

A close-up photograph of a gray cat's face. The cat has bright yellow eyes and is looking directly at the camera. It is positioned behind a yellow rectangular object, likely a book or a piece of paper, which is resting on a light-colored wooden surface. The background is a plain, light-colored wall.

Save the kitties: Factors affecting feline shelter outcomes

PBHLT 7120

Meg Rosales

25 November 2019

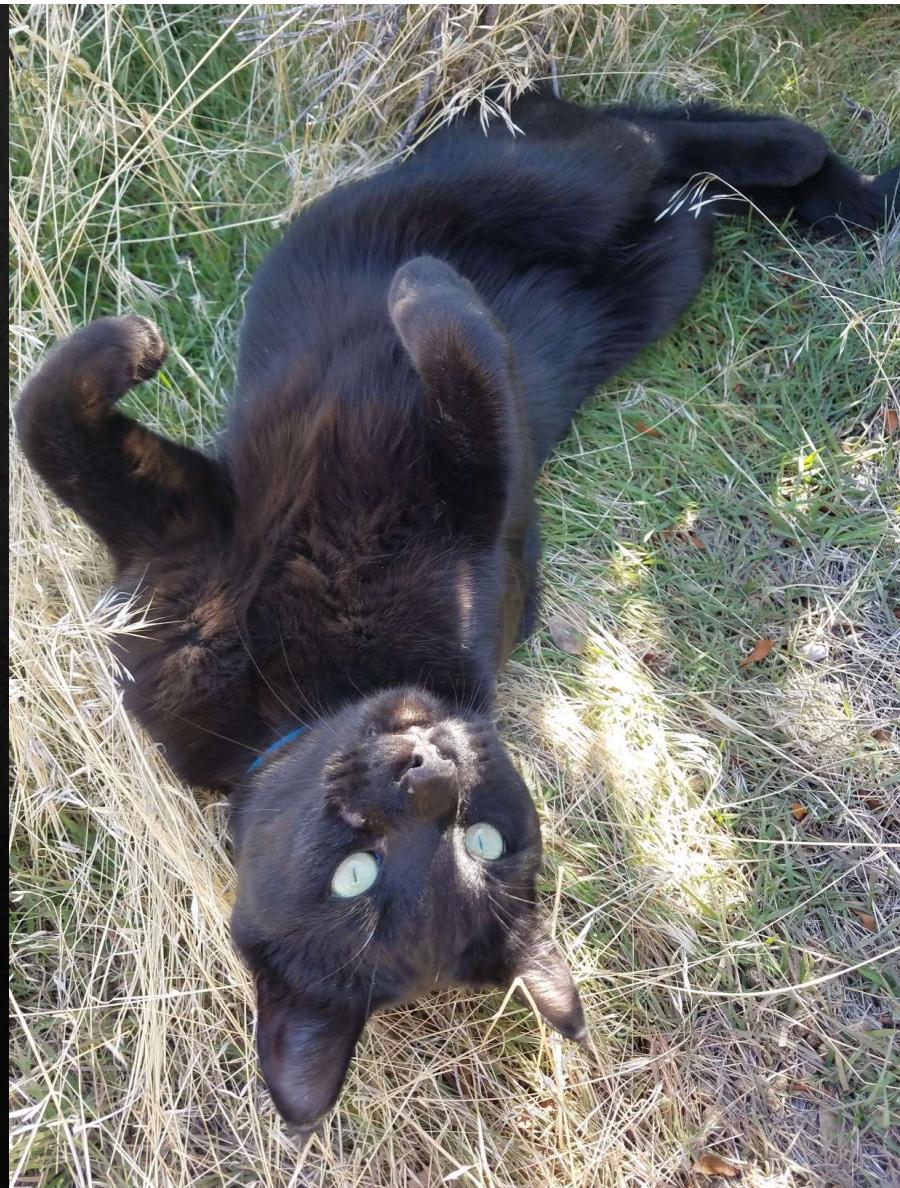
Why are you doing this?



- ❖ I love floofs
- ❖ Original data had cats, dogs, and other animals – simplified down to only cats due to time and large sample problem
- ❖ Lack of literature on factors influencing animal shelter outcomes
- ❖ According to the American Society for the Prevention of Cruelty to Animals (ASPCA), every year, about 3.2 million animals are adopted from animal shelters, and about 1.5 million animals are euthanized (ASPCA 2019).
- ❖ Research like this can help inform shelters of trends so they can take appropriate action to improve animal outcomes.

Questions

- What factors influence whether a shelter cat is adopted?
 - Examples of potential factors:
 - Color
 - Age
 - Gender
- What makes a cat more likely to be adopted?



Hypothesis

- This is an exploratory investigation, so we don't really start off with a hypothesis.
- However, some common knowledge/common sense ideas:
 - Older animals less likely to be adopted
 - People like baby animals – younger cats/kittens more likely to be adopted
 - Black cats less likely to be adopted - superstition
 - Spayed/neutered preferred.



Meet the data



- ❖ Austin Animal Center Outcomes Data – 10/1/13 – 10/22/19 (updated daily)
 - ❖ 19,826 cats who were adopted or euthanized – excluded other animal types and transfers/returns to owner/deaths
 - ❖ Eliminated all missing data prior to analysis
 - ❖ Outcome: adopted (1 = yes, 0 = no (euthanized)) – 18,365 adopted (92.6%), 1461 (7.4%) euthanized
 - ❖ Predictors:
 - ❖ Animal age at outcome in days – originally continuous, simplified to 3 categories
 - ❖ Breed – several binary variables for the most common breeds
 - ❖ Mixed breed
 - ❖ Color – several binary variables for the most common colors
 - ❖ Breed and color started out as multinomial, but I simplified after the univariable regression and started over as the model got unnecessarily complicated.
 - ❖ Mixed color
 - ❖ Gender
 - ❖ Intactness
 - ❖ Has name – some animals

Model building

Binary outcome variable -> logistic regression

(We'll check assumptions in a bit.)



Step 1: Univariable logistic regressions

Variable	Coefficient	p-value
has name	2.48975	< .001
age at outcome	-0.0004899	< .001
intact	-3.68437	< .001
female	0.23969	< .001
mixed color	0.07248	0.187
mixed breed	-0.4753	< .001
black	-0.02443	0.764
orange	-2.21776	< .001
brown	0.007893	0.894
white	-0.85808	< .001
gray	-1.14878	0.00374
multi	0.14883	0.0827
domestic shorthair	0.08541	0.2
domestic longhair	-0.38012	< .001
snowshoe	0.66996	0.19
American shorthair	0.64923	0.204

- ❖ Ran each variable individually in a logistic regression model – eliminated any variable with p > 0.25
- ❖ Only variables with p > 0.25 were color related: black and brown
- ❖ Kept all other variables

