

# Stage 2: Exploratory Data Analysis

Isaac Slagel and Jack Welsh

4/18/2019

## Main Report

### Pitbulls in Dallas Shelters

In our EDA we found a distinct difference between outcomes for pitbulls compared to other breeds of dogs. What we saw was that pitbulls tend to be killed at a higher rate than other dogs. However we also found there were other variables that influence the outcome of dogs like chip status, outcome month, and intake condition. In order to quantify the difference in treatment between dog breeds, we condensed our large dataset into a smaller form and conducted a series of binomial regressions described in Table 1.

Table 1: Modeling Dog Outcomes in Dallas Animal Shelters

	<i>Dependent variable:</i>		
	Proportion of dogs who died		
	(1)	(2)	(3)
Intercept	0.116*** (0.069,0.186)	0.111*** (0.069,0.170)	0.552*** (0.454,0.669)
Pitbull	3.440*** (1.795,6.557)	3.424*** (1.905,6.130)	3.489*** (3.022,4.027)
Scannable Chip	0.789 (0.377,1.566)	0.799 (0.412,1.483)	0.781*** (0.667,0.911)
Summer Outcome	1.461 (0.725,2.852)	1.447 (0.771,2.649)	1.478*** (1.271,1.718)
Contagious		7.286** (1.324,44.137)	3.975*** (2.568,6.168)
Treatable At Intake			0.161*** (0.133,0.196)
Overdispersion Parameter	139.72	111.46	6.27
Nested F Test		$F : 5.1142^*$	$F : 313.62^{***}$
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01	

We see in Table 1 that all of our models have variable estimates that are significant including Model 3 which accounted for whether or not an animal was deemed “treatable” at the time of intake. Further, by two F tests we see each of our models is an improvement on the one before it (Model 1 → Model 2:  $F = 5.1141$ ,  $p\text{-value} = 0.031$ ) (Model 2 → Model 3:  $F = 313.62$ ,  $p\text{-value} < < < .05$ ). Before our final draft we will continue to explore these trends and look for other variables that contribute to dog outcomes.

Using model 3 as our best model of dog outcomes from the animal shelter, we glean several insights about how certain characteristics of dogs have relationships with those animal’s outcome from the animal shelter.

For instance, if a dog is a pitbull, we expect it's odds of leaving the animal shelter dead increase by 348%, after controlling for chip status, season, and intake condition. Yikes! On the other hand, if a dog comes into the shelter with a scannable chip, we expect the animals odds of dying within the shelter system to drop by about 22%, after controlling for pitbull, season, and intake condition. Interestingly enough, if a dog has its outcome in the summer, the odds of that outcome being death are about 47% higher, controlling for breed, chip status, and intake condition.

One interesting issue that we encountered in this section of our modeling was the significance of the dataset we fit on SE estimates. If we summarized our data for each specific model, parameter estimates had increased SEs and more insignificant p-values. However, if we fit all of our models on the same, more extensive, summary table, SEs were small and parameters were more likely to be significant. We decided to proceed with the latter option, as it allows us to carry out nested F tests and would be more like a modeling process for true grouped data. This could mean that our SEs are artificially low.

## Dogs and Cats

Interestingly enough, we were not able to find any significant relationships between dog and cat outcomes. We tried multiple binomial models accounting for stray, summer, and chip status. None of these models showed any significance. This suprised us a lot, as we saw a few trends in our EDA and have heard stories about poorer outcomes for cats in animal shelters. We may consider looking into a few other outcomes for our final report (we only considered death at outcome).

## Controlling for City Council District

To control for differnces between city council districts we chose to use multilevel logistic regresison. We began by fitting two random intercepts models with no level 1 or level 2 terms, one for adopted as the response and one for death within the shelter system as a response.

While looking at adoptions as our response we found that our one fixed effect  $\alpha_0 = -0.4930$ , which if you exponentiate and convert to a proportion we see that the average proportion of dogs being adopted is 0.379. It is sad to think that only about one out of every three dogs gets adopted. We fonud a  $\sigma_u = 0.2438$  so the average deviance between city council districts 0.2438. Our one ficed effect is considered significant, see table 2.

