the unincluded breeds and colors.

- So instead, I decided to capture the variability of different breeds and colors using a heretical structure for the coefficient β_{type} of the factor variable animal type:

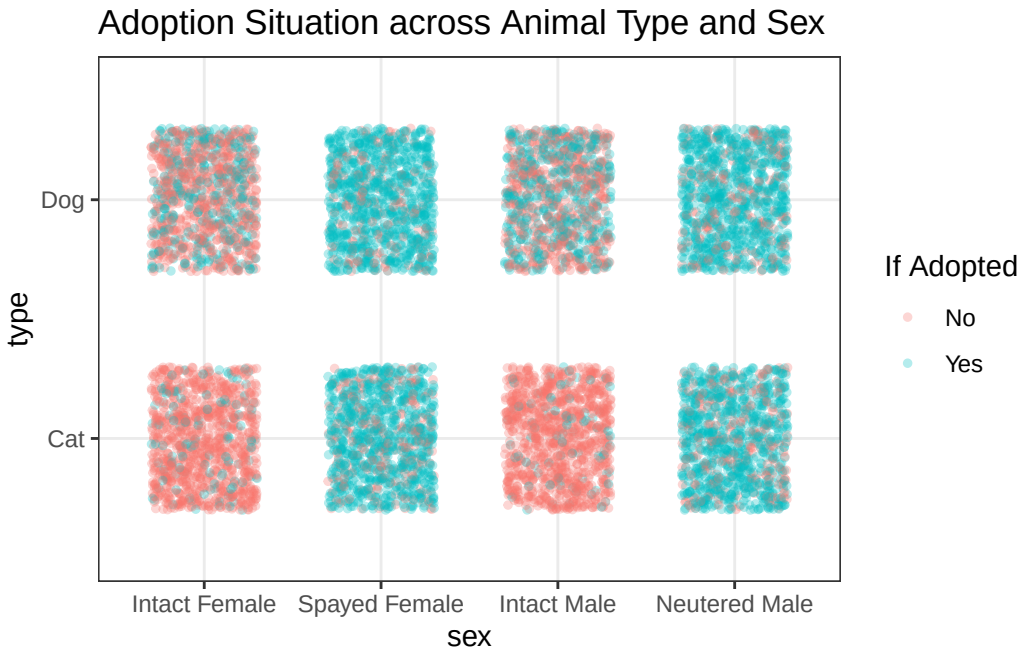
$$logit(\pi) = \beta_0 + \beta_{type} + ...$$

- β_{type} will be heretically structured and drew from $N(\mu_{cat}, \sigma_{cat})$ for cat type, and drew from $N(\mu_{dog}, \sigma_{dog})$ for dog type animals.
- For animals' age, variable "age_upon_outcome" is the animals age in either years, months or days; and variable "age" is the age of the animal all converted in days, which is more accurate and is the one kept.
- As for the outcome of the animals, unique outcome_type are as follows:

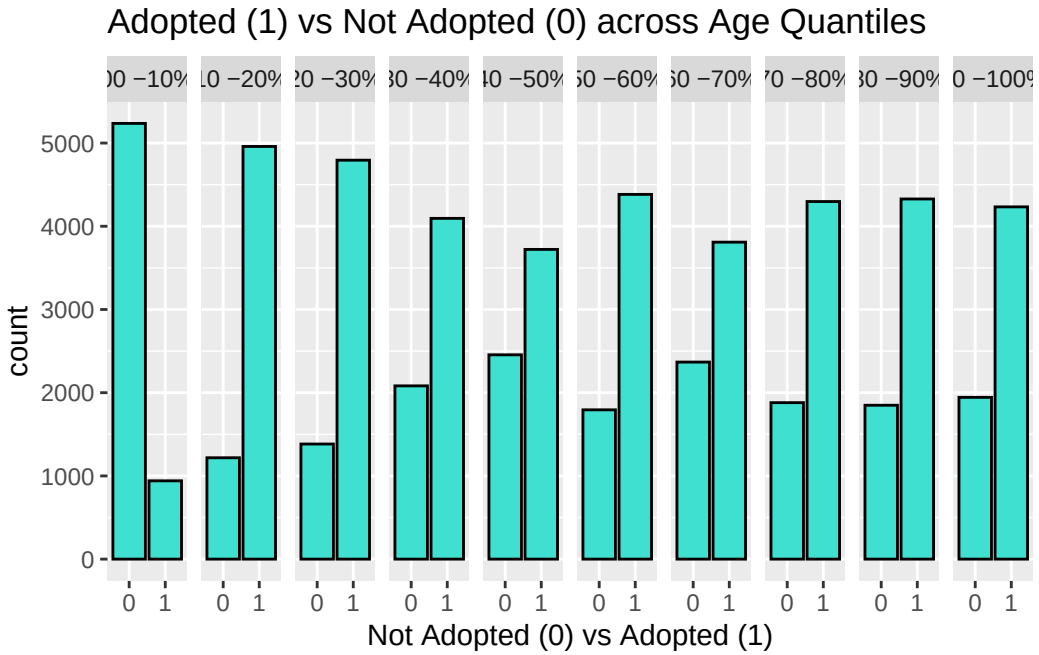
```
[1] "Return to Owner" "Euthanasia"      "Adoption"        "Transfer"
[5] "Died"            "Missing"         "Rto-Adopt"       "Disposal"
[9] NA
```