Step 2: Full model

- ❖ Added all variables to the regression model.
- ❖ Eliminated any with $p > 0.05$:
mixed color, white, gray,
multicolored, domestic
longhair, snowshoe,
American shorthair

```
# Call:  
# glm(formula = adopted ~ has_name + outcomeage + intact + female +  
#   mixedcolor + mixedbreed + orange + white + gray + multi +  
#   domsh + domlong + snowshoe + amsh, family = binomial(link = "logit"),  
#   data = cat)  
#  
# Deviance Residuals:  
#   Min     1Q  Median     3Q    Max  
# -3.8022  0.0511  0.0612  0.2328  3.0532  
#  
# Coefficients:  
#              Estimate Std. Error z value Pr(>|z|)  
# (Intercept) 4.072e+00 2.137e-01 19.058 < 2e-16 ***  
# has_name    3.458e+00 1.044e-01 33.126 < 2e-16 ***  
# outcomeage  -1.214e-03 3.371e-05 -36.016 < 2e-16 ***  
# intact      -4.885e+00 1.105e-01 -44.197 < 2e-16 ***  
# female      3.631e-01 9.321e-02  3.895 9.81e-05 ***  
# mixedcolor   1.659e-01 8.808e-02  1.883 0.059675 .  
# mixedbreed  -9.333e-01 1.722e-01 -5.420 5.98e-08 ***  
# orange      -3.086e+00 9.335e-01 -3.306 0.000947 ***  
# white       -2.228e-01 4.149e-01 -0.537 0.591271  
# gray        9.682e-02 7.424e-01  0.130 0.896241  
# multi      -6.482e-02 1.457e-01 -0.445 0.656380  
# domsh      -2.463e-01 1.216e-01 -2.026 0.042808 *  
# domlong    -2.332e-01 2.030e-01 -1.149 0.250628  
# snowshoe   7.113e-01 7.371e-01  0.965 0.334516  
# amsh       3.614e-01 6.765e-01  0.534 0.593216  
# ---  
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
#  
# (Dispersion parameter for binomial family taken to be 1)  
#  
# Null deviance: 10431.8  on 19825  degrees of freedom  
# Residual deviance: 4105.9  on 19811  degrees of freedom  
# AIC: 4135.9  
#  
# Number of Fisher Scoring iterations: 7
```

Step 2: Reduced model

- ❖ Checked if reduced model was a better fit with likelihood ratio test (LRT)
- ❖ p-value non-significant - Reduced model is better

```
## Call:  
## glm(formula = adopted ~ has_name + outcomeage + intact + female +  
##       mixedbreed + orange + domsh, family = binomial(link = "logit"),  
##       data = cat)  
##  
## Deviance Residuals:  
##      Min        Q1        Median         Q3        Max  
## -3.8623   0.0505   0.0597   0.2512   3.0870  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 4.090e+00 2.026e-01 20.194 < 2e-16 ***  
## has_name    3.455e+00 1.042e-01 33.163 < 2e-16 ***  
## outcomeage -1.214e-03 3.352e-05 -36.231 < 2e-16 ***  
## intact     -4.880e+00 1.103e-01 -44.261 < 2e-16 ***  
## female      3.561e-01 8.620e-02  4.132 3.60e-05 ***  
## mixedbreed -9.334e-01 1.720e-01 -5.427 5.72e-08 ***  
## orange     -3.136e+00 9.296e-01 -3.374 0.000741 ***  
## domsh      -1.956e-01 1.044e-01 -1.873 0.061037 .  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
## Null deviance: 10431.8 on 19825 degrees of freedom  
## Residual deviance: 4112.8 on 19818 degrees of freedom  
## AIC: 4128.8  
##  
## Number of Fisher Scoring iterations: 7  
  
#compare to full model  
anova(cat_reduced, cat_full, test = "Chisq")  
  
## Analysis of Deviance Table  
##  
## Model 1: adopted ~ has_name + outcomeage + intact + female + mixedbreed +  
##          orange + domsh  
## Model 2: adopted ~ has_name + outcomeage + intact + female + mixedcolor +  
##          mixedbreed + orange + white + gray + multi + domsh + domlong +  
##          snowshoe + amsh  
##      Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1      19818      4112.8  
## 2      19811      4105.9  7    6.9375  0.4354
```

Step 3: Check for 20% coefficient change

- ❖ Visual inspection and calculation of change for suspicious coefficients
- ❖ > 20% change in coefficient for domestic shorthair – add back mixed color as it was close to significant in previous model

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	4.072e+00	2.137e-01	19.058	< 2e-16	***
has_name	3.458e+00	1.044e-01	33.126	< 2e-16	***
outcomeage	-1.214e-03	3.371e-05	-36.016	< 2e-16	***
intact	-4.885e+00	1.105e-01	-44.197	< 2e-16	***
female	3.631e-01	9.321e-02	3.895	9.81e-05	***
mixedcolor	1.659e-01	8.808e-02	1.883	0.059675	.
mixedbreed	-9.333e-01	1.722e-01	-5.420	5.98e-08	***
orange	-3.086e+00	9.335e-01	-3.306	0.000947	***
white	-2.228e-01	4.149e-01	-0.537	0.591271	
gray	9.682e-02	7.424e-01	0.130	0.896241	
multi	-6.482e-02	1.457e-01	-0.445	0.656380	
domsh	-2.463e-01	1.216e-01	-2.026	0.042808	*
domlong	-2.332e-01	2.030e-01	-1.149	0.250628	
snowshoe	7.113e-01	7.371e-01	0.965	0.334516	
amsh	3.614e-01	6.765e-01	0.534	0.593216	

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	4.090e+00	2.026e-01	20.194	< 2e-16	***
has_name	3.455e+00	1.042e-01	33.163	< 2e-16	***
outcomeage	-1.214e-03	3.352e-05	-36.231	< 2e-16	***
intact	-4.880e+00	1.103e-01	-44.261	< 2e-16	***
female	3.561e-01	8.620e-02	4.132	3.60e-05	***
mixedbreed	-9.334e-01	1.720e-01	-5.427	5.72e-08	***
orange	-3.136e+00	9.296e-01	-3.374	0.000741	***
domsh	-1.956e-01	1.044e-01	-1.873	0.061037	.

