Final Report

Isaac Slagel and Jack Welsh

Introduction

According to an estimate by the American Society for the Prevention of Cruelty to Animals (ASPCA), over 6.5 millions animals enter a shelter system every year. (Shelter Intake and Surrender, ASPCA). After entering the shelter system, animals are either rehabilitated and sent to homes or euthanized. Animal shelters run mostly on revenue from donations and adoption fees but they routinely run into permanent deficits and must be supported with money from local governments (Goselin et. al. 2011). Knowing more about the outcomes of animals from these shelters could help minimize the percent of animals that need to be euthanized. Further, by identifying which animals are less likely to be adopted, we can help shelters allocate resources to minimize the time these animals spend in the shelter system.

Our research looks into animal shelters in the Dallas, Texas area using a dataset provided by the city of Dallas on dallasopendata.com. This data includes information about every animal that has been brought into an animal shelter in the Dallas area. There are 61634 individual animals in this dataset with information about when and where the animal was found, the health of the animal on arrival, and whether the animal was adopted, euthanized, or transferred.

Recent research on animal shelters have shown that outcomes for pitbulls are consistently worse than other dog breeds (Patronek & Crowe 2018, Lepper et al. 2002). We are interested in seeing if these this trends exist in the Dallas animal shelter system. If these trends do occur, how do they manifest in the quantitative insights we glean from our analysis. Do pitbulls face greater odds of euthanization than other breeds? Are pitbulls adopted at lower rates than other species of dogs. Do certain animal shelters handle pitbulls better or worse than others?

Materials and Methods

We accessed our data from the Dallas Open Data website. This website contains many public datasets, including animal shelter records. Anyone can request an access key to this dataset and then download it. We downloaded 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. The initial variables from this dataset are described below in Table 1.

Table 1: Description of Variables

Variable Name	Variable Role	Variable Type	Description
out_dead	response	bianary	Whether or not an animal was dead at the outcome
days_in_shelter	response	time	The number of days an animal spent in the animal shelter
summer	explanatory	bianary	Whether or not an animal had its outcome in May-Sept
pitbull	explanatory	bianary	Whether or not a dog was a pitbull
chip_status	explanatory	bianary	Whether or not an animal had a scannable chip
contagious	potential confounder	bianary	Whether or not an animal was described as 'contagious' at intake
treatable	potential confounder	bianary	Whether or not an animal was described as 'treatable' at intake

After we imported our dataset, we used a variety of variables that were coded as factors and string to create a fleet of indicator variables: `adopted`, `chip_status`, `summer`, `treatable_intake`, `adopted`, `dead`. These indicator variables were created to streamline our modeling process. They allow us to increase the interpretability of variables with multiple categories (like outcome_type). We also modified the intake date and outcome date variables to create a `days_in_shelter` variable. This variable is a necessary outcome for the survival analysis portion of our modeling.

We used quasi-binomial regression to compare how the outcomes of pitbulls differ from other breeds of dog. In fitting this model we had to control for a plethora of confounding variables including season of outcome,

chip status, and intake condition of the animal.

In order to fit a quasi-binomial model, we had to summarize our dataset with respect to certain variables we controlled for in our model. One interesting issue that we encountered was the effect that the dataset used had 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.

We also fit a random intercepts logistic regression model to compare outcomes after accounting for correlation within city council districts. Our modeling process took on a similar approach to the binomial regression, we examined differences in odds of dying for pitbulls and non-pitbulls after accounting for differences in chip status, intake condition, and season.

To investigate the number of days that dogs spend in the animal shelter before dying, we used a Cox proportional hazards model. Cox proportional hazards is a semi parametric model used in survival analysis to explain how variables of interest affect rate of the occurance of an event (death). The model does this by comparing death rates of different groups. By carrying out this modeling process, we can learn about how rate of euthanization differers between pitbulls and other dog species.

The Cox proportional hazards model is a semi-parametric model, this means that this model does not make assumptions about the underlying distribution of data, but it does assume that the hazard ratios for any two individuals is constant over time. In other words, the risk of the outcome of interest happening for one individual at every time point must be proportionally constant to all other individuals.

