Final_Project_Submission_Code

Jonathan Lawrence

01 June, 2019

OVERVIEW I intend to identify:
o Q1: What type of animal receives the most incidents at the shelter?
o Q2: What breed of dog has the most incidents?
o Q3: What type of intake does the shelter see the most?
o Q4: What type of outtake does the shelter see the most?
o Q5: During which time of the year does the shelter receive the most animals?
o Q6: Are the majority of the animals taken in healthy or sick?

Q1: What type of animal receives the most incidents at the shelter?

```r
# Set up pie chart
data <- table(dat_org$Animal.Type)

# Histogram of animal types. Use column labeled 'Animal.Type'

ggplot(dat_org, aes(x = Animal.Type)) +
  geom_histogram(stat = "count", color = "black", fill =
rgb(0.2,0.7,0.1,0.4)) +
  labs(title = "Incidents Reported by ANIMAL TYPE\n in 2018", x = "Animal
Type", y = "# of Incidents")
```
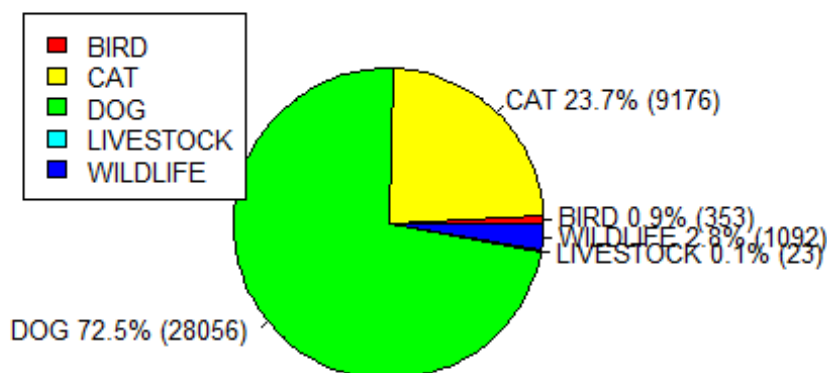
Incidents Reported by ANIMAL TYPE in 2018
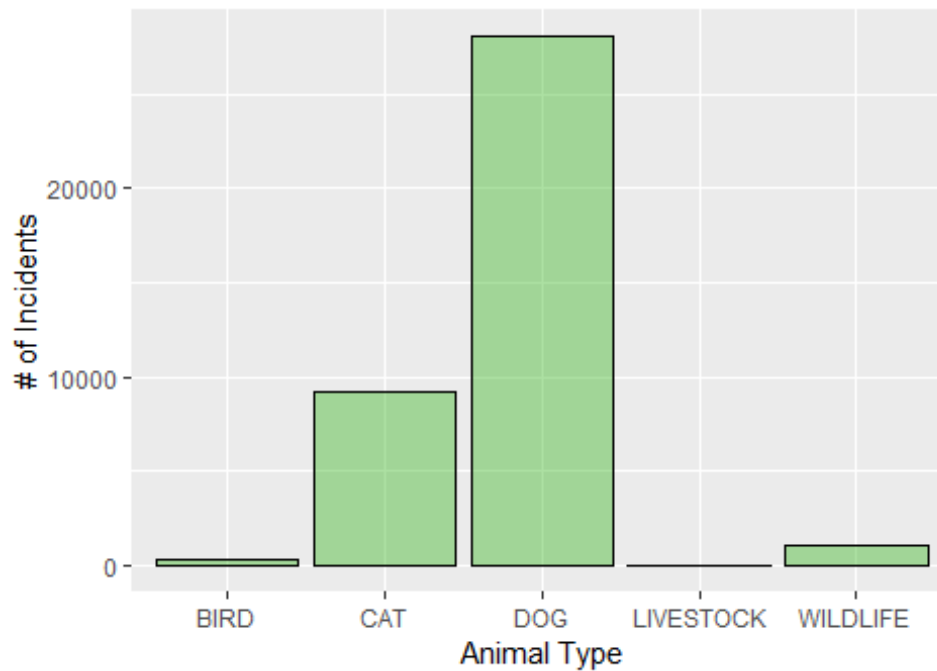

```r
# Pie chart of animal types. Use column labeled 'Animal.Type'
pie(table(dat_org$Animal.Type), main = "Incidents Reported by ANIMAL TYPE\n
in 2018", col = rainbow(6), labels =
paste(c("BIRD","CAT","DOG","LIVESTOCK","WILDLIFE"), " ",
round(prop.table(table(dat_org$Animal.Type))*100,1), "%", sep = "", " (",
table(dat_org$Animal.Type),  ") "), cex = 0.8)

legend("topleft", c("BIRD","CAT","DOG","LIVESTOCK","WILDLIFE"), cex = 0.8,
fill = rainbow(6))
```