Back to Step 2: Add back mixed color

- Compared to the full model, domestic shorthair is only reduced by 16% this time.
- LRT very non-significant, this model is a better fit.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 4.072e+00 2.137e-01 19.058 < 2e-16 ***
has_name     3.458e+00 1.044e-01 33.126 < 2e-16 ***
outcomeage   -1.214e-03 3.371e-05 -36.016 < 2e-16 ***
intact      -4.885e+00 1.105e-01 -44.197 < 2e-16 ***
female       3.631e-01 9.321e-02  3.895 9.81e-05 ***
mixedcolor    1.659e-01 8.808e-02  1.883 0.059675 .
mixedbreed   -9.333e-01 1.722e-01 -5.420 5.98e-08 ***
orange      -3.086e+00 9.335e-01 -3.306 0.000947 ***
white        -2.228e-01 4.149e-01 -0.537 0.591271
gray         9.682e-02 7.424e-01  0.130 0.896241
multi        -6.482e-02 1.457e-01 -0.445 0.656380
domsh        -2.463e-01 1.216e-01 -2.026 0.042808 *
domlong      -2.332e-01 2.030e-01 -1.149 0.250628
snowshoe     7.113e-01 7.371e-01  0.965 0.334516
amsh         3.614e-01 6.765e-01  0.534 0.593216
```

Analysis of Deviance Table

```
Model 1: adopted ~ has_name + outcomeage + intact + female + mixedcolor +
          mixedbreed + orange + white + gray + multi + domsh + domlong +
          snowshoe + amsh
Model 2: adopted ~ has_name + outcomeage + intact + female + mixedcolor +
          mixedbreed + orange + domsh
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      19811    4105.9
2      19817    4109.3 -6   -3.4528   0.7502
```

```
Call:
glm(formula = adopted ~ has_name + outcomeage + intact + female +
     mixedcolor + mixedbreed + orange + domsh, family = binomial(link = "logit"),
     data = cat)

Deviance Residuals:
    Min      1Q  Median      3Q      Max
-3.8451  0.0515  0.0612  0.2395  3.0645

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 4.033e+00 2.044e-01 19.731 < 2e-16 ***
has_name     3.456e+00 1.042e-01 33.155 < 2e-16 ***
outcomeage   -1.217e-03 3.361e-05 -36.214 < 2e-16 ***
intact      -4.883e+00 1.103e-01 -44.259 < 2e-16 ***
female       3.472e-01 8.636e-02  4.021 5.79e-05 ***
mixedcolor    1.616e-01 8.674e-02  1.863 0.06247 .
mixedbreed   -9.309e-01 1.717e-01 -5.423 5.88e-08 ***
orange      -3.071e+00 9.315e-01 -3.296 0.00098 ***
domsh        -2.060e-01 1.046e-01 -1.969 0.04892 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10431.8 on 19825 degrees of freedom
Residual deviance: 4109.3 on 19817 degrees of freedom
AIC: 4127.3
```

Number of Fisher Scoring iterations: 7

Step 4: Add back variables deleted in Step 1

- Only brown and black were eliminated in Step 1 – add these back to see if they contribute to the model when combined with the other variables
- Both were significant – add both to the model

```

Call:
glm(formula = adopted ~ black + has_name + outcomeage + intact +
    female + mixedcolor + mixedbreed + orange + domsh, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-3.8668  0.0488  0.0588  0.2415  3.0685 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 4.111e+00  2.064e-01 19.917 < 2e-16 ***
black       -4.283e-01  1.354e-01 -3.164 0.001556 **  
has_name     3.471e+00  1.046e-01 33.189 < 2e-16 ***  
outcomeage  -1.220e-03  3.367e-05 -36.234 < 2e-16 ***  
intact      -4.896e+00  1.107e-01 -44.241 < 2e-16 ***  
female      3.395e-01  8.647e-02  3.926 8.65e-05 ***  
mixedcolor   6.968e-02  9.232e-02  0.755 0.4580431  
mixedbreed   -9.263e-01  1.718e-01 -5.392 6.97e-08 ***  
orange      -3.167e+00  9.333e-01 -3.393 0.000691 ***  
domsh       -1.866e-01  1.049e-01 -1.780 0.075137 .  
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10431.8 on 19825 degrees of freedom
Residual deviance: 4099.6 on 19816 degrees of freedom
AIC: 4119.6

Number of Fisher Scoring iterations: 7

```

```

Call:
glm(formula = adopted ~ brown + has_name + outcomeage + intact +
    female + mixedcolor + mixedbreed + orange + domsh, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-3.8172  0.0503  0.0601  0.2384  3.0436 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 3.873e+00  2.162e-01 17.917 < 2e-16 ***
brown       2.232e-01  9.964e-02  2.240  0.02508 *  
has_name     3.461e+00  1.043e-01 33.169 < 2e-16 ***  
outcomeage  -1.217e-03  3.361e-05 -36.191 < 2e-16 ***  
intact      -4.891e+00  1.105e-01 -44.249 < 2e-16 ***  
female      3.953e-01  8.921e-02  4.432 9.34e-06 ***  
mixedcolor   1.078e-01  9.020e-02  1.195  0.23193  
mixedbreed   -9.427e-01  1.722e-01 -5.474 4.40e-08 ***  
orange      -2.925e+00  9.316e-01 -3.139  0.00169 **  
domsh       -1.881e-01  1.050e-01 -1.792  0.07319 .  
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10431.8 on 19825 degrees of freedom
Residual deviance: 4104.3 on 19816 degrees of freedom
AIC: 4124.3

Number of Fisher Scoring iterations: 7

```

```

Call:
glm(formula = adopted ~ black + brown + has_name + outcomeage +
    intact + female + mixedcolor + mixedbreed + orange + domsh,
    family = binomial(link = "logit"), data = cat)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-3.8549  0.0492  0.0596  0.2438  3.0614 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 4.049e+00 2.307e-01 17.553 < 2e-16 ***
black       -3.717e-01 1.650e-01 -2.252 0.024305 *  
brown        7.265e-02 1.216e-01  0.598 0.550031    
has_name     3.470e+00 1.046e-01 33.186 < 2e-16 ***
outcomeage  -1.220e-03 3.368e-05 -36.209 < 2e-16 *** 
intact      -4.897e+00 1.107e-01 -44.241 < 2e-16 *** 
female       3.556e-01 9.067e-02   3.921 8.81e-05 *** 
mixedcolor   6.382e-02 9.285e-02   0.687 0.491857    
mixedbreed   -9.308e-01 1.721e-01  -5.409 6.34e-08 *** 
orange       -3.107e+00 9.378e-01  -3.313 0.000922 *** 
domsh        -1.834e-01 1.050e-01  -1.746 0.080768 .  
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10431.8 on 19825 degrees of freedom
Residual deviance: 4099.3 on 19815 degrees of freedom
AIC: 4121.3

Number of Fisher Scoring iterations: 7

```

Step 4: Add black and brown back to the model together

- ❖ Added both brown and black to the model – only black was significant. Proceeded with model with only black added

Step 4: Examine the new model

- With black added, domestic shorthair and mixed color were non-significant – remove from model

```
Call:  
glm(formula = adopted ~ black + has_name + outcomeage + intact +  
    female + mixedcolor + mixedbreed + orange + domsh, family = binomial(link = "logit"),  
    data = cat)  
  
Deviance Residuals:  
    Min      1Q  Median      3Q     Max  
-3.8668   0.0488   0.0588   0.2415   3.0685  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) 4.111e+00 2.064e-01 19.917 < 2e-16 ***  
black       -4.283e-01 1.354e-01 -3.164 0.001556 **  
has_name     3.471e+00 1.046e-01 33.189 < 2e-16 ***  
outcomeage   -1.220e-03 3.367e-05 -36.234 < 2e-16 ***  
intact      -4.896e+00 1.107e-01 -44.241 < 2e-16 ***  
female      3.395e-01 8.647e-02  3.926 8.65e-05 ***  
mixedcolor   6.968e-02 9.232e-02  0.755 0.450431  
mixedbreed   -9.263e-01 1.718e-01 -5.392 6.97e-08 ***  
orange      -3.167e+00 9.333e-01 -3.393 0.000691 ***  
domsh       -1.866e-01 1.049e-01 -1.780 0.075137 .  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 10431.8 on 19825 degrees of freedom  
Residual deviance: 4099.6 on 19816 degrees of freedom  
AIC: 4119.6  
  