- 2 rows of NA are dropped. Then values "Return to Owner", "Adoption" and "Rto-Adopt" (return to owner - adopt) are merged into value 1 (adopted), and the rest of the values are merged into 0 (not adopted).
- outcome_subtype are most missing values as shown from the data skim at the beginning, and therefore is dropped.
- Animal ID and name are then dropped as they will not be used as predictors in the model. Data of Birth is also dropped as it is collinear with age.
- As for the candidate predictor adoption date, since there are only 4 years (2013 - 2017) of data, I decided to index the time by month and code month 2013/10 to 2017/12 to be month 1 to 51, there are no missing months in between.
- The final dataset contains animals' type, sex and age and the corresponding outcome and outcome date.

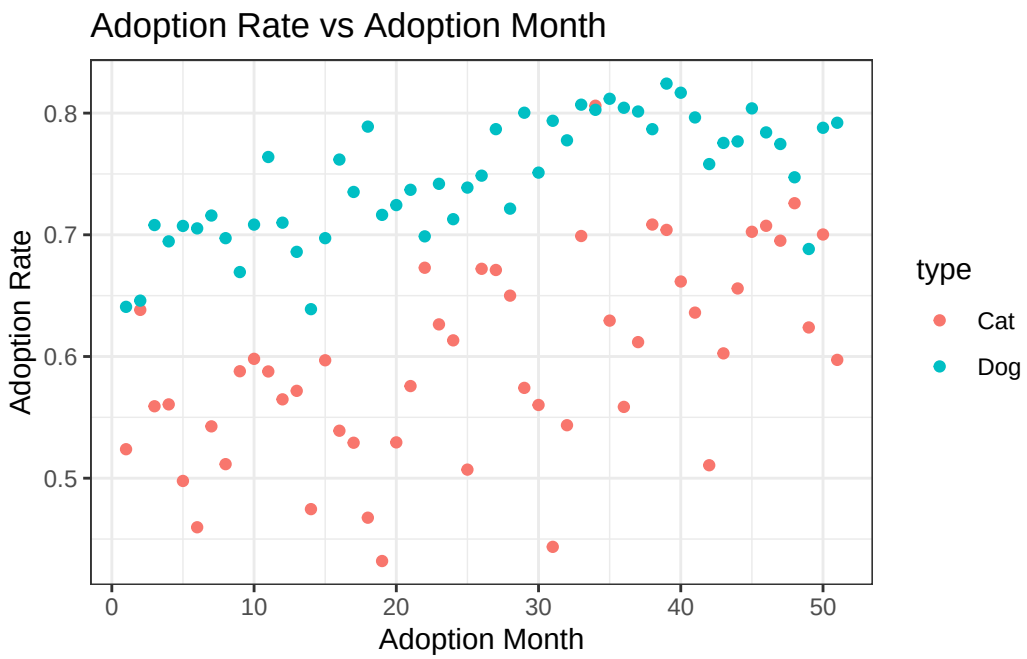
EDA



- Spaying/Neutering greatly increases the chance of an animal getting adopted.
- When not spayed/neutered, dogs are more likely to be adopted than cats. When spayed/neutered, the adoption rate is not visually different for cats and dogs.