Next we repeated the same model with out\_dead as the outcome. We fonud a fixed effect for our  $\alpha_0 = -1.65465$ , which after you exponentiate and convert to a proportion we get 0.160. So only 16% of the dogs that enter the animal shelter system exit dead. We found a  $\sigma_u = 0.1395$ , so the average deciance between city council districts is 0.1395. This fixed effect was also found to be significant, see table 2.

Table 2: Predicting Dead Outcomes and Adopted outcomes with Multilevel Structure No level 1 Predictors

	<i>Dependent variable:</i>	
	out_dead	adopted
	(1)	(2)
Constant	-1.655*** (0.041)	-0.493*** (0.067)
Between District $\sigma$	0.1395	0.2438
Observations	45,487	45,487
Log Likelihood	-19,759.010	-29,340.940
Akaike Inf. Crit.	39,522.010	58,685.880
Bayesian Inf. Crit.	39,539.460	58,703.330
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

We repeated the same procedure as above just this time adding in summer, chip status and treatable intake to level 1. We found with adopted as the response the average proportion adopted is 0.126 not durring the summer for dogs that are untreatable when they arrive and the chip was not readable or not present. We found that all of our predictors are significant, see table 3. We found that on average the intake during the summer increases the odds of a dog being adopted by 7.4% compared not during the summer, controlling for chip status, and treatable intake. Suprisingly we found that if a dog has a readable chip on average the odds of them being adopted decreases by 11.5% compared to a dog with a unreadable chip or no chip controlling for summer and treatable intake. This suprised us at first, but then we realized that dogs that have a chip tend to be dogs that are returned to owner. Unsuprisingly, a dog with a treatable intake odds of being adopted are 393% higher than dogs without a treatable intake controlling for whether it is summer and chip status.

When looking at death as our response we found that the average proportion of dogs dying in the animal shelter system not durring the summer, without a chip, and with a non treatable intake is 0.52. We found that on average summer increases the odds of death increase by 45.5% controlling for chip status and treatable intake. On average having a chip decreases the odds of death by 21% acconuting for summer and treatable intake. On average treatable intake decreases the odds of death ny 89.3% accounting for summer and chip status.

Table 3: Predicting Dead Outcomes and Adoption Outcomes with Multilevel Structure with Level 1 Predictors

	<i>Dependent variable:</i>	
	out_dead	adopted
	(1)	(2)
Constant	0.083 (0.053)	-1.936*** (0.093)
summer	0.375*** (0.030)	0.072*** (0.023)
chip_status	-0.236*** (0.031)	-0.122*** (0.022)
treatable_intake	-2.121*** (0.036)	1.597*** (0.051)
Between District $\sigma$	0.1408	0.2896
Observations	45,487	45,487
Log Likelihood	-17,995.950	-28,653.040
Akaike Inf. Crit.	36,001.900	57,316.070
Bayesian Inf. Crit.	36,045.530	57,359.700
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

## Annotated Appendix

### Pitbulls

One of our research interests was to look into how different species of dogs fare in the animal shelter system. Specifically, we were interested in why pitbulls are more likely to be dead at the outcome of their time in the shelter. Does this trend continue after controlling for other variables in our dataset?

Initially we fit a model which factors in the chip status, whether or not the outcome was in the summer, and whether the dog was a pitbull or not.

```

pitbull_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0)) %>%
  mutate(chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0)) %>%
  group_by(pitbull, chip_status, summer, month) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())

pitbull_model1_binom <- glm(prop_dead ~ pitbull + chip_status + summer, weights = count, family = binomial)

pitbull_model1_quasi <- glm(prop_dead ~ pitbull + chip_status + summer, weights = count, family = quasipoisson)

summary(pitbull_model1_binom)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer, family = binomial,
##      data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4184  -1.8174  -0.4575   1.5859   9.4807
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.15259    0.02131 -100.994 < 2e-16 ***
## pitbull      1.23537    0.02783  44.387 < 2e-16 ***
## chip_status -0.23649    0.03050  -7.753 8.97e-15 ***
## summer       0.37879    0.02938  12.891 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2641.29  on 51  degrees of freedom
## Residual deviance:  544.19  on 48  degrees of freedom
##      (2 observations deleted due to missingness)
## AIC: 867.48
##
## Number of Fisher Scoring iterations: 4
exp(confint(pitbull_model1_binom))

## Waiting for profiling to be done...
##              2.5 %    97.5 %
## (Intercept) 0.1114096 0.1211179
## pitbull     3.2569778 3.6324244
## chip_status 0.7434426 0.8378715
## summer      1.3786281 1.5469365
summary(pitbull_model1_quasi)

##