Number of Fisher Scoring iterations: 7
```

```
Call:  
glm(formula = adopted ~ black + has_name + outcomeage + intact +  
  female + mixedbreed + orange, family = binomial(link = "logit"),  
  data = cat)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8373	0.0479	0.0579	0.2512	3.0515

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.990e+00	1.828e-01	21.827	< 2e-16 ***
black	-4.722e-01	1.271e-01	-3.715	0.000203 ***
has_name	3.471e+00	1.045e-01	33.219	< 2e-16 ***
outcomeage	-1.211e-03	3.321e-05	-36.468	< 2e-16 ***
intact	-4.893e+00	1.105e-01	-44.263	< 2e-16 ***
female	3.422e-01	8.635e-02	3.963	7.41e-05 ***
mixedbreed	-9.272e-01	1.714e-01	-5.408	6.37e-08 ***
orange	-3.179e+00	9.209e-01	-3.452	0.000557 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 10431.8 on 19825 degrees of freedom  
Residual deviance: 4103.3 on 19818 degrees of freedom  
AIC: 4119.3
```

Number of Fisher Scoring iterations: 7

Step 4: Yet another model

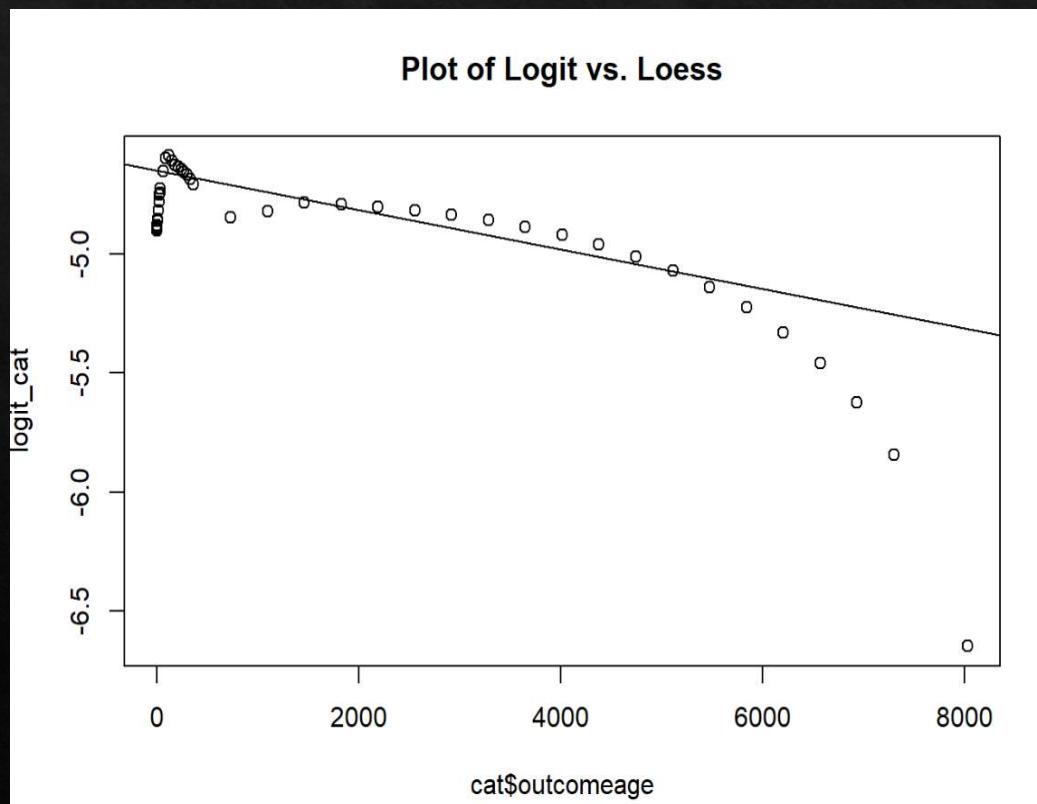
- ◊ LRT comparing model without domestic shorthair and mixed color was non-significant – this model is a better fit
- ◊ AIC for this model was also lower than the previous model – keep this model

Analysis of Deviance Table

```
Model 1: adopted ~ black + has_name + outcomeage + intact + female + mixedcolor +  
  mixedbreed + orange + domsh  
Model 2: adopted ~ black + has_name + outcomeage + intact + female + mixedbreed +  
  orange  
Resid. Df Resid. Dev Df Deviance Pr(>chi)  
1      19816    4099.6  
2      19818    4103.3 -2     -3.616    0.164
```

Step 5: Check linearity of age at outcome with logit of adopted

- ❖ Does not appear to be linear
- ❖ Sharp rise at young age followed by short, sharp drop, gradual decline, and then another more rapid drop
- ❖ Reparametrized into binary categories:
 - ❖ Implied reference: kittens 1 year or younger
 - ❖ age1 - > 365 days (1 year) and <= 913 days (2.5 years)
 - ❖ age2 - > 913 days (2.5 years) and < 5900 days (16 years)
 - ❖ age3 - > 5900 days (16 years)
 - ❖ Estimated based on visual drop offs and rounding to logical ages



Step 5: Add age categories individually to model

- ◊ Added each age category individually to model – all were significant
- ◊ Removed continuous age first

```

Call:
glm(formula = adopted ~ black + has_name + intact + female +
    mixedbreed + orange + age1, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-3.4720  0.1086  0.1251  0.2377  2.5136 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  3.16467   0.15314 20.665 < 2e-16 ***
black       -0.29171   0.11052 -2.640 0.008302 **  
has_name     2.57714   0.07949 32.420 < 2e-16 ***
intact      -3.80753   0.07992 -47.639 < 2e-16 ***
female      0.28322   0.07455  3.799 0.000145 *** 
mixedbreed   -0.89485   0.14763 -6.061 1.35e-09 *** 
orange       -2.25955   0.82840 -2.728 0.006379 **  
age1        -1.57798   0.13740 -11.484 < 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: 10431.8 on 19825 degrees of freedom
Residual deviance: 4281.4 on 19818 degrees of freedom
AIC: 4297.4

Number of Fisher Scoring iterations: 8

Call:
glm(formula = adopted ~ black + has_name + intact + female +
    mixedbreed + orange + age2, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-3.9327  0.0458  0.0528  0.2460  3.3645 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  4.06731   0.18703 21.747 < 2e-16 ***
black       -0.48087   0.12508 -3.844 0.000121 *** 
has_name     3.37867   0.09804 34.462 < 2e-16 ***
intact      -5.06344   0.11681 -43.348 < 2e-16 ***
female      0.28653   0.08455  3.389 0.000702 *** 
mixedbreed   -0.87129   0.17177 -5.073 3.93e-07 *** 
orange       -3.12652   0.92016 -3.398 0.000679 *** 
age2        -4.07562   0.12760 -31.942 < 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: 10431.8 on 19825 degrees of freedom
Residual deviance: 5472.1 on 19818 degrees of freedom
AIC: 5488.1