- For animals with extremely young ages (the youngest 10%), the adoption rate is very low.
- For the rest of the ages, the adoption rate starts high for young animals, then gradually decreases as animals' age increase, until reaching the 50% age quantile, when the adoption rate starts to remain about constant.
- Base on the plot, the youngest 10% animals is excluded from the model data, resulting in 5.5k + rows of data in the dataset.



- The adoption rate goes up as month increases.
- Dogs have higher adoption rates than cats at almost all month values.

Appendix: A Draft Model

$$Logit(\pi_i) = \beta_0 + \beta_{type,i} + \beta_{sex,i} + \beta_{age} \cdot age_i + \beta_{month} \cdot month_i$$

Where type can be either cat or dog, and:

$$\beta_{type,i} \sim N(\mu_{cat}, \sigma_{cat}), \text{ for cat}$$

$$\beta_{type,i} \sim N(\mu_{dog}, \sigma_{dog}), \text{ for dog}$$

Similarly, sex can be intact female, spayed female, intact male or neutered male, and: for intact female or spayed female,

$$\beta_{sex,i} \sim N(\mu_{female}, \sigma_{female})$$

for intact male or neutered male,

$$\beta_{sex,i} \sim N(\mu_{male}, \sigma_{male})$$

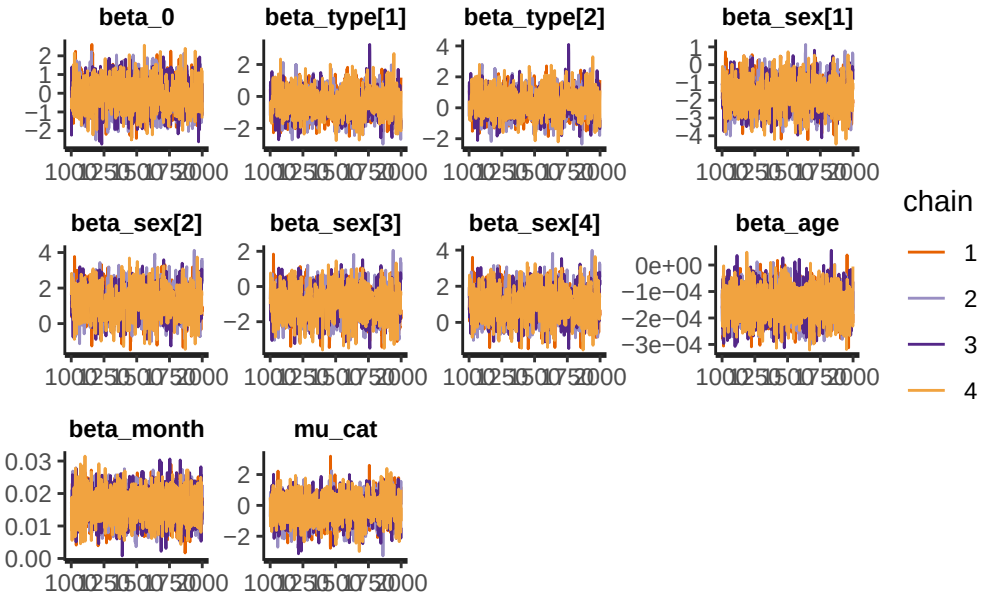
- I don't have the technical knowledge about if spayed for female and neutered for male are equivalent and can be considered to have the same effect, so I decided to leave them within the sex factor instead of extract them out as a separate factor of intact or not.

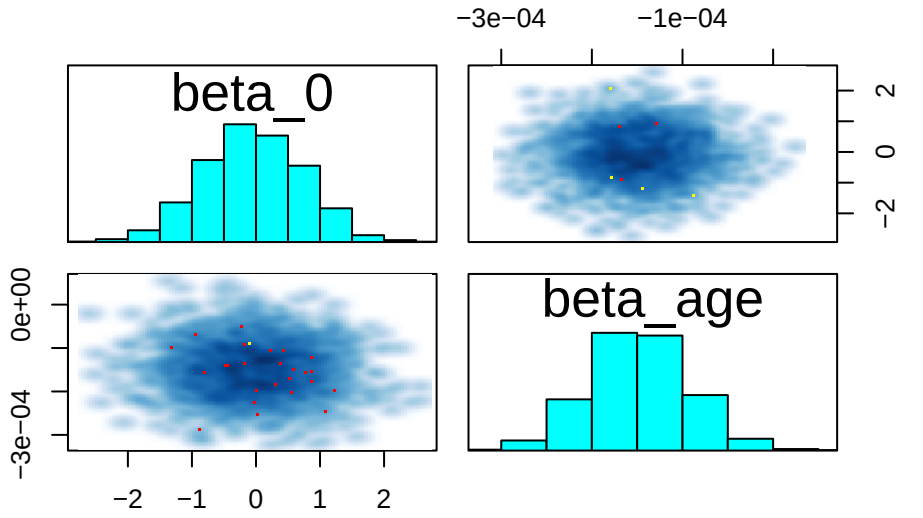
- The coefficients are as follows:

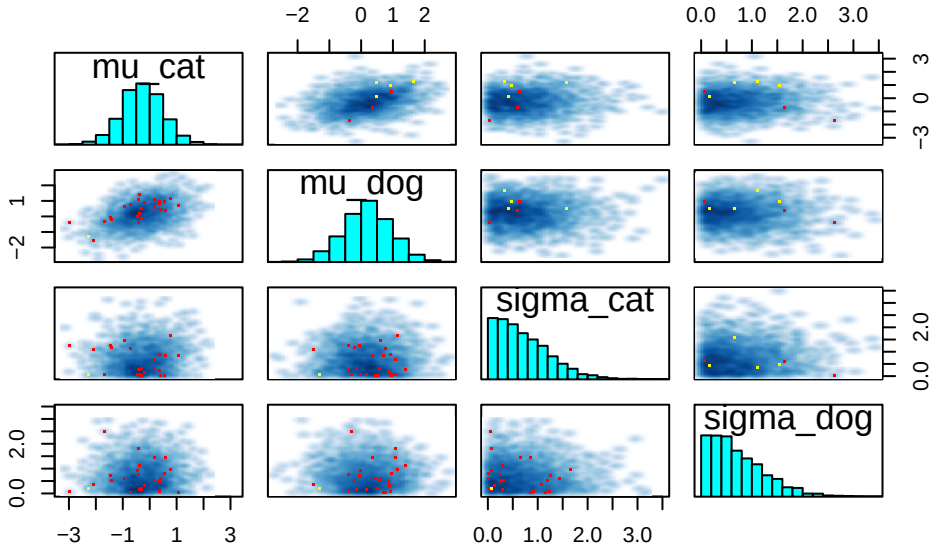
	mean	se_mean	sd	2.5%
beta_0	-0.0541208258	2.294831e-02	7.833954e-01	-1.5922792434
beta_type[1]	-0.3944461990	2.500882e-02	7.609838e-01	-1.9127530726
beta_type[2]	0.3029035144	2.535147e-02	7.621276e-01	-1.2158419271
beta_sex[1]	-1.6703397182	2.909026e-02	8.403448e-01	-3.3513184686
beta_sex[2]	1.2103520956	2.887459e-02	8.250833e-01	-0.4080554637
beta_sex[3]	-0.9583712276	2.905979e-02	8.296223e-01	-2.6159687507
beta_sex[4]	1.0348367920	2.886793e-02	8.229989e-01	-0.6137136336
beta_age	-0.0001486257	8.774927e-07	5.492762e-05	-0.0002547319
beta_month	0.0164042223	6.425509e-05	4.314881e-03	0.0076584978
mu_cat	-0.2685900704	1.948813e-02	7.753335e-01	-1.8228548049
mu_dog	0.1952430425	2.016318e-02	7.689392e-01	-1.3366654511
mu_female	-0.1141651900	2.091990e-02	8.279888e-01	-1.6959023306
mu_male	0.0138180489	1.982647e-02	7.920609e-01	-1.6006593092
sigma_cat	0.7191674072	1.193039e-02	5.373428e-01	0.0343981059
sigma_dog	0.7249580153	1.112530e-02	5.449204e-01	0.0433528798
sigma_female	1.4555420665	8.583261e-03	4.728266e-01	0.7377108364
sigma_male	1.2231627729	8.690518e-03	4.608709e-01	0.5577663658