This model follows the general form $h(t) = h_0(t) * \exp\{b_1 * x_1 + b_2 * x_2 + ... + b_p * x_p\}$. where $h_0(t)$ is the baseline hazard rate and is an estimation of the hazard rate of an individual at time t with x_1 to x_p all equal to 0 having the outcome of interest already happened. The parameters b_1 to b_p are the hazard ratio estimates. These are estimations of how much the hazard rate will change at compared to the baseline hazard rate at any time t. This model also allows us to censor our data. We do not know if an animal that is transferred dies in the animal shelter, so that variable is censored at the time of the outcome. This us allows us to consider all of our data without making assumptions on whether the animal is dead or alive.

Results

EDA Results: Pitbull

To look into how pitbulls are handled within animal shelters, we decided to explore how the outcomes of pitbulls may differ from non-pitbulls. Table 2 describes the differences in outcome rates for non-pitbulls compared to pitbulls.

Table 2: Pitbull Outcomes

Outcome	Pitbull (%)	Non-Pitbull (%)
Adoption	32.03	35.73
Euthanized	28.42	10.25
Returned to owner	22.25	30.96
Transfer	10.92	17.56
Foster	2.25	1.64
Other	1.88	1.59
Dead on arrival	0.93	0.82
Treatment	0.77	1.01
Died	0.51	0.39
Missing	0.04	0.06

We see that pitbulls, while only adopted at a slightly lower rate (~5%), are euthanized at well over double the rate of other dogs. Further, we see that other dogs have a much higher chance of being transferred to

another facility or returned to their owner. This prompted us to create a variable `out_dead` which identifies whether a dog had an outcome of death ("EUTHANIZED" or "DIED") or not ("TRANSFER", "RETURNED TO OWNER", "ADOPTION"). To get a better understanding of the relationship between `pitbull` and `out_dead`, we decided to control for chip status, intake condition, and season of outcome in our quasi-binomial regression.

Quasi-Binomial Models for Dog Outcomes

Table 3: Modeling Dog Outcomes in Dallas Animal Shelters

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

We see in Table 3 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. Furthermore, by two nested F tests we see each of our models is an improvement on the one before it (Model 1 → Model 2: F = 5.1141, p-value = 0.031) (Model 2 → Model 3: F = 313.62, p-value $<<<$ .05).

Proceeding with Model 3 as our best model of dog outcomes from the animal shelter, we glean several insights about the relationships between certain characteristics of dogs and those animals' outcomes from the shelter. For instance, if a dog is a pitbull, on average its odds of leaving the animal shelter dead increase by 348%, after controlling for chip status, season, and intake condition. Furthermore , if a dog is a pitbull we are 95% confident that the true increase in odds of death at outcome is between 302% and 402%. 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.

EDA Results: Effect of Council District

An additional variable that was available in our dataset was `council_district`. The city of Dallas has 14 council districts. We are concerned that there may be some correlation between outcomes for each council district.
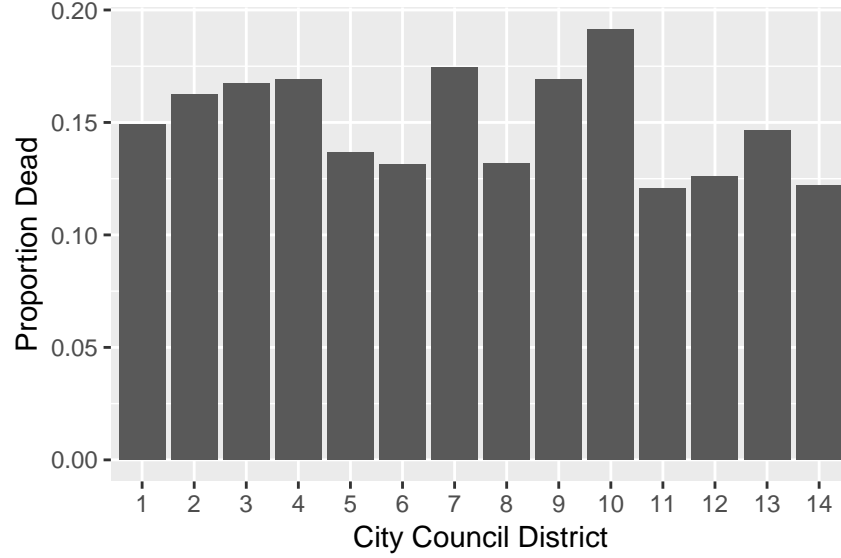

Figure 1: Proportion of dogs with dead outcomes by council district.