```

```
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer, family = quasibinomial,
##      data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4184  -1.8174  -0.4575   1.5859   9.4807
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.15259    0.06971 -30.880 < 2e-16 ***
## pitbull      1.23537    0.09103  13.572 < 2e-16 ***
## chip_status -0.23649    0.09976  -2.371 0.021826 *
## summer       0.37879    0.09610   3.942 0.000262 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 10.69677)
##
## Null deviance: 2641.29 on 51 degrees of freedom
## Residual deviance: 544.19 on 48 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(pitbull_model1_quasi))
```

```
## Waiting for profiling to be done...
```

```
##              2.5 %    97.5 %
## (Intercept) 0.1011467 0.1329404
## pitbull      2.8769183 4.1109439
## chip_status  0.6478111 0.9580232
## summer       1.2082342 1.7612914
```

Initially we see that all variables included have a significant relationship to the proportion of dogs who are dead at the end of their time in the shelter, even after inflating our standard errors to account for the variance structure of our data.

Next lets try to control for whether or not a animal was contagious when it was brought into the shelter.

```
pitbull_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0),
         chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0),
         contagious=ifelse(grepl(".*[~NON-]CONTAGIOUS", intake_condition),1,0)) %>%
  group_by(pitbull, chip_status, summer, contagious) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())

pitbull_model2_binom <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious, weights = count,
  family = binomial, data = pitbull_binom)

pitbull_model2_quasi <- glm(prop_dead ~ pitbull + chip_status +
```

```

        summer + contagious , weights = count,
        family = quasibinomial, data = pitbull_binom)

summary(pitbull_model2_binom)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious,
##      family = binomial, data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89655  -1.32251  -0.08517   1.08867   2.12765
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.20165    0.02169 -101.518 < 2e-16 ***
## pitbull       1.23078    0.02814  43.742 < 2e-16 ***
## chip_status  -0.22483    0.03077  -7.306 2.74e-13 ***
## summer        0.36978    0.02969  12.456 < 2e-16 ***
## contagious    1.98593    0.08113  24.477 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2699.882  on 15  degrees of freedom
## Residual deviance:   32.758  on 11  degrees of freedom
## (2 observations deleted due to missingness)
## AIC: 143.79
##
## Number of Fisher Scoring iterations: 3

exp(confint(pitbull_model2_binom))

## Waiting for profiling to be done...
##
##              2.5 %    97.5 %
## (Intercept) 0.1059964 0.1154017
## pitbull     3.2401347 3.6179798
## chip_status 0.7517627 0.8481433
## summer      1.3654369 1.5339545
## contagious  6.2166833 8.5454709

summary(pitbull_model2_quasi)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious,
##      family = quasibinomial, data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89655  -1.32251  -0.08517   1.08867   2.12765
##
## Coefficients:

```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.20165    0.03766 -58.455 4.54e-15 ***
## pitbull      1.23078    0.04887  25.187 4.45e-11 ***
## chip_status -0.22483    0.05344  -4.207 0.00147 **
## summer       0.36978    0.05156   7.172 1.82e-05 ***
## contagious   1.98593    0.14091  14.094 2.19e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 3.01608)
##
## Null deviance: 2699.882 on 15 degrees of freedom
## Residual deviance: 32.758 on 11 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
```

```
exp(confint(pitbull_model2_quasi))
```

```
## Waiting for profiling to be done...
```

```
##               2.5 %    97.5 %
## (Intercept) 0.1026891 0.1190278
## pitbull      3.1110085 3.7678996
## chip_status  0.7188083 0.8863497
## summer       1.3078316 1.6007855
## contagious   5.5323688 9.6176550
```

Next lets try to control for whether or not a animal was treatable when it was brought into the shelter.

