Austin Animal Center: Dog Statistics

Courtney Schoen

Introduction

The following pages will be an exploration into the statistics surrounding the Austin Animal Center. This is a place in Austin, Texas that takes in lost, stray, or unwanted animals. Through the Austin Animal Center I was provided with two data sets: Intakes and Outcomes. All of the data pertaining to the center can be found at https://data.austintexas.gov/browse?q=austin%20animal%20center&sortBy=relevance&utf8=%E2%9C%93.

Within these sets of data is a large variety of animals and the details of their time at the center. Every time an animal is brought into the shelter the date, location, reason for intake, their condition, and their breed are recorded. Additionally every animal that leaves the shelter the date, where the animal is going/what is happening to the animal, and the breed are recorded. Each animal has a numeric identifier that corresponds to all of this information.

The data also stretches across several years as it begins in October 2013 and ends in August 2019. As a whole there were no specific trends or patterns that stuck out. Even so, by organizing the data in specific ways, the Center staff could use various summaries of the data to better improve the shelter and share an overview with the public.

Load Packages

For this project a variety of packages were used that aided in the exploration of the data in unique ways. Unfamiliar to what has been used in class in the past are flexdashboard and kableExtra. The package flexdashboard allowed for a better presentation of the data and kableExtra allowed tables to be visually edited.

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("janitor")
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library("stringr")
library("flexdashboard")
library("kableExtra")
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

Load the Data

As mentioned, there are two connected data sets utilized. One of the data sets deal with intakes and the other with outcomes. In both sets there is an animal id that uniquely identifies each animal and is consistent across data sets. Having both of the data sets allows for a more full picture look of the animals and the center as one can then see how long an animal is in the shelter, how intake specifics could influence outcome, etc.

```
intakes <- read_csv("austin-animal-center-intakes-1.csv")
```

```
## Rows: 107289 Columns: 12
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (12): Animal ID, Name, DateTime, MonthYear, Found Location, Intake Type,...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
outcomes <- read_csv("austin-animal-center-outcomes-1.csv")
```

```
## Rows: 95367 Columns: 12
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (12): Animal ID, Name, DateTime, MonthYear, Date of Birth, Outcome Type,...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Tidy the Data

At this point I decided to refine my exploration to the animal type dog. So in both intake and outcome I filtered for just dogs and selected only the variables I may work with in order to make the data as clean and simple as possible. I then cleaned the names so that they would be in a format more suitable for work in R. I also had to create a date column as date was in a combined date/time column. I also put the date in the format most suitable work working this time. At this point both of my data sets were relatively tidy and ready to be worked with. Before moving on I wanted to look for any possible repeats among the data. So, I

counted the animal ID and filtered for counts greater than one. There were a variety of animals that were repeated across the data set. However, looking at the data set itself I found that these dogs were entering into the shelter on multiple occasions. Since the animal was not necessarily repeated but readmitted, I decided to count the repeats as individuals.

```
dog_intakes <- intakes %>%
  filter(`Animal Type` == "Dog") %>%
  select(`Animal ID`, DateTimeIntake = DateTime,
         `Intake Type`, `Intake Condition`,
         Breed, `Found Location`) %>%
  clean_names() %>%
  separate(date_time_intake,
           into = c("intake_date"),
           sep = " ", remove = FALSE) %>%
  mutate(intake_date = as.Date(intake_date, format = "%m / %d / %Y"))
```

```
## Warning: Expected 1 pieces. Additional pieces discarded in 60813 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
dog_outcomes <- outcomes %>%
  filter(`Animal Type` == "Dog") %>%
  select(`Animal ID`, DateTimeOutcome = DateTime,
         `Outcome Type`, `Outcome Subtype`, Breed) %>%
  clean_names() %>%
  separate(date_time_outcome,
           into = c("outcome_date"),
           sep = " ", remove = FALSE) %>%
  mutate(outcome_date = as.Date(outcome_date, format = "%m / %d / %Y"))
```

```
## Warning: Expected 1 pieces. Additional pieces discarded in 54327 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

Analyze the Data

The main goal of the following analysis of the data is to provide a resource for the Center to provide to the public. I will create an easily accessible page that allows people to get a quick glance at the level of success and the level of failure occurring at the shelter. The tables and graphs are intended to be as simple as possible given the audience being the common individual, most of which have limited statistical and mathematical knowledge. With this, the Center could potentially gain support and funding from those who access the data.