Incidents Reported by ANIMAL TYPE in 2018

Legend:
- BIRD
- CAT
- DOG
- LIVESTOCK
- WILDLIFE

CAT 23.7% (9176)
BIRD 0.9% (353)
WILDLIFE 2.8% (1092)
LIVESTOCK 0.1% (23)
DOG 72.5% (28056)

ANSWER: In 2018, the data reported by the shelter indicates that dog-related incidents were 72.5% of all incidents for that year. This number is significantly higher than cats which were the second most reported, coming in at only 23.7%. Livestock-related incidents are so rare that the corresponding slice of the pie chart is very difficult to see, considering it only makes up 0.1% of the total.

Q2: What breed of dog has the most incidents?

```
# Subset the data for DOG only.

dat_org_DOG <- subset(dat_org,  Animal.Type == "DOG")

# Eliminate breeds that have less than 500 incident reports.
dat_org_DOG_500 <- dat_org_DOG %>%
  group_by(Animal.Breed) %>%
  filter(n() > 500)

# Plot 'Animal.Breed' vs # of Incidents greater than 500

ggplot(dat_org_DOG_500, aes(x = Animal.Breed)) +
  geom_histogram(stat = "count", color = "black", fill =
rgb(0.2,0.7,0.1,0.4), binwidth = 50) +
  labs(title = "Dog Breeds Reported in 2018", x = "Dog Breed", y = "# of
Incidents") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Dog Breeds Reported in 2018
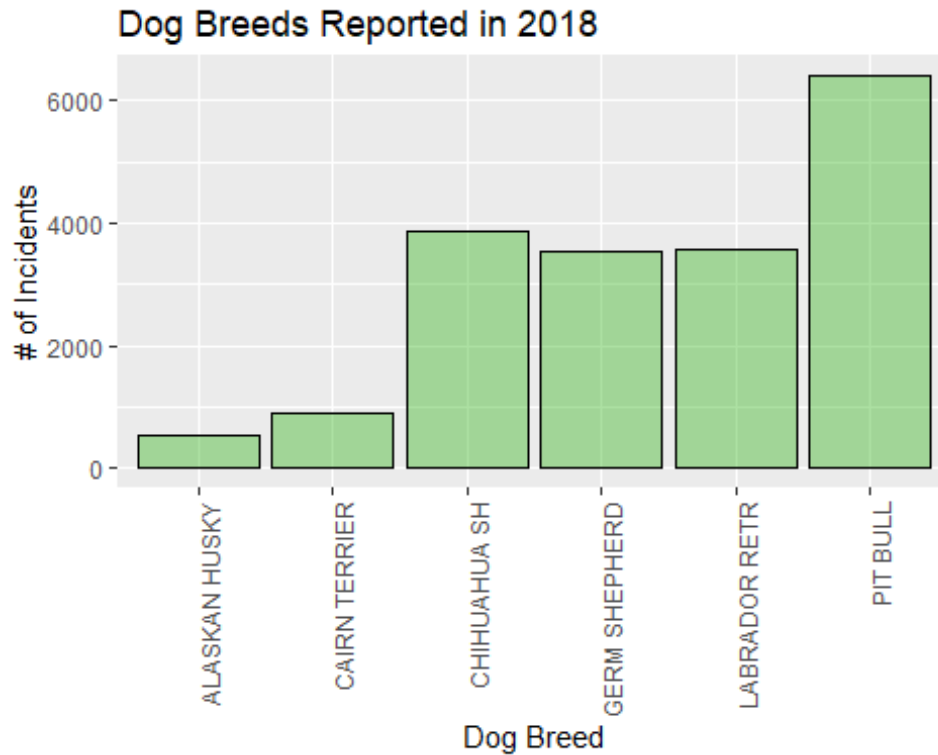
ANSWER: The data shows that Pit Bulls had the highest number of incidents reported to the shelter in 2018, with just over 6000. The second highest, Chihuahua Shorthair, was just under 4000. That is a difference of more than 2000 incidents reported for Pit Bulls over the second, third, and forth place breeds.
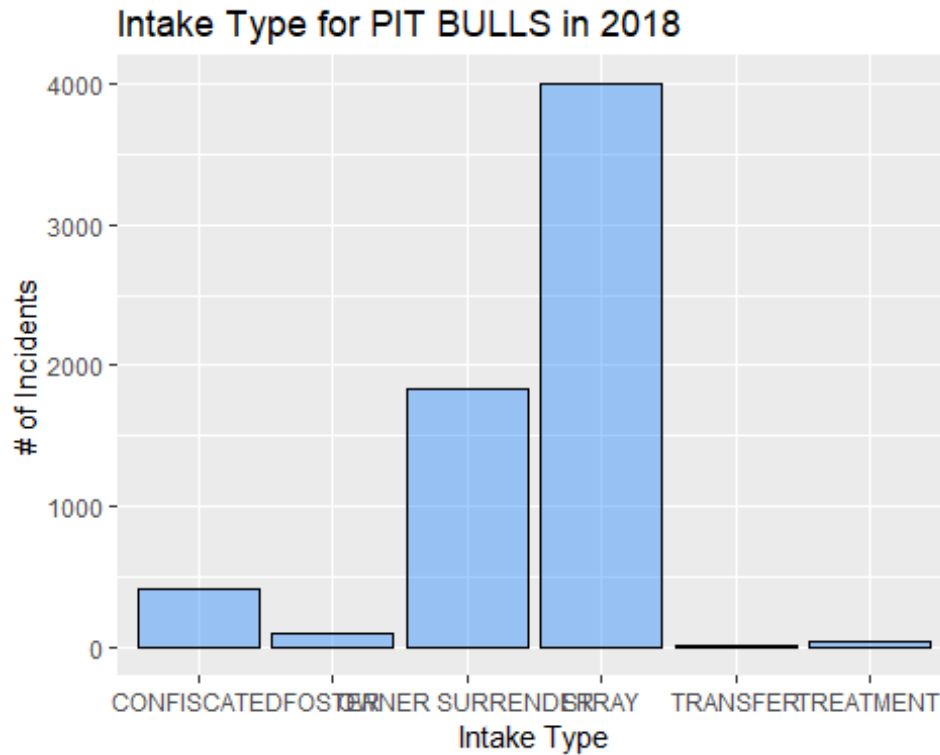
Q3: What type of intake does the shelter see the most?