Number of Fisher Scoring iterations: 7

Call:
glm(formula = adopted ~ black + has_name + intact + female +
    mixedbreed + orange + age3, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-3.4635  0.1111  0.1283  0.1481  2.0289 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  3.08436   0.15053 20.490 < 2e-16 ***
black       -0.28833   0.11062 -2.606 0.009149 **  
has_name     2.62119   0.08047 32.572 < 2e-16 ***
intact      -3.80742   0.07934 -47.986 < 2e-16 ***
female      0.28974   0.07436  3.897 9.75e-05 *** 
mixedbreed   -0.91014   0.14558 -6.252 4.05e-10 *** 
orange       -2.66478   0.74409 -3.581 0.000342 *** 
age3        -5.21570   0.39917 -13.066 < 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: 10431.8 on 19825 degrees of freedom
Residual deviance: 5472.1 on 19818 degrees of freedom
AIC: 5488.1

Number of Fisher Scoring iterations: 7

```

```

Call:
glm(formula = adopted ~ black + has_name + intact + female +
    mixedbreed + orange + age1 + age2 + age3, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-4.1405  0.0306  0.0352  0.2093  3.5687 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 4.71756   0.21198 22.255 < 2e-16 ***
black       -0.50200  0.13240 -3.791  0.00015 ***
has_name     3.57424  0.10255 34.852 < 2e-16 ***
intact      -5.62533  0.14152 -39.749 < 2e-16 ***
female      0.28002  0.08943  3.131  0.00174 **  
mixedbreed  -0.90721  0.18272 -4.965 6.87e-07 ***
orange      -2.01012  1.12370 -1.789  0.07364 .  
age1        -3.33388  0.18945 -17.598 < 2e-16 ***
age2        -4.83111  0.15222 -31.737 < 2e-16 ***
age3        -7.76344  0.42019 -18.476 < 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: 10431.8 on 19825 degrees of freedom
Residual deviance: 3824.8 on 19816 degrees of freedom
AIC: 3844.8

Number of Fisher Scoring iterations: 8

```

Step 5: Add all age categories to the model together

- ❖ Different coefficients for each category consistent with the graph – all ages significant
- ❖ Orange is now close but non-significant - remove

Step 5: Finalized main effects model

- ❖ LRT barely non-significant – Model fits better without orange
- ❖ This is our finalized main effects model.
- ❖ Variables included – black, has name, intact, female, mixed breed, age1, age2, age3

Analysis of Deviance Table

```
Model 1: adopted ~ black + has_name + intact + female + mixedbreed + orange +
  age1 + age2 + age3
Model 2: adopted ~ black + has_name + intact + female + mixedbreed + age1 +
  age2 + age3
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     19816    3824.8
2     19817    3828.4 -1   -3.5756  0.05863 .
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Call:
glm(formula = adopted ~ black + has_name + intact + female +
  mixedbreed + age1 + age2 + age3, family = binomial(link = "logit"),
  data = cat)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-4.1419  0.0306  0.0352  0.2098  3.5709 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  4.71400   0.21226 22.209 < 2e-16 ***
black       -0.49788   0.13244 -3.759  0.00017 ***
has_name     3.58001   0.10252 34.919 < 2e-16 ***
intact      -5.62718   0.14153 -39.759 < 2e-16 ***
female      0.28331   0.08937  3.170  0.00152 **  
mixedbreed   -0.90872   0.18299 -4.966 6.84e-07 ***
age1        -3.35601   0.18871 -17.784 < 2e-16 ***
age2        -4.83530   0.15223 -31.763 < 2e-16 ***
age3        -7.76635   0.42022 -18.482 < 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: 10431.8 on 19825 degrees of freedom
Residual deviance: 3828.4 on 19817 degrees of freedom
AIC: 3846.4
```

Number of Fisher Scoring iterations: 8

Step 6: Interactions

- ❖ Compute all the interactions and add each to the main effects model individually
- ❖ Significant interactions: black:mixedbreed, has_name:intact, has_name:mixedbreed, has_name:age2, intact:mixedbreed, intact:age1, intact:age2

Interaction	Coefficient	p-value
black:mixedbreed	-0.09	0.04
has_name:intact	0.55	0.006
has_name:mixedbreed	0.11	< .001
has_name:age2	0.63	0.005
intact:mixedbreed	-1.04	<0.002
intact:age1	1.10	0.01
intact:age2	1.59	< .001

Step 6: Add in all the interactions

- ❖ Non-significant: Drop
black:mixedbreed, intact:mixedbreed,
intact:age1, intact:age2

```
call:  
glm(formula = adopted ~ black + has_name + intact + female +  
mixedbreed + age1 + age2 + age3 + black:mixedbreed + has_name:intact +  
has_name:mixedbreed + has_name:age2 + intact:mixedbreed +  
intact:age1 + intact:age2, family = binomial(link = "logit"),  
data = cat)  
  
Deviance Residuals:  
    Min      1Q   Median      3Q     Max  
-3.7152   0.0449   0.0512   0.1798   3.6672  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)  5.43076  0.35863 15.143 < 2e-16 ***  
black        -0.95479  0.42965 -2.222 0.026267 *  
has_name      1.05642  0.38778  2.724 0.006444 **  
intact       -5.65821  0.36514 -15.496 < 2e-16 ***  
female        0.26397  0.09191  2.872 0.004077 **  
mixedbreed   -1.31353  0.34272 -3.833 0.000127 ***  
age1          -3.25269  0.20689 -15.722 < 2e-16 ***  
age2          -5.51212  0.21672 -25.434 < 2e-16 ***  
age3          -6.88301  0.44114 -15.603 < 2e-16 ***  
black:mixedbreed  0.49775  0.45207  1.101 0.270873  
has_name:intact  1.50542  0.26075  5.773 7.77e-09 ***  
has_name:mixedbreed  1.46285  0.34604  4.227 2.36e-05 ***  
has_name:age2     1.54145  0.28094  5.487 4.10e-08 ***  
intact:mixedbreed -0.65445  0.34805 -1.880 0.060064 .  
intact:age1       0.78805  0.44527  1.770 0.076756 .  
intact:age2       0.72047  0.48732  1.478 0.139291  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 10431.8 on 19825 degrees of freedom  
Residual deviance: 3747.8 on 19810 degrees of freedom  
AIC: 3779.8  
  
Number of Fisher Scoring iterations: 8
```

Step 6: Inspect the reduced interaction model

- ❖ AIC increased – need to go back and add some interactions back