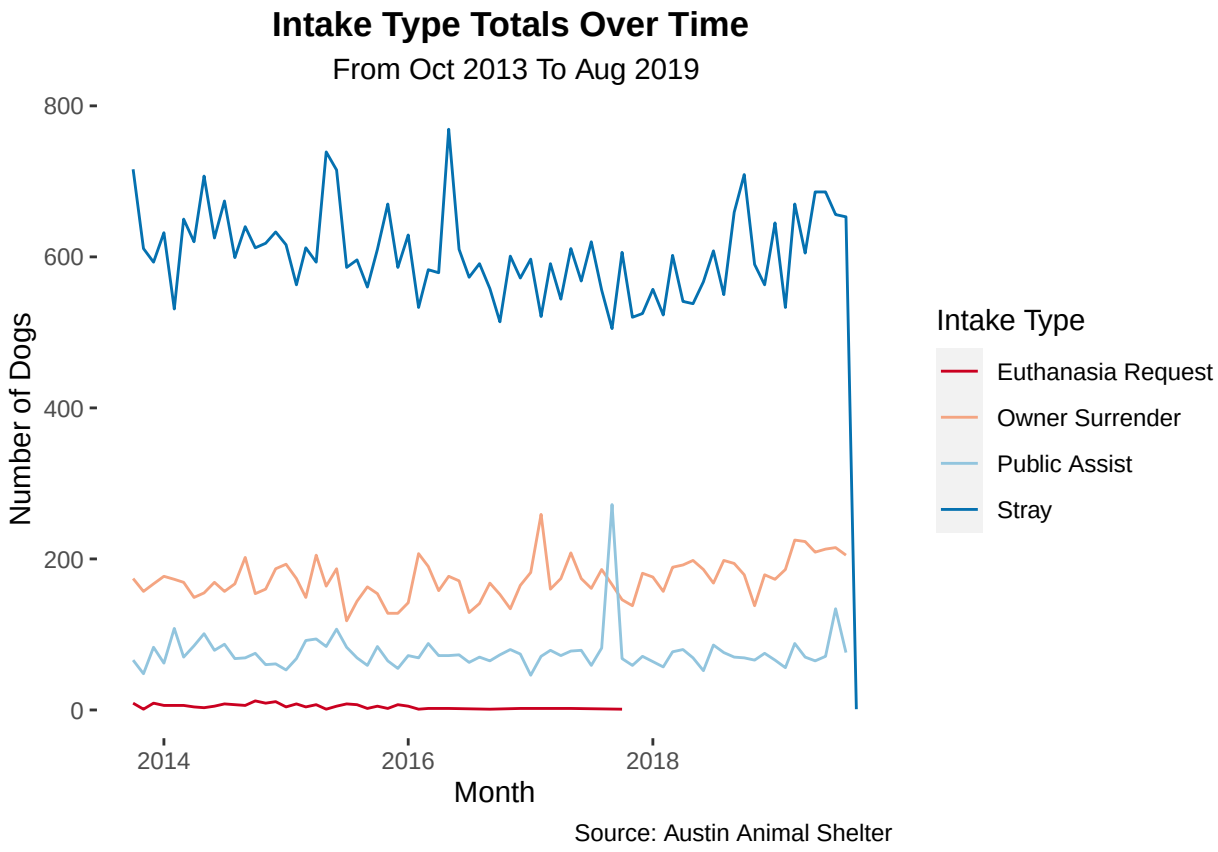
Overall Trend

The first part of my data summary consists of a look at the overall trends of dogs coming into the shelter, dogs leaving the shelter, and the reasons for doing so. Both graphs look at the count of each type of intake and each type of outcome per month over the course of 5 years.