From Figure 1, we see that the proportion of dogs dying in animals shelters differs by city council district. City council district 12 has the highest rate of dog death at 19.2%. City council district 11 has the lowest at only 12.1% of the dogs dying in the shelter system.

Random Intercepts Model for Dog Outcomes

We see from above that there is a difference in the proportion of dogs who die between city council districts. To account for this, we used a random intercepts logistic regression model.

Our model has two levels,

- Level 1

$$\log(\frac{p_{ij}}{1-p_{ij}}) = a_i + b_i(\text{summer})_{ij} + c_i(\text{pitbull})_{ij} + d_i(\text{chip status})_{\text{ij}}$$
$$f_i(\text{treatable intake})_{ij} + g_i(\text{contagious intake})_{ij}$$

- Level 2

$$
\begin{aligned}
a_i &= \alpha_0 + u_i \\
b_i &= \beta_0 \\
c_i &= \gamma_0 \\
d_i &= \delta_0 \\
f_i &= \zeta_0 \\
g_i &= \theta_0
\end{aligned}
$$

The variance component is described by:

$$u_i = N(0, \sigma_u^2).$$

In Table 4 below, we note that the odds of a pitbull dying in the animal shelter are 3.47 times higher than for non pitbulls accounting for season, chip status, and intake condition. As expected, the odds of a contagious dog dying at the animal shelter are 3.96 time higher than a non contagious dog accounting for whether the dog is treatable, the season, chip status and pitbull. Also, unsurprisingly, if a dog enters the animal shelter with a treatable intake the odds of dying in the animal shelter decrease by 84% controlling for whether the dog is contagious on intake, the season, chip status and pitbull.

Table 4: Modeling Dog Deaths in Animal Shelters using Random Intercepts Logistic Regression

|	*Dependent variable:*
	Proportion of dogs who died
Intercept	0.549***
	(0.496,0.607)
Pitbull	3.474***
	(3.280,3.680)
Scannable Chip	0.795***
	(0.746,0.847)
Summer Outcome	1.483***
	(1.396,1.575)
Contagious	3.964***
	(3.325,4.724)
Treatable At Intake	0.159***
	(0.147,0.172)
Observations	45,098
Log Likelihood	-16,733.970
Akaike Inf. Crit.	33,481.950
Bayesian Inf. Crit.	33,542.970
Note:	*p<0.1; **p<0.05; ***p<0.01

EDA Results: Time Until Death for Dogs
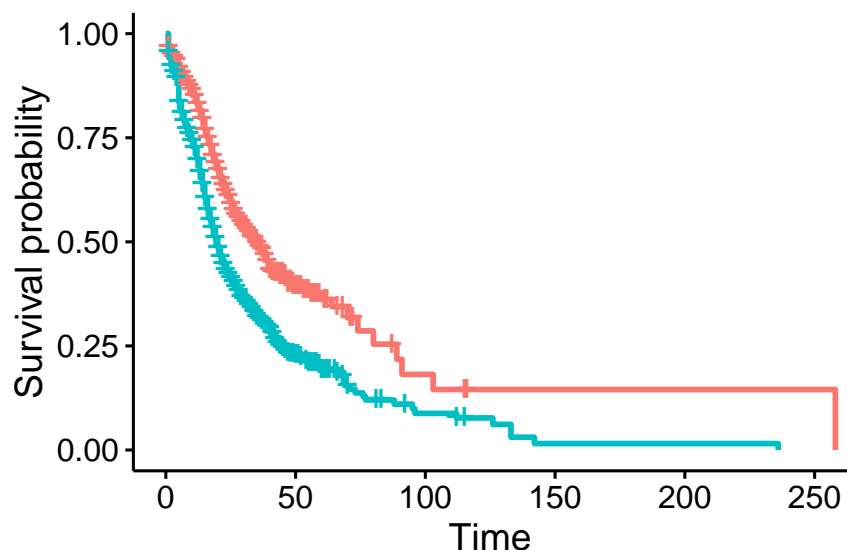
Figure 2: Kaplan–Meier Curves for pitbull (blue) and non-pitbull survival (pink).