```

Call:
glm(formula = adopted ~ black + has_name + intact + female +
    mixedbreed + age1 + age2 + age3 + has_name:intact + has_name:mixedbreed +
    has_name:age2, family = binomial(link = "logit"), data = cat)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-3.6691  0.0493  0.0564  0.1865  3.8330 

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)    
(Intercept)                      5.78966   0.27243 21.252 < 2e-16 ***
black                         -0.50842   0.13443 -3.782 0.000156 ***  
has_name                        0.67010   0.36955  1.813 0.069788 .  
intact                          -6.18285   0.18043 -34.267 < 2e-16 ***  
female                          0.27026   0.09212  2.934 0.003349 **  
mixedbreed                      -1.74661   0.23991 -7.280 3.33e-13 ***  
age1                            -3.05264   0.18275 -16.704 < 2e-16 ***  
age2                            -5.47563   0.21448 -25.530 < 2e-16 ***  
age3                            -6.72041   0.42819 -15.695 < 2e-16 ***  
has_name:intact                 1.64660   0.24621  6.688 2.27e-11 ***  
has_name:mixedbreed              1.73003   0.32952  5.250 1.52e-07 ***  
has_name:age2                     1.70031   0.27083  6.278 3.43e-10 ***  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10431.8  on 19825  degrees of freedom
Residual deviance: 3757.5  on 19814  degrees of freedom
AIC: 3781.5

Number of Fisher Scoring iterations: 8

> #ORs and confidence intervals
> cbind(exp(coef(interactions2)),exp(confint(interactions2)))
Waiting for profiling to be done...
                                         2.5 %      97.5 %
(Intercept)                  3.269031e+02 1.928318e+02 5.613608e+02
black                       6.014475e-01 4.639270e-01 7.859992e-01
has_name                      1.954434e+00 9.539660e-01 4.069555e+00
intact                        2.064527e-03 1.436663e-03 2.915603e-03
female                        1.310300e+00 1.093985e+00 1.570033e+00
mixedbreed                    1.743638e-01 1.085681e-01 2.784667e-01
age1                          4.723426e-02 3.295718e-02 6.753316e-02
age2                          4.187592e-03 2.722454e-03 6.316361e-03
age3                          1.206046e-03 5.115180e-04 2.777477e-03
has_name:intact                5.189286e+00 3.197570e+00 8.404422e+00
has_name:mixedbreed            5.640847e+00 2.927351e+00 1.068040e+01
has_name:age2                   5.475628e+00 3.222581e+00 9.330181e+00

```

```

Call:
glm(formula = adopted ~ black + has_name + intact + female +
    mixedbreed + age1 + age2 + age3 + has_name:intact + has_name:mixedbreed +
    has_name:age2 + intact:mixedbreed, family = binomial(link = "logit"),
    data = cat)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-3.6701   0.0488   0.0558   0.1850   3.8372 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  5.30242   0.35080 15.115 < 2e-16 ***
black        -0.50088   0.13459 -3.721 0.000198 *** 
has_name      0.90380   0.38123  2.371 0.017752 *  
intact       -5.54279   0.36085 -15.360 < 2e-16 *** 
female        0.26984   0.09222  2.926 0.003431 ** 
mixedbreed   -1.24270   0.33563 -3.703 0.000213 *** 
age1          -3.05747   0.18251 -16.752 < 2e-16 *** 
age2          -5.44745   0.21271 -25.609 < 2e-16 *** 
age3          -6.71181   0.42842 -15.666 < 2e-16 *** 
has_name:intact  1.65527   0.24658  6.713 1.91e-11 *** 
has_name:mixedbreed 1.50015   0.34648  4.330 1.49e-05 *** 
has_name:age2   1.66954   0.26949  6.195 5.82e-10 *** 
intact:mixedbreed -0.70066   0.34638 -2.023 0.043096 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10431.8 on 19825 degrees of freedom
Residual deviance: 3753.4 on 19813 degrees of freedom
AIC: 3779.4

Number of Fisher Scoring iterations: 8

>
> #see if this model fits better
> anova(interactions2, interactions3, test = "chisq")
Analysis of Deviance Table

Model 1: adopted ~ black + has_name + intact + female + mixedbreed + age1 +
    age2 + age3 + has_name:intact + has_name:mixedbreed + has_name:age2
Model 2: adopted ~ black + has_name + intact + female + mixedbreed + age1 +
    age2 + age3 + has_name:intact + has_name:mixedbreed + has_name:age2 +
    intact:mixedbreed
Resid. Df Resid. Dev Df Deviance Pr(>chi)
1      19814      3757.5
2      19813  3753.4  1   4.0924  0.04308 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step 6: Add back interactions

- ❖ intact:mixedbreed was the closest to significant, so add it back
- ❖ LRT is significant for test between model with and model without intact:mixedbreed, so model fits better with intact:mixedbreed
- ❖ LRT non-significant for test between full interaction model and model with intact:mixedbreed, so intact:mixed breed model fits better

Analysis of Deviance Table

Model 1:	adopted ~ black + has_name + intact + female + mixedbreed + age1 + age2 + age3 + black:mixedbreed + has_name:intact + has_name:mixedbreed + has_name:age2 + intact:mixedbreed + intact:age1 + intact:age2			
Model 2:	adopted ~ black + has_name + intact + female + mixedbreed + age1 + age2 + age3 + has_name:intact + has_name:mixedbreed + has_name:age2 + intact:mixedbreed			
Resid. Df	Resid. Dev	Df	Deviance	Pr(>chi)
1	19810		3747.8	
2	19813	-3	3753.4	-5.6367 0.1307



Hooray, we have
the final model!

```
Call:  
glm(formula = adopted ~ black + has_name + intact + female +  
     mixedbreed + age1 + age2 + age3 + has_name:intact + has_name:mixedbreed +  
     has_name:age2 + intact:mixedbreed, family = binomial(link = "logit"),  
     data = cat)  
  
Deviance Residuals:  
    Min      1Q  Median      3Q     Max  
-3.6701  0.0488  0.0558  0.1850  3.8372  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)  5.30242  0.35080 15.115 <2e-16 ***  
black        -0.50088  0.13459 -3.721 0.000198 ***  
has_name      0.90380  0.38123  2.371 0.017752 *  
intact       -5.54279  0.36085 -15.360 <2e-16 ***  
female        0.26984  0.09222  2.926 0.003431 **  
mixedbreed   -1.24270  0.33563 -3.703 0.000213 ***  
age1          -3.05747  0.18251 -16.752 <2e-16 ***  
age2          -5.44745  0.21271 -25.609 <2e-16 ***  
age3          -6.71181  0.42842 -15.666 <2e-16 ***  
has_name:intact  1.65527  0.24658  6.713 1.91e-11 ***  
has_name:mixedbreed 1.50015  0.34648  4.330 1.49e-05 ***  
has_name:age2   1.66954  0.26949  6.195 5.82e-10 ***  
intact:mixedbreed -0.70066  0.34638 -2.023 0.043096 *  
---  
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 10431.8 on 19825 degrees of freedom  
Residual deviance: 3753.4 on 19813 degrees of freedom  
AIC: 3779.4  
  
Number of Fisher scoring iterations: 8
```