Looking at intake we can see that there is around an average of 500 stray dogs coming into the shelter and this is the highest sub type for intake. We also see that there are owner surrenders and public assists with a very low number of euthanasia requests. This gives the viewer an idea of the overall case load and the reasons for which animals are entering.

Looking at the outcome we can see that there is around an average of 400 dogs adopted from the shelter per month which is the highest outcome. We also see that there is an overall increasing trend in adoption but also some drastic dips and spikes. There are also a rather high amount of dogs transferred to other shelters and dogs who are returned to their owners. There are a variety of other reasons but they are all very low and rather insignificant in terms of the overall trend. With this information and a glimpse into the overall workings of the shelter, we can move onto more specific ideas.

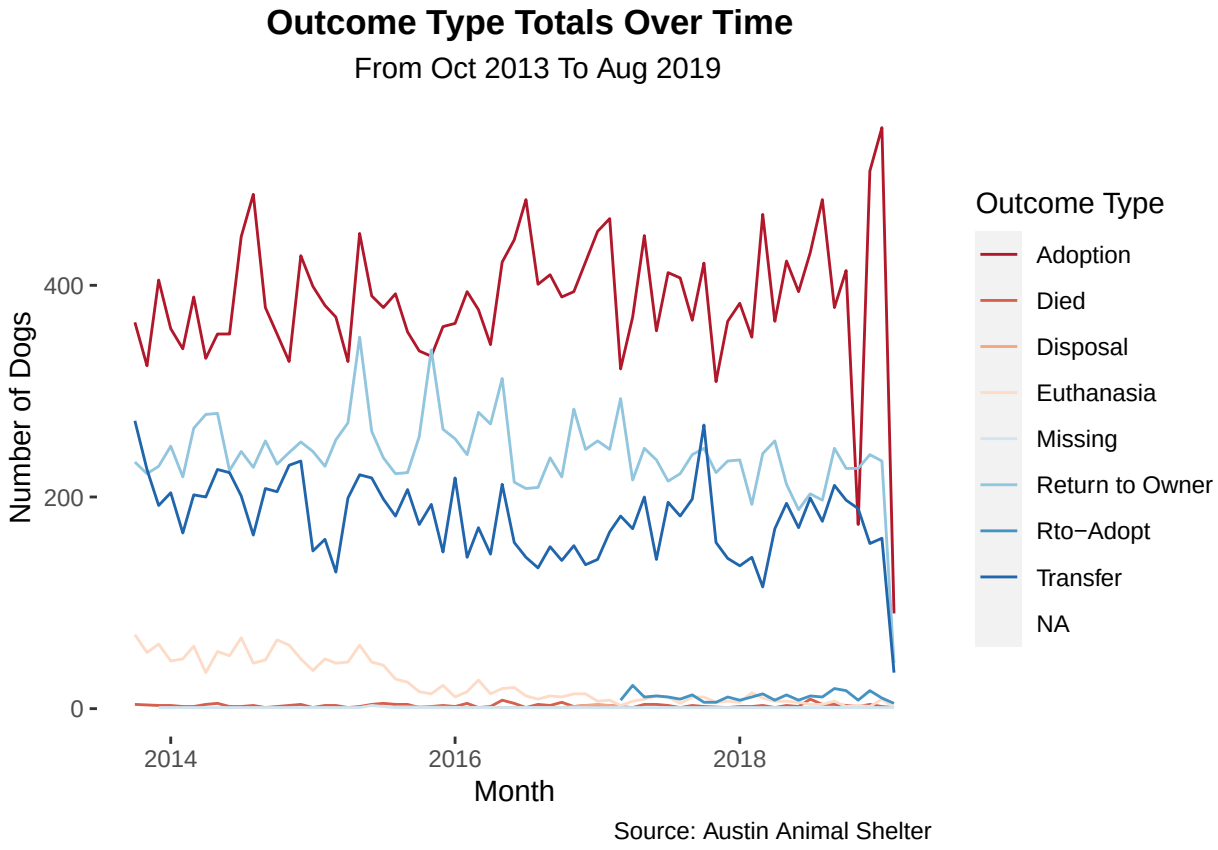
```
dog_intakes %>%
  mutate(date = round_date(intake_date, unit = "month")) %>%
  group_by(date) %>%
  count(intake_type) %>%
  ggplot(aes(date, n, color = intake_type)) +
  geom_line() +
  scale_color_brewer(palette = "RdBu") +