```
pitbull_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0),
         chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0),
         contagious=ifelse(grepl("[^NON-]CONTAGIOUS", intake_condition),1,0),
         treatable=ifelse(grepl("^TREATABLE.*", intake_condition),1,0)) %>%
  group_by(pitbull, chip_status, summer, month, contagious, treatable) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())

pitbull_model3_binom <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious + treatable, weights = count,
  family = binomial, data = pitbull_binom)

pitbull_model3_quasi <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious + treatable,
  weights = count,
  family = quasibinomial, data = pitbull_binom)

summary(pitbull_model3_binom)
```

```
##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious +
```

```

##      treatable, family = binomial, data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.0491  -1.4236  -0.1897   0.9336   9.5936
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.59457    0.03949  -15.06 < 2e-16 ***
## pitbull      1.24955    0.02925   42.72 < 2e-16 ***
## chip_status -0.24774    0.03176   -7.80 6.21e-15 ***
## summer       0.39095    0.03070   12.73 < 2e-16 ***
## contagious   1.38010    0.08913   15.48 < 2e-16 ***
## treatable    -1.82378    0.03953  -46.14 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5475.42  on 194  degrees of freedom
## Residual deviance:  842.76  on 189  degrees of freedom
##      (2 observations deleted due to missingness)
## AIC: 1514.1
##
## Number of Fisher Scoring iterations: 4
exp(confint(pitbull_model3_binom))

## Waiting for profiling to be done...

##              2.5 %    97.5 %
## (Intercept) 0.5106406 0.5961339
## pitbull      3.2943807 3.6946299
## chip_status  0.7333041 0.8305406
## summer       1.3918868 1.5699049
## contagious   3.3384762 4.7351594
## treatable    0.1493824 0.1744212
summary(pitbull_model3_quasi)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious +
##      treatable, family = quasibinomial, data = pitbull_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.0491  -1.4236  -0.1897   0.9336   9.5936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.59457    0.08240  -7.216 1.26e-11 ***
## pitbull      1.24955    0.06104  20.472 < 2e-16 ***
## chip_status -0.24774    0.06628  -3.738 0.000246 ***
## summer       0.39095    0.06407   6.102 5.80e-09 ***

```



```
## contagious    1.38010    0.18600    7.420 3.86e-12 ***
## treatable     -1.82378    0.08249   -22.110 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 4.354402)
##
## Null deviance: 5475.42 on 194 degrees of freedom
## Residual deviance: 842.76 on 189 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(pitbull_model3_quasi))
```

```
## Waiting for profiling to be done...
```

```
##              2.5 %    97.5 %
## (Intercept) 0.4692407 0.6482124
## pitbull     3.0952904 3.9321196
## chip_status 0.6848701 0.8881191
## summer      1.3032732 1.6754195
## contagious  2.7616713 5.7307395
## treatable   0.1373228 0.1897619
```

Note that these models have different confidence intervals than the ones mentioned in the main report of our paper. While each of the three models in the appendix was fit on a dataset summarised specifically for each model, the models in the main report were all fit using the same dataset. By using the same dataset we can compare models using nested F tests, but SEs may be artificially small.

## Dogs vs Cats

```
cat_dog_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1 | cat == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0)) %>%
  mutate(chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0)) %>%
  group_by(dog, chip_status, summer, stray) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())
```

```
##### MODEL 1: dog #####
```

```
dogcat_model1_binom <- glm(prop_dead ~ dog, weights = count, family = binomial, data = cat_dog_binom)
```

```
dogcat_model1_quasi <- glm(prop_dead ~ dog, weights = count, family = quasibinomial, data = cat_dog_binom)
```

```
summary(dogcat_model1_binom)
```

```
##
## Call:
## glm(formula = prop_dead ~ dog, family = binomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
```

```

##      Min      1Q   Median      3Q      Max
## -19.837  -7.385  -2.250   4.985  15.639
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.43741    0.02199  -65.38  <2e-16 ***
## dog         -0.30509    0.02563  -11.90  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1745.6  on 16  degrees of freedom
## AIC: 1866
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model1_binom))

## Waiting for profiling to be done...
##              2.5 %    97.5 %
## (Intercept) 0.2274800 0.2479563
## dog         0.7010303 0.7751213
summary(dogcat_model1_quasi)