When investigating the amount of time that dogs spend in the animal shelter before dying or having another outcome, we noticed that pitbulls die much more often (see Table 2). In the Kaplan Meier curve above, we can see that the survival probability for pitbulls falls below 50% sooner than that of non pitbulls.

Table 5: Average Time Spent in Animal Shelter

Pitbull	Days
No	6.023040
Yes	9.168684

From Table 5 we see that the average time a pitbull spends in a shelter is longer than the average time a non pitbull spends in a shelter. Pitbulls are spending more time in the animal shelter and therefore having more opportunities to be adopted, yet have a higher proportion die in the animal shelter.

Modeling Time Until Death for Pitbulls

In order to model differences in dog survival over time, we used Cox proportional hazards. We wanted to control for `chip_status` and `summer`, but the proportional hazards assumption was no longer met when we added these confounding variables to our model (see appendix). To get around this, we stratified our data by both `summer` and `chip_status`. For each of these 4 strata we fit the model $h(t) = h_0(t) * \exp\{b_1 * \text{Pitbull}\}$. Our findings are summarized in Table 6 below.

Table 6: exp(b_1) Estimates for Each Strata

	Strata			
	Summer and Chip	Summer and No Chip	Not Summer and Chip	Neither Summer or Chip
pitbull	1.698***	1.717***	1.694***	1.941***
	(1.431,2.016)	(1.553,1.897)	(1.509,1.902)	(1.509,1.902)

Note: *p<0.1; **p<0.05; ***p<0.01

We found that in all 4 strata the coefficient for pitbull is significant. The hazard ratio ranged from 1.69 in the strata where it is not the summer and all the animals have chips to 1.94 in the not summer and all animals do not have a chip dataset. In layman's terms, at any time during an animals stay at the animal shelter, we expect the death rate of pitbulls to be 1.69 to 1.94 times higher than non pitbulls, depending on the strata.

Discussion

We have demonstrated clear evidence that pitbulls that come into the Dallas animal shelter system have a far higher risk of dying in the shelter system. We demonstrated this consistently through a binomial approach, a multilevel approach, and a survival analysis based approach. We urge the Dallas animal shelter system to consider allocating more resources to pitbulls coming into the shelter. Lampe and Witte (2015) evidenced the effect that photogenicity can have on animal adoption. Perhaps by better marketing at risk animals, implementing more training programs for these animals, or advocating for more pitbulls to be fostered, fewer pitbulls would need to be euthanized in the Dallas animal shelter system.

In our modeling approach, we used our findings in our exploratory data analysis and previous literature models to drive our choices for confounding variables to include in our models. In both the quasi-binomial and random intercepts model we controlled for chip status, intake condition, and season of outcome. Further, in the random intercepts model, we adjusted our model for possible correlation within council districts. In our Cox proportional hazards modeling, we stratified our model on season and chip status to better meet the proportional hazards assumption. We would like to better our understanding of the intake condition and outcome condition variables in future research. In these variables, shelter workers described animals as being "healthy", "treatable", "contagious", etc. but the exact definitions and consistencies of these descriptions are not known.

Conclusions about outcomes of animals in shelters from this report can only be extended to the city of Dallas. While our sample size is totally representative for this population, we cannot account for differences that would be encountered in other municipalities, say higher adoption rates or less stigma against certain breeds of dog. Further, we cannot be sure that the trends observed in this paper will continue into the future. Anti-pitbull advocacy groups like National Pitbull Victims Awareness and pro-pitbull advocacy groups like Love-a-Bull are at odds trying to sway public opinion and policy regarding the handling of pitbulls in animal shelters. If changes in public perception of pitbulls or the policy that surrounds them changes, our model will lose it's generalizability to Dallas.

A limitation of our study is that we do not know what happens to animals after they have been transferred out of the animal shelter. In our analysis, we have viewed this outcome as a live outcome, but for all we know these animals could go on to another shelter where they are euthanized. Depending on the validity of our assumption, we could be adding some bias to our conclusions. Additionally, there is not consistent information on the criteria used by shelters to evaluate animals intake conditions. We looked for key words that indicated the health of the animal. However, we are not sure how consistent these evaluations are across all Dallas animal shelters. We are also unsure of how the decision is made to euthanize an animal. We do not know if an animal is euthanized if it has been in the shelter for more than a certain number of days. We also don't know if this is consistent across all of the animal shelters.