  labs(title = "Intake Type Totals Over Time",
       subtitle = "From Oct 2013 To Aug 2019",
       x = "Month",
       y = "Number of Dogs",
       color = "Intake Type",
       caption = "Source: Austin Animal Shelter")  +
  theme(plot.title = element_text(face = "bold", hjust = .55),
        plot.subtitle = element_text(hjust = .55),
        panel.background = element_blank())
```




```
dog_outcomes %>%
  mutate(date = round_date(outcome_date, unit = "month")) %>%
  group_by(date) %>%
  count(outcome_type) %>%
  ggplot(aes(date, n, color = outcome_type)) +
  geom_line() +
  scale_color_brewer(palette = "RdBu") +
  labs(title = "Outcome Type Totals Over Time",
       subtitle = "From Oct 2013 To Aug 2019",
       x = "Month",
```

```
       y = "Number of Dogs",
       color = "Outcome Type",
       caption = "Source: Austin Animal Shelter")  +
  theme(plot.title = element_text(face = "bold", hjust = .55),
        plot.subtitle = element_text(hjust = .55),
        panel.background = element_blank())
```

```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```



Dogs Entering Homes

The next portion is an exploration into the Center's successes. When dogs enter an animal shelter the goal is for them to enter into a forever home, whether that be getting returned to their owner or being adopted. As a result the success of the shelter can be represented by the proportion of the total number of dogs who come into the shelter that enter into a home. To represent this data I produced a bar graph that looks at the percent of dogs that enter home per quarter, per year. I chose to use percent as opposed to proportion as that is more readily understood. While I expected to see some level of quarterly trend. Even so, we see that around an average of 75% of dogs are going into homes. This is a very good statistic for the Center considering that every months around 500 dogs enter the shelter and so every 3 months there are around 1500 dogs and 75% of them are entering home. Additionally looking at the trend graph there are a considerable amount of dogs also being sent to other shelters where they could still successfully find a home.

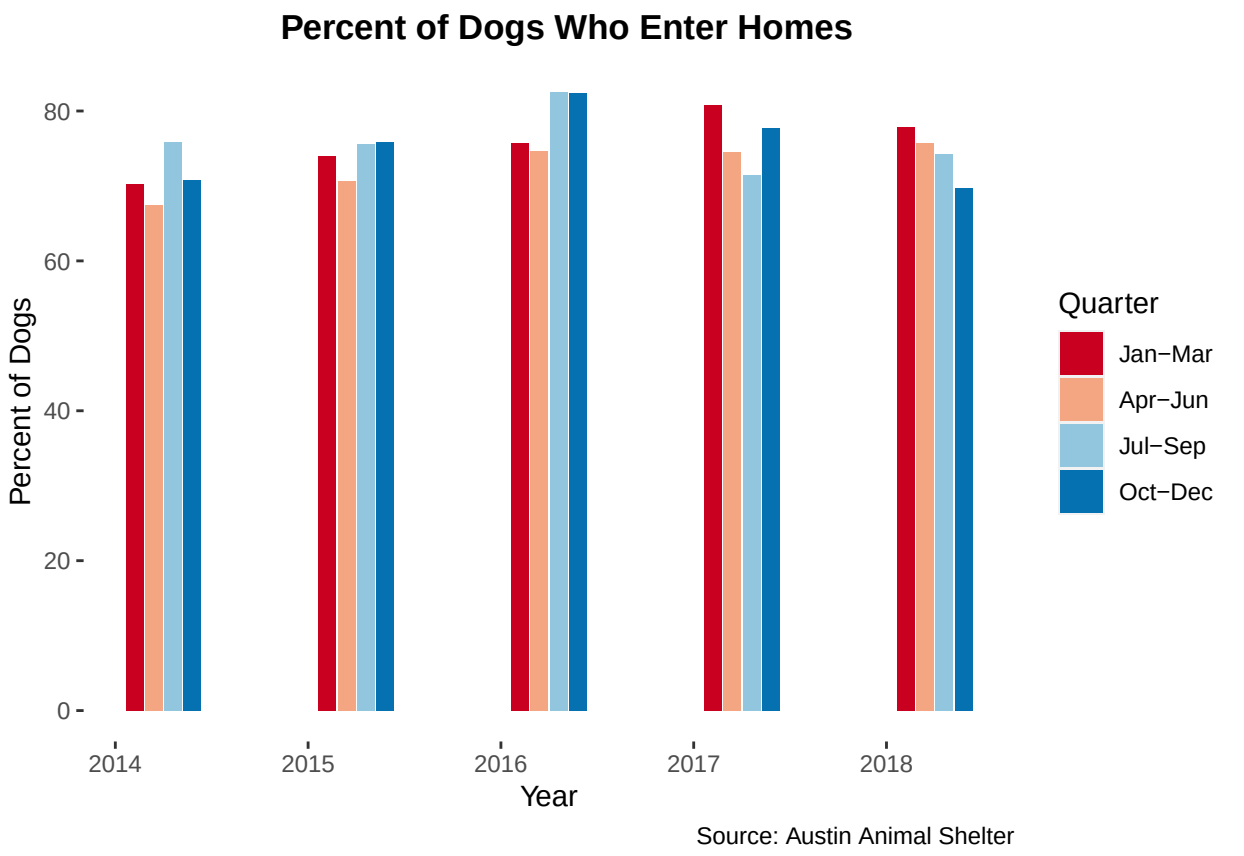
```
intake_per_quarter <- dog_intakes %>%
  mutate(date = round_date(intake_date, unit = "month"),
         date_quarter = quarter(date, with_year = TRUE)) %>%
  filter(date_quarter != 2013.4,
         date_quarter != 2019.1,
```