```
# Most common intake type for all dog breeds
ggplot(dat_org_DOG, aes(x = Intake.Type)) +
  geom_histogram(stat = "count", color = "black", fill =
rgb(0.2,0.7,0.1,0.4)) +
  labs(title = "Intake Type in 2018", x = "Intake Type", y = "# of
Incidents")
```

Intake Type in 2018
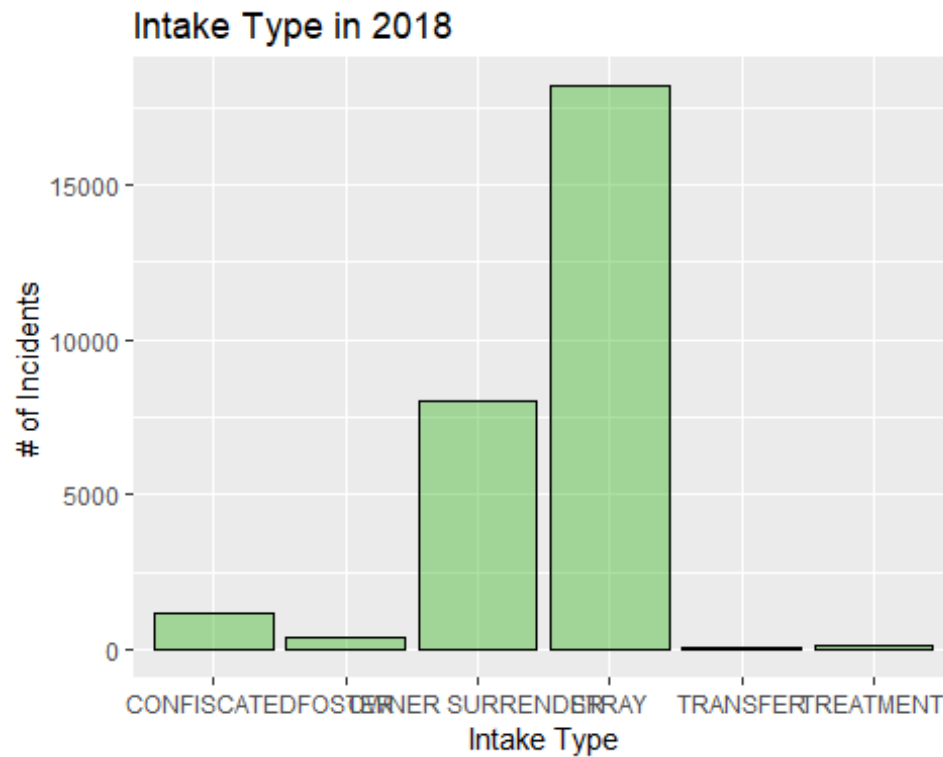