Our analysis was effective in approaching a particular question from a number of unique approaches. We implemented a binomial regression, a random intercepts logistic regression, and a Cox proportional hazards model to investigate how pitbull outcomes differ from other breeds of dogs after accounting for confounding variables. Further, the results from each of these investigations tell a similar story about the increase rates of death for pitbulls in the Dallas animal shelter system.

One weakness of our analysis came in the quasi-binomial modeling of dog outcomes. We found that the predicted SE and significance of our parameter estimates depended extensively on how we summarized our dataset. We chose to run our models off of a single summarized dataset with all of our confounders of interest. We justify this because grouped data would hopefully be collected with all confounders contained. By doing this we can use nested-F tests to compare our models, which helps us verify which confounders must be controlled for. Nonetheless, this approach may be giving us underestimated SEs and overstated precision. The reason for this phenomenon remains unclear and could be investigated in future research.

Future work could be done in Dallas to help minimize the number of animals which need to be euthanized. By identifying specific at-risk groups of animals, more resources could be allocated to advocating for these animals or transferring them to No-Kill shelters. We recommend any additional research to work directly with the Dallas Animal Shelter group. By having a direct contact with the group doing data collection, a statistician will be able to better identify pressing issues of the group and verify certain assumptions that we currently cannot.

Works Cited

Goselin, C., Letson, K., McGuinniss, K., Walters, A., & Fagan, J. M. (2011). Animal shelter funding.

Shelter Intake and Surrender. (n.d.). Retrieved from https://www.aspca.org/animal-homelessness/shelter-intake-and-surrender

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.

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.

Patronek, G. J., & Crowe, A. (2018). Factors Associated with High Live Release for Dogs at a Large, Open-Admission, Municipal Shelter. Animals : an open access journal from MDPI, 8(4), 45. doi:10.3390/ani8040045

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.

Annotated Appendix

All of our work can be accessed in a public Github repository at https://github.com/islagel/adoptions.

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 = binom

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

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
```

```r
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
```

```r
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
```

```r
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.

```r
pitbull_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1) %>%
```

```r
    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
```

```r
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
```

```r
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

```r
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 #########
```

```r
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_bino

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
```

```r
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
```

```r
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
```

```r
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
```

```r
####### 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_

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
```

```r
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
```

```r
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
```

```r
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 ######
```

```r
dogcat_model3_binom <- glm(prop_dead ~ dog + chip_status, weights = count, family = binomial, data = ca
```

```r
dogcat_model3_quasi <- glm(prop_dead ~ dog + chip_status, weights = count, family = quasibinomial, data
```

```r
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
```

```r
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
```

```r
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
```

```r
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

dogcat_model4_quasi <- glm(prop_dead ~ summer+dog+ summer:dog, weights = count, family = quasibinomial,