```
         date_quarter != 2019.2,
         date_quarter != 2019.3) %>%
  group_by(date_quarter) %>%
  count(date_quarter) %>%
  mutate(Quarter = str_sub(date_quarter, -1))

outcome_per_quarter <- dog_outcomes %>%
  mutate(date = round_date(outcome_date, unit = "month"),
         date_quarter = quarter(date, with_year = TRUE)) %>%
  filter(date_quarter != 2013.4,
         date_quarter != 2019.1,
         date_quarter != 2019.2,
         date_quarter != 2019.3) %>%
  filter(outcome_type == "Adoption" | outcome_type == "Return to Owner") %>%
  group_by(date_quarter) %>%
  count(date_quarter) %>%
  mutate(Quarter = str_sub(date_quarter, -1))

quarter_results <- intake_per_quarter %>%
  full_join(outcome_per_quarter, by = "date_quarter") %>%
  mutate(quarter_proportion = n.y / n.x) %>%
  mutate(quarter_percent = quarter_proportion*100)

ggplot(quarter_results) +
  geom_col(aes(date_quarter, quarter_percent, fill = Quarter.x)) +
  scale_fill_brewer(palette = "RdBu",
                    labels = c("Jan-Mar", "Apr-Jun", "Jul-Sep", "Oct-Dec")) +
  labs(title = "Percent of Dogs Who Enter Homes",
       x = "Year",
       y = "Percent of Dogs",
       fill = "Quarter",
       caption = "Source: Austin Animal Shelter") +
  theme(plot.title = element_text(face = "bold", hjust = .55),
        panel.background = element_blank())
```



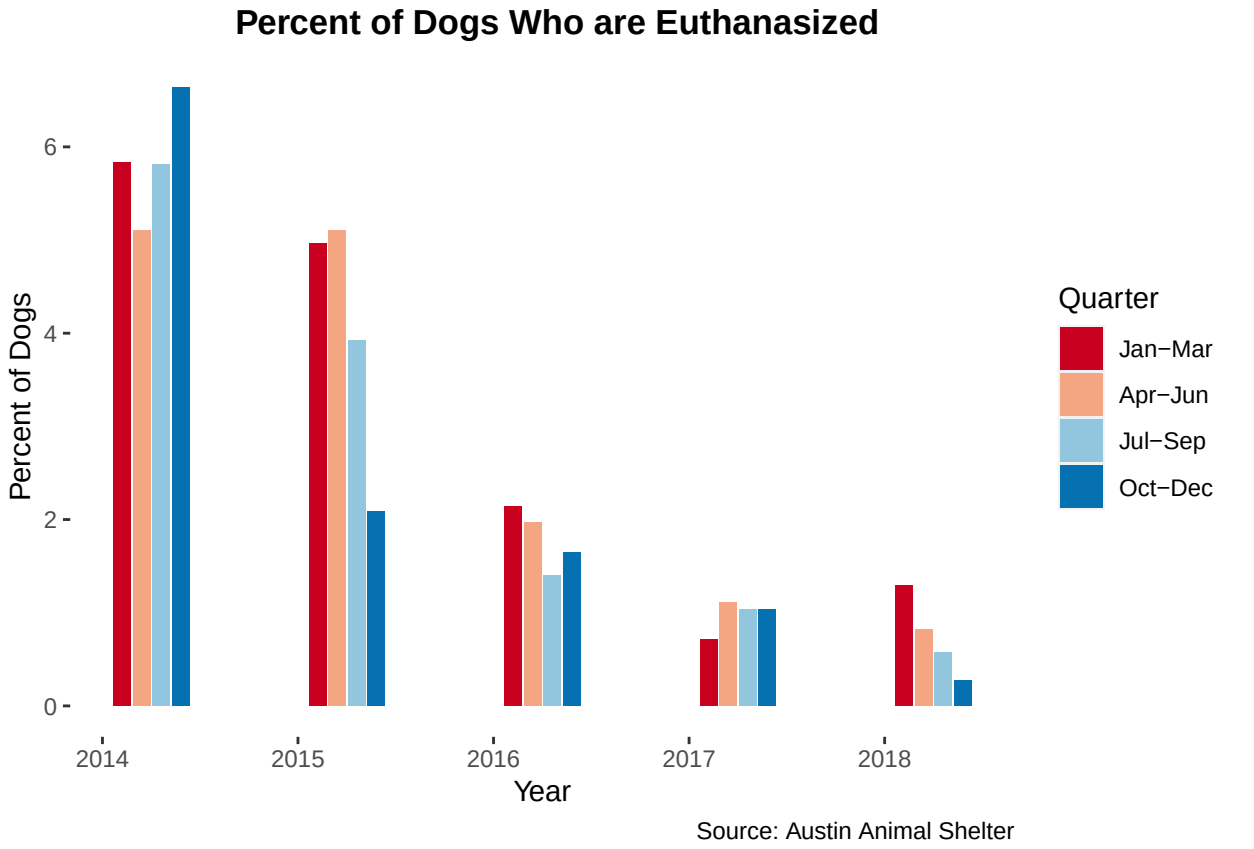
Dogs Being Euthanized

I feel to give an accurate representation of the shelter it is important to represent both successes and failures. While successes are marked by dogs entering homes, failures are marked by dogs never exiting the Center. For that reason I created another bar chart that looked at the proportion of the total number of dogs each quarter, each month who were euthanized as a percentage. Again we see really no quarterly trend but we do see an overall downward trend as year increases. We also see that even at it's highest, euthanasia was only 6% and in most recent years is under 2%. With the numbers low the next step is looking at why dogs are being euthanized and if it is at all breed specific.