##
## Call:
## glm(formula = prop_dead ~ dog, family = quasibinomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -19.837  -7.385  -2.250   4.985  15.639
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4374    0.2249  -6.393 8.91e-06 ***
## dog          -0.3051    0.2621  -1.164   0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 104.5899)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1745.6  on 16  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model1_quasi))

## Waiting for profiling to be done...
##              2.5 %    97.5 %

```

```

## (Intercept) 0.1495373 0.3624932
## dog          0.4457027 1.2501118
##### MODEL 2: summer #####

dogcat_model2_binom <- glm(prop_dead ~ summer, weights = count, family = binomial, data = cat_dog_binom)

dogcat_model2_quasi <- glm(prop_dead ~ summer, weights = count, family = quasibinomial, data = cat_dog_binom)

summary(dogcat_model2_binom)

##
## Call:
## glm(formula = prop_dead ~ summer, family = binomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.202   -7.386   -4.440    5.671   17.663
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.81259     0.01405  -129.0  <2e-16 ***
## summer       0.45539     0.02372   19.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1524.9  on 16  degrees of freedom
## AIC: 1645.3
##
## Number of Fisher Scoring iterations: 4

exp(confint(dogcat_model2_binom))

## Waiting for profiling to be done...

##              2.5 %    97.5 %
## (Intercept) 0.1587829 0.1677739
## summer      1.5050770 1.6517349

summary(dogcat_model2_quasi)

##
## Call:
## glm(formula = prop_dead ~ summer, family = quasibinomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.202   -7.386   -4.440    5.671   17.663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.8126      0.1344 -13.490 3.71e-10 ***

```

```

## summer          0.4554      0.2268    2.008    0.0619 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 91.44736)
##
## Null deviance: 1883.0 on 17 degrees of freedom
## Residual deviance: 1524.9 on 16 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model2_quasi))

## Waiting for profiling to be done...
##
## 2.5 % 97.5 %
## (Intercept) 0.1243408 0.2107409
## summer      1.0043970 2.4491693
##### MODEL 3: chip #####
dogcat_model3_binom <- glm(prop_dead ~ dog + chip_status, weights = count, family = binomial, data = cat_dog_binom)
dogcat_model3_quasi <- glm(prop_dead ~ dog + chip_status, weights = count, family = quasibinomial, data = cat_dog_binom)
summary(dogcat_model3_binom)

##
## Call:
## glm(formula = prop_dead ~ dog + chip_status, family = binomial,
## data = cat_dog_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6677  -4.1391   0.1846  10.8318  14.4792
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.40636    0.02211 -63.599 <2e-16 ***
## dog          -0.24565    0.02620  -9.377 <2e-16 ***
## chip_status -0.29305    0.02828 -10.363 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1750.8 on 15 degrees of freedom
## Residual deviance: 1503.0 on 13 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: 1625.4
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model3_binom))

## Waiting for profiling to be done...

```

```
##           2.5 %    97.5 %
## (Intercept) 0.2345951 0.2558385
## dog         0.7431337 0.8235017
## chip_status 0.7056151 0.7883420
```

```
summary(dogcat_model3_quasi)
```

```
##
## Call:
## glm(formula = prop_dead ~ dog + chip_status, family = quasibinomial,
##      data = cat_dog_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6677  -4.1391   0.1846  10.8318  14.4792
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4064     0.2391  -5.883 5.38e-05 ***
## dog           -0.2456     0.2832  -0.867   0.401
## chip_status  -0.2931     0.3057  -0.959   0.355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 116.867)
##
##      Null deviance: 1750.8  on 15  degrees of freedom
## Residual deviance: 1503.0  on 13  degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(dogcat_model3_quasi))
```

```
## Waiting for profiling to be done...
```

```
##           2.5 %    97.5 %
## (Intercept) 0.1496550 0.3838132
## dog         0.4540004 1.3839138
## chip_status 0.3987708 1.3312323
```

```
##### MODEL 4: summer + dog + summer:dog #####
```

```
dogcat_model4_binom <- glm(prop_dead ~ summer+dog+ summer:dog, weights = count, family = binomial, data = cat_dog_binom)
dogcat_model4_quasi <- glm(prop_dead ~ summer+dog+ summer:dog, weights = count, family = quasibinomial, data = cat_dog_binom)
summary(dogcat_model4_binom)
```