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
```

```r
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
```

```r
####### 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
```

```r
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
```

```r
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
```

```r
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
```

```r
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),
         rehabitable_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")
```
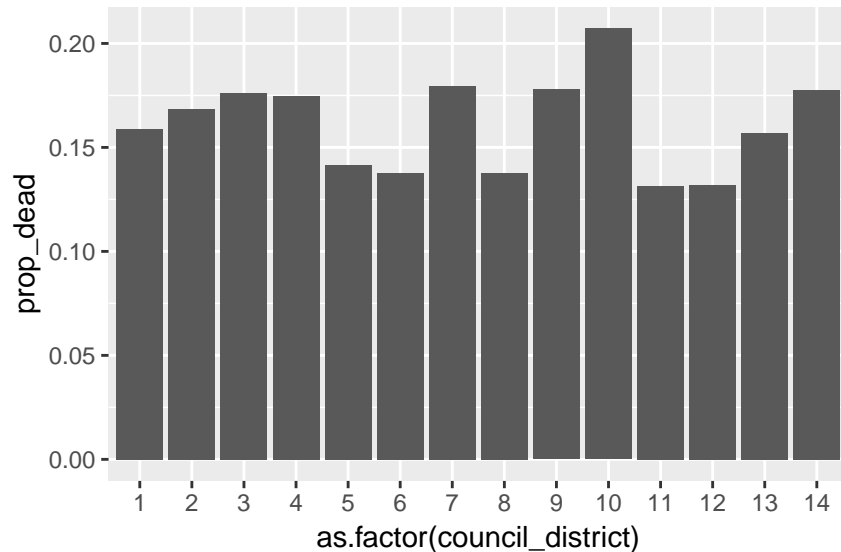

Above you can see that certain city council districts have higher rates of dogs dying in the shelter system. City council district 12 seems to be the highest at 20.7% of the dogs are dying in the shelter system, and city council district 11 has the lowest at only 13.1% of the dogs dying in the shelter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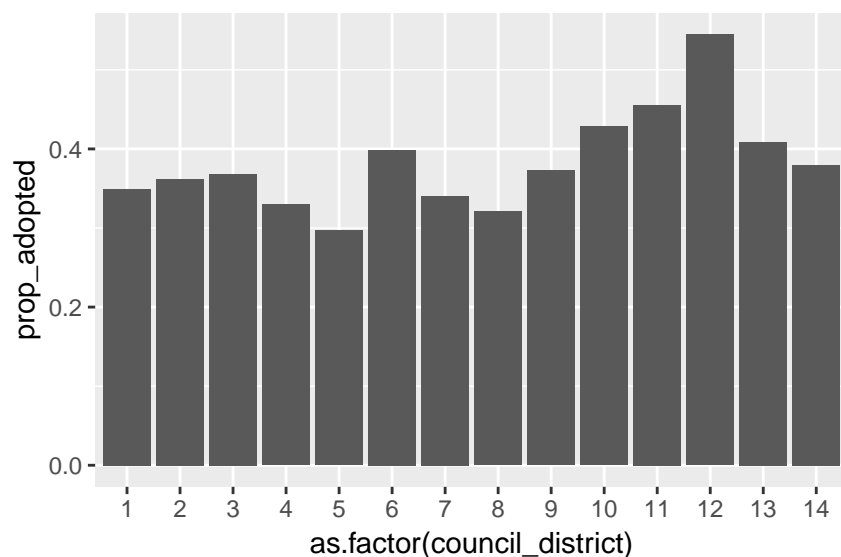

For proportion adopted we see a range of values. City council district 12 has 54.5% of the dogs adopted,

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
```

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

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

Multi-Level:

```r
mod.2.adopted=glmer(adopted~summer+chip_status+treatable_intake+(1|council_district), data=just_dog_ado
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
## treatbl_ntk -0.519  0.002 -0.014
```

```r
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+pitbull+(1|council_district), data=just_dog_adoptions, fam
summary(mod.2.dead)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## out_dead ~ summer + chip_status + pitbull + (1 | council_district)
##    Data: just_dog_adoptions
##
##      AIC      BIC   logLik deviance df.resid
##  37520.3  37564.0 -18755.2  37510.3    45482
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.8645 -0.4041 -0.3588 -0.3244  3.4188
##
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  council_district (Intercept) 0.01504  0.1226
## Number of obs: 45487, groups:  council_district, 14
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.03183    0.04088 -49.703   <2e-16 ***
## summer       0.36052    0.02878  12.527   <2e-16 ***
## chip_status -0.23750    0.03024  -7.855    4e-15 ***
## pitbull      1.18989    0.02741  43.411   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) summer chp_st
## summer      -0.228
## chip_status -0.214  0.019
## pitbull     -0.246  0.034 -0.018
exp(fixef(mod.2.dead))

## (Intercept)       summer chip_status      pitbull
##   0.1310956    1.4340745    0.7885970    3.2867036
exp(fixef(mod.2.dead))/(1+exp(fixef(mod.2.dead)))

## (Intercept)       summer chip_status      pitbull
##   0.1159015    0.5891662    0.4409026    0.7667205
```

Cox Proportional Hazards

As you can see below when we include chip_status and summer in the cox proportional hazards model the

kaplan meir curves begin to overlap and cross which means the the hazards are no longer proportional.

```
kp_curve.all.confounders=survfit(Surv(days_in_shelter,censored)~pitbull+summer+chip_status,
    data=adoptions_surv)

ggsurvplot(kp_curve.all.confounders, data=adoptions_surv)
```