```
euthanasia_per_quarter <- dog_outcomes %>%
  mutate(date = round_date(outcome_date, unit = "month"),
         date_quarter = quarter(date, with_year = TRUE)) %>%
  filter(date_quarter != 2013.4,
         date_quarter != 2019.1,
         date_quarter != 2019.2,
         date_quarter != 2019.3) %>%
  filter(outcome_type == "Euthanasia") %>%
  group_by(date_quarter) %>%
  count(date_quarter) %>%
  mutate(Quarter = str_sub(date_quarter, -1))

euthanasia_quarter_results <- euthanasia_per_quarter %>%
  full_join(intake_per_quarter, by = "date_quarter") %>%
  mutate(quarter_proportion = n.x / n.y) %>%
  mutate(quarter_percent = quarter_proportion*100)
```

```
ggplot(euthanasia_quarter_results) +
  geom_col(aes(date_quarter, quarter_percent, fill = Quarter.x)) +
  scale_fill_brewer(palette = "RdBu",
                    labels = c("Jan-Mar", "Apr-Jun", "Jul-Sep", "Oct-Dec")) +
  labs(title = "Percent of Dogs Who are Euthanasized",
       x = "Year",
       y = "Percent of Dogs",
       fill = "Quarter",
       caption = "Source: Austin Animal Shelter") +
  theme(plot.title = element_text(face = "bold", hjust = .55),
        panel.background = element_blank())
```

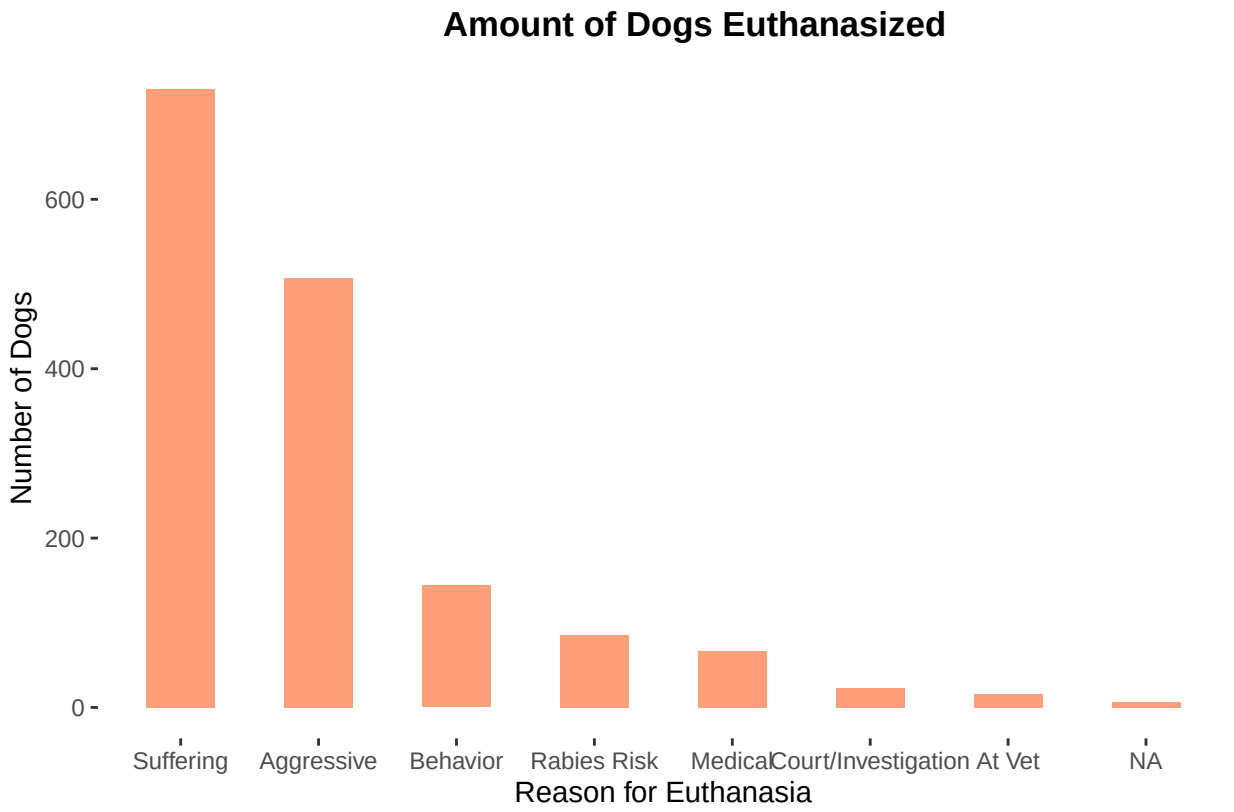


Reasons for Euthanasia

While the level of euthanasia is low, it is important to verify that there is no underlying negative trend. In order to do that I created a bar plot of the count for all reasons for euthanasia. This showed that almost all accounts are are due to health or high levels of aggression. With this there is still no reason for concern. Another possible concern is that euthanasia is breed specific so I created a chart that looks at at the percent of all dogs of a breed that are euthanized. Even looking at only the 6 highest percentages we do not see any disturbing or concerning statistics. Thus, the failures of the Center were not only low, but had valid reasoning.