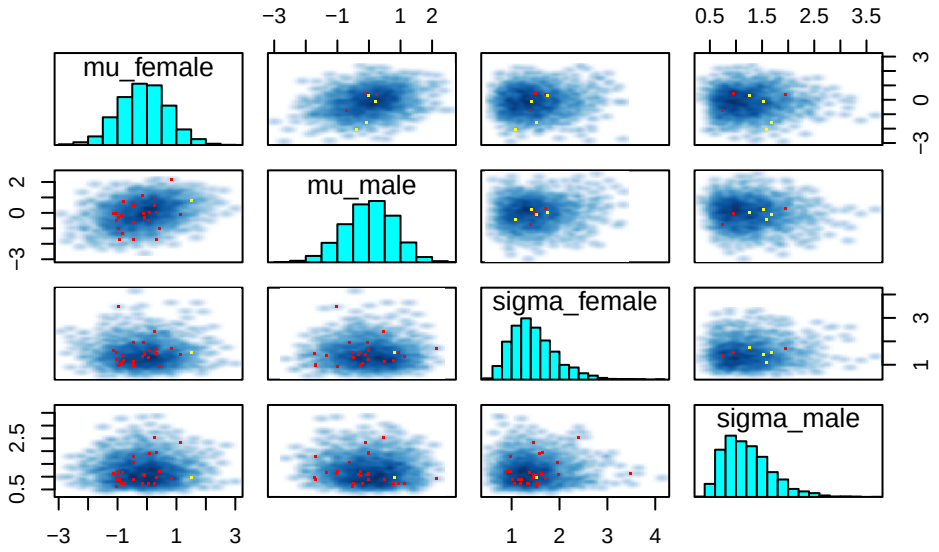
	25%	50%	75%	97.5%	n_eff
beta_0	-0.5790098703	-0.0570649499	0.4945410165	1.418613e+00	1165.3614
beta_type[1]	-0.8729171461	-0.4065753943	0.0803728664	1.157343e+00	925.9005
beta_type[2]	-0.1725955499	0.2826142116	0.7820438784	1.852165e+00	903.7514
beta_sex[1]	-2.2334237453	-1.6592757635	-1.1066934629	-3.729996e-02	834.4873
beta_sex[2]	0.6478584786	1.2284529389	1.7572066670	2.797131e+00	816.5145
beta_sex[3]	-1.5265513486	-0.9371655556	-0.4072848017	6.640335e-01	815.0343
beta_sex[4]	0.4671695794	1.0500111650	1.5718021288	2.650187e+00	812.7691
beta_age	-0.0001862356	-0.0001490774	-0.0001122081	-4.048146e-05	3918.2713
beta_month	0.0135886446	0.0164389325	0.0191800148	2.483215e-02	4509.4387
mu_cat	-0.7727609377	-0.2725398371	0.2549375062	1.244685e+00	1582.8391
mu_dog	-0.3066871493	0.2131168686	0.7076852307	1.684228e+00	1454.3395
mu_female	-0.6827180914	-0.1011752615	0.4410367875	1.534117e+00	1566.4983
mu_male	-0.5126843203	0.0427595815	0.5724953669	1.504678e+00	1595.9763
sigma_cat	0.2938946128	0.6143451408	1.0416727065	2.006604e+00	2028.5882
sigma_dog	0.2929978245	0.6014074278	1.0416375558	2.001546e+00	2399.0695

```
sigma_female  1.1159671912  1.3868343042  1.7197733979  2.563761e+00 3034.5840
sigma_male    0.8759479963  1.1588087170  1.4875886665  2.314225e+00 2812.3357
                    Rhat
beta_0         1.0031711
beta_type[1]   1.0035752
beta_type[2]   1.0035248
beta_sex[1]    1.0042619
beta_sex[2]    1.0037383
beta_sex[3]    1.0042099
beta_sex[4]    1.0038274
beta_age       1.0004926
beta_month     0.9997796
mu_cat         1.0006366
mu_dog         1.0035245
mu_female      1.0018832
mu_male        1.0011900
sigma_cat      1.0001032
sigma_dog      1.0004009
sigma_female   1.0000777
sigma_male     1.0008781
```




beta_0
beta_age
−3e−04
−1e−04
0e+00
−3e−04
−2
−1
0
1
2


mu_cat
mu_dog
sigma_cat
sigma_dog
−2
0
1
2
0.0
1.0
2.0
3.0
−3
−1
1
3


mu_female
mu_male
sigma_female
sigma_male
−3
−1
1
2
0.5
1.5
2.5
3.5
3
0
−3
2
0
−3
3
1
2.5
0.5
−3
−1
1
3
1
2
3
4