```r
# Filter data to just show pit bulls
dat_org__DOG_PITBULL <- subset(dat_org,  Animal.Breed == "PIT BULL")

# Most common intake type for all dog breeds
ggplot(dat_org_DOG, aes(x = Intake.Type)) +
  geom_histogram(data = dat_org__DOG_PITBULL, stat = "count", color =
"black", fill = rgb(0.1,0.5,1,0.4)) +
  labs(title = "Intake Type for PIT BULLS in 2018", x = "Intake Type", y = "#
of Incidents")
```
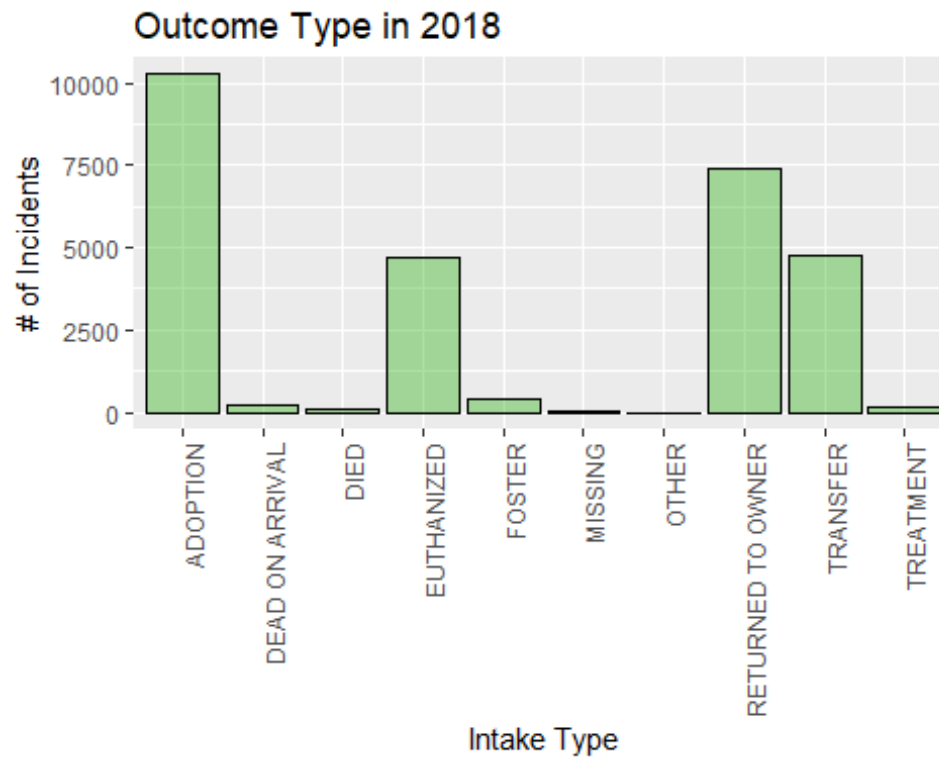
Intake Type for PIT BULLS in 2018

ANSWER: The shelter reported more incidents related to strays than any other type of intake. Strays were more than twice the number of owner surrenders which was the second most reported incident. Comparing the graphs for intake type between 'All Dogs' and 'Pit Bulls', it is apparent that the graphs look the same, albeit different values on the y-axis. From this, we can safely assume that the intake type of pit bulls will be a good representation of the total intakes from all dogs.

Q4: What type of outtake does the shelter see the most?