```
dog_outcomes %>%
  filter(outcome_type == "Euthanasia") %>%
  group_by(outcome_subtype) %>%
  ggplot(aes(fct_infreq(outcome_subtype))) +
  geom_bar(fill = "lightsalmon1", width = 0.5) +
```

```
  labs(title = "Amount of Dogs Euthanasized",
       x = "Reason for Euthanasia",
       y = "Number of Dogs",
       caption = "Source: Austin Animal Shelter") +
  theme(panel.background = element_blank(),
        plot.title = element_text(face = "bold", hjust = .55))
```




```
breed_stats <- dog_outcomes %>%
  group_by(breed) %>%
  count(outcome_type) %>%
  filter(outcome_type == "Adoption" | outcome_type == "Euthanasia")

adopted <- breed_stats %>%
  filter(outcome_type == "Adoption") %>%
  arrange(desc(n))

euthanasized <- breed_stats %>%
  filter(outcome_type == "Euthanasia") %>%
  arrange(desc(n))

total_breed <- dog_outcomes %>%
  group_by(breed) %>%
  count(breed) %>%
  filter(n > 100)

outcome_proportions <- adopted %>%
  full_join(euthanasized, by = "breed") %>%
```

Dog Breed	Total Euthanized	Breed Total	Percent Euthanized
Chow Chow Mix	19	153	12.418301
Pit Bull	22	273	8.058608
Pit Bull/Labrador Retriever	8	128	6.250000
American Staffordshire Terrier Mix	15	247	6.072875
Pit Bull Mix	451	7495	6.017345
Rottweiler Mix	23	385	5.974026

```
  inner_join(total_breed, by = "breed") %>%
  mutate(adopted_proportion = n.x / n,
         euthanasia_percent = (n.y / n)*100) %>%
  filter(n > 100) %>%
  select("Dog Breed" = breed,
         "Total Euthanized" = n.y,
         "Breed Total" = n,
         "Percent Euthanized" = euthanasia_percent) %>%
  arrange(desc(`Percent Euthanized`)) %>%
  head(6L)
outcome_proportions %>%
  kbl() %>%
  kable_paper("hover", full_width = F)
```

Final Result

My final result consists of a link that organizes these findings and is easily accessible. Thank link can be found here: https://rpubs.com/cs8445/766157.

Limitations: In this data, I am fully relying on the fact that the shelter has accurately provided data. Additionally when trying to join the data I found that there are not the same number of dogs in the intake data as in the outcome data. This must mean that in some place there are dogs missing. This not only questions the validity of the data but since I did not account for it, it also would throw off the statistics I found. Another problem is the y axis scales of the euthanasia percent graph and the percent of dogs in homes graph is very different. The one involving dogs in homes has a scale from 0% to 80% where the one involving euthanasia has a scale from 0% to 6%. Without explanation or a close eye, the euthanasia statistics look much higher than in reality which could result very negatively for the shelter.