Step 7: Goodness of fit

- ❖ Used Hosmer-Lemeshow goodness of fit test to assess model fit
- ❖ Test is significant – fit is very poor (why?????)
- ❖ Let's check the assumptions first

```
Hosmer and Lemeshow goodness of fit (GOF) test  
data: cat$adopted, fitted(interactions3)  
X-squared = 24.458, df = 8, p-value = 0.00192
```

Assumptions

- ◊ Binary outcome variable: adopted (1= yes, 0 = no)
- ◊ Continuous predictors (age) linear with logit of outcome (adopted): took care of this already
- ◊ No multicollinearity among predictor variables – ran VIFs, nothing over 5, looks good

```
vif(cat_main2)
   black has_name intact female mixedbreed      age1      age2      age3
1.012662  1.308941  2.458383  1.011268  1.033101  1.368032  2.740839  1.129768
```

- ◊ Check for influential values – obtained Cook's distances for each point, checked the ones which were $> 4/n$
 - ◊ 624 values (yikes, that seems like a lot) – maybe this is why the fit is bad?

```
#Check for influential values in continuous variable outcomeage
#anything with a Cook's distance larger than 4/n is suspect
#add Cook's distance to dataframe
cat$cooks <- cooks.distance(interactions3)

#look at the outliers
high_lev <- cat%>%
  filter(cooks > 4/nrow(cat))
high_lev
```

has_name	outcomeage	intact	female	mixedcolor	mixedbreed	black	orange	brown	white
1	6	1	0	0	1	0	0	1	0
1	6570	0	1	1	0	0	0	0	0
0	2920	0	0	1	0	0	0	1	0
0	60	1	0	1	0	0	0	1	0
1	150	0	0	0	0	1	0	0	0
0	.30	1	1	1	0	0	0	0	0
0	120	1	0	0	0	0	0	1	0
0	60	1	1	0	0	0	0	0	0
0	60	1	1	1	0	0	0	1	0
1	120	1	0	1	0	0	0	1	0

1-10 of 624 rows | 1-10 of 21 columns

Previous 1 2 3 4 5 6 ... 63 Next

```

> #try final model again without outliers
> final_no_01 <- glm(formula = adopted ~ black + has_name + intact + female + mixedbreed + age1 + age2 + age3 + ha
s_name:intact + has_name:mixedbreed + has_name:age2 + intact:mixedbreed,
+   family = binomial(link = "logit"), data = no_01)
glm.fit: algorithm did not convergeglm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(final_no_01)

call:
glm(formula = adopted ~ black + has_name + intact + female +
  mixedbreed + age1 + age2 + age3 + has_name:intact + has_name:mixedbreed +
  has_name:age2 + intact:mixedbreed, family = binomial(link = "logit"),
  data = no_01)

Deviance Residuals:
    Min      1Q      Median      3Q      Max 
-8.401e-05  2.100e-08  2.100e-08  2.100e-08  4.004e-05 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 6.333e+01  1.243e+04  0.005   0.996    
black       -6.347e-01  3.428e+03  0.000   1.000    
has_name     3.811e+00  1.418e+04  0.000   1.000    
intact      -8.422e+01  7.405e+04 -0.001   0.999    
female      1.874e-01  2.494e+03  0.000   1.000    
mixedbreed   4.617e-01  1.214e+04  0.000   1.000    
age1        -4.238e+01  4.182e+03 -0.010   0.992    
age2         -8.890e+01  1.478e+04 -0.006   0.995    
age3        -4.806e+01  1.015e+05  0.000   1.000    
has_name:intact 4.479e+01  1.194e+04  0.004   0.997    
has_name:mixedbreed 7.350e-02  1.437e+04  0.000   1.000    
has_name:age2  4.558e+01  1.480e+04  0.003   0.998    
intact:mixedbreed -5.321e+00  7.306e+04  0.000   1.000    

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 7.9265e+03 on 19201 degrees of freedom
Residual deviance: 5.0230e-07 on 19189 degrees of freedom
AIC: 26

Number of Fisher scoring iterations: 25

> cbind(exp(coef(final_no_01)),exp(confint(final_no_01)))
Waiting for profiling to be done...
glm.fit: fitted probabilities numerically 0 or 1 occurredglm.fit: fitted probabilities numerically 0 or 1 occurred
glm.fit: fitted probabilities numerically 0 or 1 occurredglm.fit: fitted probabilities numerically 0 or 1 occurred
glm.fit: fitted probabilities numerically 0 or 1 occurredglm.fit: fitted probabilities numerically 0 or 1 occurred
glm.fit: fitted probabilities numerically 0 or 1 occurredglm.fit: fitted probabilities numerically 0 or 1 occurred
glm.fit: fitted probabilities numerically 0 or 1 occurredglm.fit: fitted probabilities numerically 0 or 1 occurred
glm.fit: fitted probabilities numerically 0 or 1 occurredglm.fit: fitted probabilities numerically 0 or 1 occurred

```

Assumptions: try again without influential points

- ❖ Attempted to run model again without influential values and was planning to run Hosmer-Lemeshow's goodness of fit test to check for improvement
- ❖ Nonsense model: getting rid of the influential values seems to have introduced some separation (see errors and ridiculous standard errors)
- ❖ Thoughts on how to fix this/poor fit?



Model interpretation

Main effects model

With all other variables held constant:

- ❖ Black cats have 0.6 of the odds of being adopted compared to non-black cats (OR = 0.60, CI = [0.47, 0.79]).
- ❖ Cats that are named have 3.59 times the odds of cats without names of being adopted (OR = 3.59, CI = [29.4, 44.0]).
- ❖ Cats who are not spayed/neutered have 0.004 of the odds of cats who are spayed/neutered of being adopted (OR = 0.004, CI = [0.0027, 0.0047]).
- ❖ Female cats have 1.33 times the odds of male cats of being adopted (OR = 1.33, CI = [1.11, 1.58]).
- ❖ Mixed breed cats have 0.40 of the odds of purebred cats of being adopted (OR = 0.40, CI = [0.28, 0.57]).
- ❖ Cats between 1 year old and 2.5 years old have 0.03 of the odds of cats 1 year old or younger of being adopted (OR = 0.03, CI = [0.02, 0.05]).
- ❖ Cats between 2.5 years old and 16 years old have 0.008 of the odds of cats 1 year old or younger of being adopted (OR = 0.008, CI = [0.006, 0.011]).
- ❖ Cats older than 16 years old have 0.0004 of the odds of cats 1 year old or younger of being adopted (OR = 0.0004, CI = [0.0002, 0.0009]).

```
## Call:  
## glm(formula = adopted ~ black + has_name + intact + female +  
##       mixedbreed + age1 + age2 + age3, family = binomial(link = "logit"),  
##       data = cat)  
##  
## Deviance Residuals:  
##      Min        1Q     Median        3Q       Max  
## -4.1419    0.0306   0.0352   0.2098   3.5709  
##  
## Coefficients:  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept)  4.71400   0.21226 22.209 < 2e-16 ***  
## black      -0.49788   0.13244 -3.759  0.00017 ***  
## has_name    3.58001   0.10252 34.919 < 2e-16 ***  
## intact     -5.62718   0.14153 -39.759 < 2e-16 ***  
## female      0.28331   0.08937  3.170  0.00152 **  
## mixedbreed  -0.90872   0.18299 -4.966 6.84e-07 ***  
## age1       -3.35601   0.18871 -17.784 < 2e-16 ***  
## age2       -4.83530   0.15223 -31.763 < 2e-16 ***  
## age3       -7.76635   0.42022 -18.482 < 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: 10431.8  on 19825  degrees of freedom  
## Residual deviance: 3828.4  on 19817  degrees of freedom  
## AIC: 3846.4  
##  
## Number of Fisher Scoring iterations: 8  
  
cbind(exp(coef(cat_main2)),exp(confint(cat_main2)))  
  
## Waiting for profiling to be done...  
  
##                  2.5 %      97.5 %  
## (Intercept) 1.114972e+02 74.157970115 1.704017e+02  
## black      6.078155e-01  0.470442817 7.907877e-01  
## has_name   3.587389e+01 29.415658636 4.397109e+01  
## intact     3.598727e-03  0.002708703 4.719518e-03  
## female     1.327521e+00  1.114375296 1.582117e+00  
## mixedbreed 4.030392e-01  0.279980310 5.736658e-01  
## age1       3.487416e-02  0.024071727 5.047948e-02  
## age2       7.944283e-03  0.005859266 1.064595e-02  
## age3       4.237579e-04  0.000182604 9.621111e-04
```

Final model

- ❖ With all other variables held constant:
 - ❖ Black cats have 0.6 of the odds of being adopted compared to non-black cats (OR = 0.60, CI = [0.47, 0.79]).
 - ❖ Female cats have 1.33 times the odds of male cats of being adopted (OR = 1.33, CI = [1.11, 1.58]).
 - ❖ Cats between 1 year old and 2.5 years old have 0.03 of the odds of cats 1 year old or younger of being adopted (OR = 0.03, CI = [0.02, 0.05]).
 - ❖ Cats older than 16 years old have 0.0004 of the odds of cats 1 year old or younger of being adopted (OR = 0.0004, CI = [0.0005, 0.003]).

		2.5 %	97.5 %
(Intercept)	2.008220e+02	1.031017e+02	4.075050e+02
black	6.060001e-01	4.672989e-01	7.922267e-01
has_name	2.468975e+00	1.172790e+00	5.236925e+00
intact	3.915576e-03	1.915749e-03	7.893414e-03
female	1.309759e+00	1.093321e+00	1.569681e+00
mixedbreed	2.886049e-01	1.472071e-01	5.492576e-01
age1	4.700654e-02	3.281522e-02	6.717970e-02
age2	4.307271e-03	2.810036e-03	6.474745e-03
age3	1.216463e-03	5.157323e-04	2.802833e-03
has_name:intact	5.234507e+00	3.223360e+00	8.484757e+00
has_name:mixedbreed	4.482375e+00	2.254634e+00	8.789735e+00
has_name:age2	5.309748e+00	3.132927e+00	9.023073e+00
intact:mixedbreed	4.962583e-01	2.508556e-01	9.783666e-01

Final model

- ❖ Considering interaction terms:
 - ❖ Cats who are spayed/neutered, purebred, between 2.5 and 16 years old and have a name have 2.47 times the odds of cats with all these qualities who don't have a name (OR = 2.47, CI = [1.17, 5.23]).
 - ❖ Cats who are purebred, have no name, and are not spayed/neutered have 0.004 of the odds of cats with the same background who are spayed/neutered of being adopted (OR = 0.004, CI = [0.002, 0.008]).
 - ❖ Cats who have no name, are spayed/neutered, and are a mixed breed have 0.29 of the odds of purebred cats of a similar background of being adopted (OR = 0.29, CI = [0.15, 0.55]).
 - ❖ And so forth... (it's hard to interpret interaction terms).
 - ❖ Just know that there are significant interactions between having a name and being intact, being a mixed breed, or being between 2.5 and 16 years old and between being intact and being a mixed breed.

		2.5 %	97.5 %
(Intercept)	2.008220e+02	1.031017e+02	4.075050e+02
black	6.060001e-01	4.672989e-01	7.922267e-01
has_name	2.468975e+00	1.172790e+00	5.236925e+00
intact	3.915576e-03	1.915749e-03	7.893414e-03
female	1.309759e+00	1.093321e+00	1.569681e+00
mixedbreed	2.886049e-01	1.472071e-01	5.492576e-01
age1	4.700654e-02	3.281522e-02	6.717970e-02
age2	4.307271e-03	2.810036e-03	6.474745e-03
age3	1.216463e-03	5.157323e-04	2.802833e-03
has_name:intact	5.234507e+00	3.223360e+00	8.484757e+00
has_name:mixedbreed	4.482375e+00	2.254634e+00	8.789735e+00
has_name:age2	5.309748e+00	3.132927e+00	9.023073e+00
intact:mixedbreed	4.962583e-01	2.508556e-01	9.783666e-01

What did we learn?

- ❖ Factors that matter for cat adoption: sex, being black vs. another color, age, intactness, being purebred or not, and having a name
 - ❖ Naming might be something that happens when the cat has been at the shelter for a while, so I wonder if this can really be considered causal.
- ❖ If you're a cat and you want to be adopted, it is better to be:
 - ❖ Female
 - ❖ Not black
 - ❖ 1 year old or younger
 - ❖ Spayed/neutered
 - ❖ Named
 - ❖ Purebred
- ❖ Having a name interacts with the effects of being a mixed breed, being between the ages of 2.5 and 16 years old, and being intact.
- ❖ Being intact interacts with the effects of being a mixed breed.



Implications and future directions

- ❖ These results contribute to the body of literature on factors influencing outcomes for cats in animal shelters. In particular, they examine the odds of adoption based on various cat characteristics.
- ❖ Animal organizations can use this information to become more aware of trends in animal adoption and direct their resources and efforts towards cats who may otherwise be at a disadvantage. Examples of potential action include:
 - ❖ Increasing efforts to spay/neuter cats at the shelter/feral cats in the community
 - ❖ Increasing advertising for cats who don't fit the ideal demographic (i.e., for older, black, male, mixed-breed cats)
- ❖ Future directions:
 - ❖ Analyzing time of outcome in the data (tricky/ran out of time)
 - ❖ Looking more closely at the effects of age, color, and breed
 - ❖ Repeat analysis on data from other shelters – Austin is likely closer to best case scenario than other shelters

The End



References

American Society for the Prevention of Cruelty to Animals. (2019). Retrieved from

<https://www.aspca.org/animal-homelessness/shelter-intake-and-surrender/pet-statistics>

Austin Animal Center Outcomes. (2019). Retrieved from

<https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Outcomes/9t4d-g238>