```
##
## Call:
## glm(formula = prop_dead ~ summer + dog + summer:dog, family = binomial,
##      data = cat_dog_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -17.392   -7.749   -4.789    6.527   18.670
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.67600    0.03030 -55.313 < 2e-16 ***
## summer      0.56018    0.04434  12.634 < 2e-16 ***
## dog         -0.17185    0.03420  -5.024 5.05e-07 ***
## summer:dog  -0.18718    0.05275  -3.549 0.000387 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1421.8  on 14  degrees of freedom
## AIC: 1546.2
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model4_binom))

## Waiting for profiling to be done...
##             2.5 %    97.5 %
## (Intercept) 0.1762591 0.1984903
## summer      1.6052408 1.9099840
## dog         0.7877336 0.9007624
## summer:dog  0.7478053 0.9195843
summary(dogcat_model4_quasi)

##
## Call:
## glm(formula = prop_dead ~ summer + dog + summer:dog, family = quasibinomial,
##      data = cat_dog_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -17.392   -7.749   -4.789    6.527   18.670
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.6760    0.2978  -5.628 6.23e-05 ***
## summer        0.5602    0.4358   1.285  0.220
## dog          -0.1718    0.3362  -0.511  0.617
## summer:dog   -0.1872    0.5185  -0.361  0.723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 96.60753)
##
##    Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1421.8  on 14  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```

exp(confint(dogcat_model4_quasi))

## Waiting for profiling to be done...
##           2.5 %    97.5 %
## (Intercept) 0.09987833 0.324086
## summer      0.74071253 4.144038
## dog         0.44621757 1.682512
## summer:dog  0.29772816 2.292399
##### MODEL 5: stray #####

dogcat_model5_binom <- glm(prop_dead ~ stray, weights = count, family = binomial, data = cat_dog_binom)
dogcat_model5_quasi <- glm(prop_dead ~ stray, weights = count, family = quasibinomial, data = cat_dog_b
summary(dogcat_model5_binom)

##
## Call:
## glm(formula = prop_dead ~ stray, family = binomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.398   -7.178   -2.626    4.353   23.455
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.44258    0.01775  -81.28  <2e-16 ***
## stray       -0.36155    0.02303  -15.70  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1640.2  on 16  degrees of freedom
## AIC: 1760.6
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model5_binom))

## Waiting for profiling to be done...
##           2.5 %    97.5 %
## (Intercept) 0.228210 0.2446524
## stray      0.665877 0.7287805
summary(dogcat_model5_quasi)

##
## Call:
## glm(formula = prop_dead ~ stray, family = quasibinomial, data = cat_dog_binom,
##      weights = count)
##

```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.398   -7.178   -2.626    4.353   23.455
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4426     0.1803  -8.002 5.53e-07 ***
## stray        -0.3616     0.2339  -1.546   0.142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 103.1719)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1640.2  on 16  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(dogcat_model5_quasi))
```

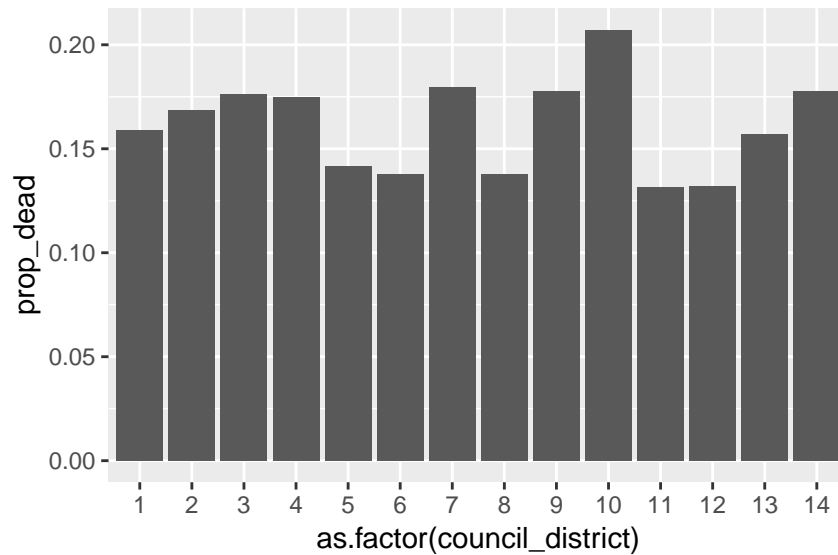
```
## Waiting for profiling to be done...
```

```
##              2.5 %    97.5 %
## (Intercept) 0.1636774 0.3324977
## stray       0.4414678 1.1069417
```

```
just_dog_adoptions=adoptions%>%
  filter(animal_type=="DOG")%>%
  mutate(animal_breed=as.factor(animal_breed))%>%
  mutate(healthy_intake=ifelse(grepl("^HEALTHY.*",intake_condition),1,0),
         contagious_intake=ifelse(grepl(".*[~NON~]CONTAGIOUS", intake_condition),1,0),
         untreatable_intake=ifelse(grepl(".*(UNTREATABLE).*", intake_condition),1,0),
         treatable_intake=ifelse(grepl("^TREATABLE.*", intake_condition),1,0),
         manageable_intake=ifelse(grepl(".*MANAGEABLE.*", intake_condition),1,0),
         rehabilitable_intake=ifelse(grepl(".*REHABILITABLE.*", intake_condition),1,0),
         normal_intake=ifelse(grepl(".*NORMAL.*", intake_condition),1,0),
         chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0))%>% #Isaac added this just to see how mod
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0))
```

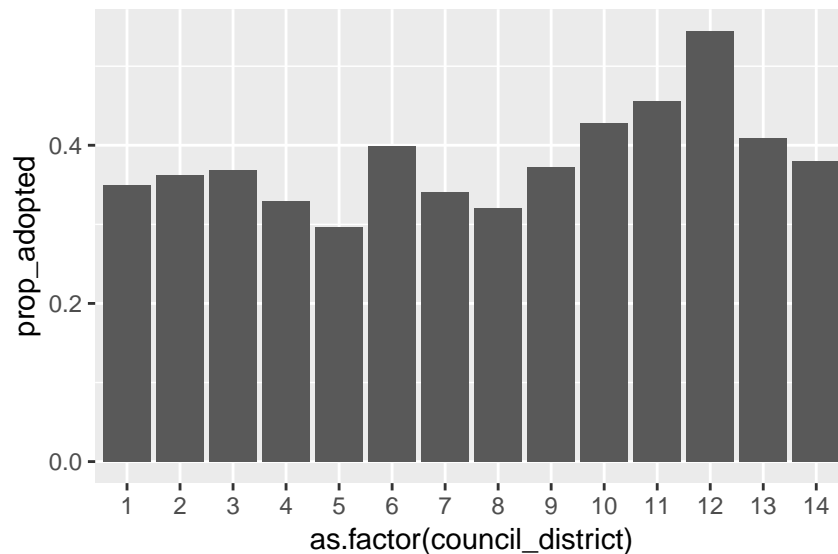
```
just_dog_adoptions %>%
  group_by(council_district)%>%
  summarise(prop_dead = sum(out_dead)/n())%>%
  filter(!is.na(council_district))%>%
  ggplot(aes(x=as.factor(council_district), y = prop_dead)) +
  geom_bar(stat = "identity")
```





Above you can see that certain city council districts have higher rates of dogs dying in the shelter system. City council district 12 seems to be the highest at 20.7% of the dogs are dying in the shelter system, and city council district 11 has the lowest at only 13.1% of the dogs dying in the shelter system.

```
just_dog_adoptions %>%
  group_by(council_district)%>%
  summarise(prop_adopted = sum(adopted)/n())%>%
  filter(!is.na(council_district)) %>%
  ggplot(aes(x=as.factor(council_district), y = prop_adopted)) +
  geom_bar(stat = "identity")
```



For proportion adopted we see a range of values. City council district 12 has 54.5% of the dogs adopted, while city council district 5 has the lowest proportion adopted at only 0.297.

### Random Intercepts:

```
mod.1=glmer(adopted~1+(1|council_district), data=just_dog_adoptions, family = "binomial")
summary(mod.1)
```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: adopted ~ 1 + (1 | council_district)
## Data: just_dog_adoptions
##
##      AIC      BIC    logLik deviance df.resid
## 58685.9 58703.3 -29340.9 58681.9    45485
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.0569 -0.7335 -0.6885  1.2963  1.5351
##
## Random effects:
## Groups             Name             Variance Std.Dev.
## council_district (Intercept) 0.05945  0.2438
## Number of obs: 45487, groups: council_district, 14
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.49300    0.06663  -7.399 1.37e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(fixef(mod.1))/(1+exp(fixef(mod.1)))

## (Intercept)
## 0.3791861

mod.1_dead=glmer(out_dead~1+(1|council_district), data=just_dog_adoptions, family = "binomial")
summary(mod.1_dead)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: out_dead ~ 1 + (1 | council_district)
## Data: just_dog_adoptions
##
##      AIC      BIC    logLik deviance df.resid
## 39522.0 39539.5 -19759.0 39518.0    45485
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.4909 -0.4592 -0.4080 -0.4014  2.4913
##
## Random effects:
## Groups             Name             Variance Std.Dev.
## council_district (Intercept) 0.01946  0.1395
## Number of obs: 45487, groups: council_district, 14
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.65465    0.04143  -39.94 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
exp(fixef(mod.1_dead))/(1+exp(fixef(mod.1_dead)))
```

```
## (Intercept)
## 0.1604818
```

## Multi-Level:

```
mod.2.adopted=glmer(adopted~summer+chip_status+treatable_intake+(1|council_district), data=just_dog_adopt,
summary(mod.2.adopted)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## adopted ~ summer + chip_status + treatable_intake + (1 | council_district)
## Data: just_dog_adoptions
##
##      AIC      BIC    logLik deviance df.resid
## 57316.1 57359.7 -28653.0 57306.1    45482
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.2426 -0.7608 -0.6787  1.2118  3.4802
##
## Random effects:
## Groups           Name          Variance Std.Dev.
## council_district (Intercept) 0.08389  0.2896
## Number of obs: 45487, groups: council_district, 14
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.93608    0.09252 -20.926 < 2e-16 ***
## summer         0.07182    0.02272   3.162  0.00157 **
## chip_status   -0.12224    0.02227  -5.490 4.03e-08 ***
## treatable_intake 1.59669    0.05142  31.050 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) summer chp_st
## summer      -0.069
## chip_status -0.073  0.027
## treatabl_nk -0.519  0.002 -0.014
```

```
exp(fixef(mod.2.adopted))
```

```
##      (Intercept)      summer      chip_status treatable_intake
##      0.1442687      1.0744612      0.8849374      4.9366449
```

```
exp(fixef(mod.2.adopted))/(1+exp(fixef(mod.2.adopted)))
```

```
##      (Intercept)      summer      chip_status treatable_intake
##      0.1260794      0.5179471      0.4694784      0.8315547
```

```
mod.2.dead=glmer(out_dead~summer+chip_status+treatable_intake+(1|council_district), data=just_dog_adopt,
summary(mod.2.dead)
```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## out_dead ~ summer + chip_status + treatable_intake + (1 | council_district)
## Data: just_dog_adoptions
##
##      AIC      BIC    logLik deviance df.resid
## 36001.9 36045.5 -17996.0 35991.9    45482
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.3865 -0.3979 -0.3527 -0.3135  3.4388
##
## Random effects:
## Groups           Name      Variance Std.Dev.
## council_district (Intercept) 0.01982  0.1408
## Number of obs: 45487, groups: council_district, 14
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.08265    0.05258   1.572   0.116
## summer         0.37470    0.02963  12.644 < 2e-16 ***
## chip_status    -0.23570    0.03116  -7.565 3.88e-14 ***
## treatable_intake -2.12055    0.03592 -59.039 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) summer chp_st
## summer      -0.149
## chip_status -0.194  0.020
## treatabl_ntk -0.530 -0.051  0.021
exp(fixef(mod.2.dead))

##      (Intercept)      summer      chip_status treatable_intake
##      1.0861591      1.4545479      0.7900191      0.1199661
exp(fixef(mod.2.dead))/(1+exp(fixef(mod.2.dead)))

##      (Intercept)      summer      chip_status treatable_intake
##      0.5206502      0.5925930      0.4413468      0.1071159

```