```
# Most common outcome type for all dog breeds
ggplot(dat_org_DOG, aes(x = Outcome.Type)) +
  geom_histogram(stat = "count", color = "black", fill =
rgb(0.2,0.7,0.1,0.4)) +
  labs(title = "Outcome Type in 2018", x = "Intake Type", y = "# of
Incidents") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Outcome Type in 2018


```
# Most common outcome type for all dog breeds
ggplot(dat_org_DOG, aes(x = Outcome.Type)) +
  geom_histogram(data = dat_org__DOG_PITBULL, stat = "count", color =
"black", fill = rgb(0.1,0.5,1,0.4)) +
  labs(title = "Outcome Type for PIT BULLS\n in 2018", x = "Outcome Type", y
= "# of Incidents") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Outcome Type for PIT BULLS in 2018
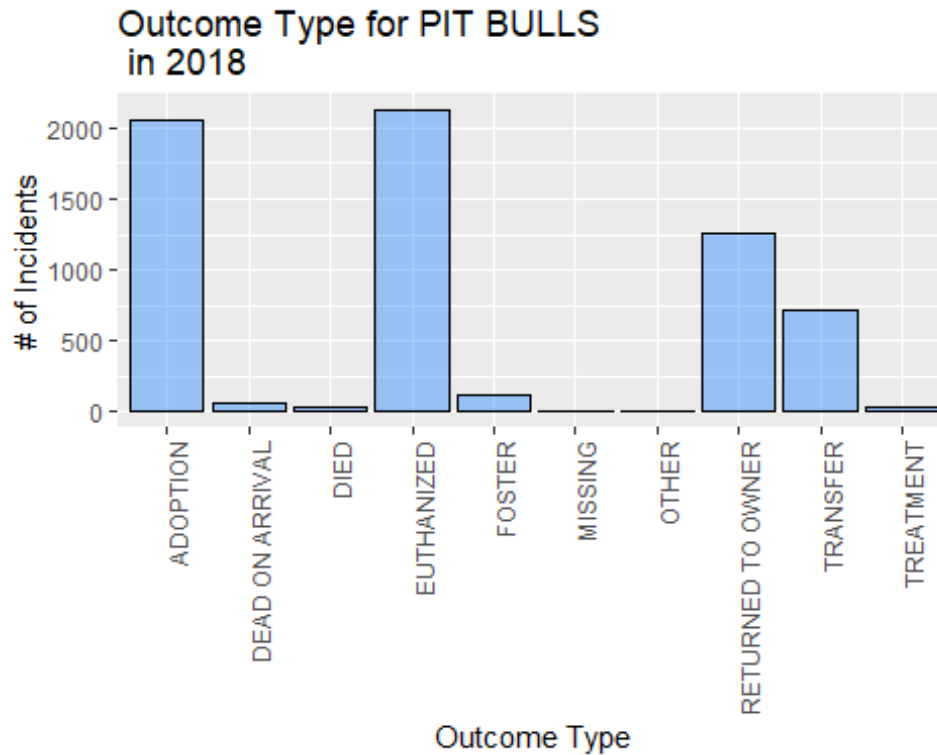
ANSWER: The data reveals that overall more dogs are adopted than any other outcome type. When looking solely at the breed with the most incidents reported, Pit Bulls are very slightly more likely to be euthanized than adopted.

Q5: During which time of the year does the shelter receive the most animals?

```
# Reorganize Month factors into proper order
dat_org_DOG$Month <- factor(dat_org_DOG$Month, levels =
c("JAN.2018","FEB.2018","MAR.2018","APR.2018","MAY.2018","JUN.2018","JUL.2018
","AUG.2018","SEP.2018","OCT.2018","NOV.2018","DEC.2018"))
levels(dat_org_DOG$Month)

## [1] "JAN.2018" "FEB.2018" "MAR.2018" "APR.2018" "MAY.2018" "JUN.2018"
## [7] "JUL.2018" "AUG.2018" "SEP.2018" "OCT.2018" "NOV.2018" "DEC.2018"

# Plot months with a histogram for best visual (gree = all dogs)
ggplot(dat_org_DOG, aes(x = Month)) +
  geom_histogram(stat = "count", binwidth = "0.1", color = "white", fill =
rgb(0.2,0.7,0.1,0.4)) +
  labs(title = "Incident Reported by Month in 2018") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Incident Reported by Month in 2018
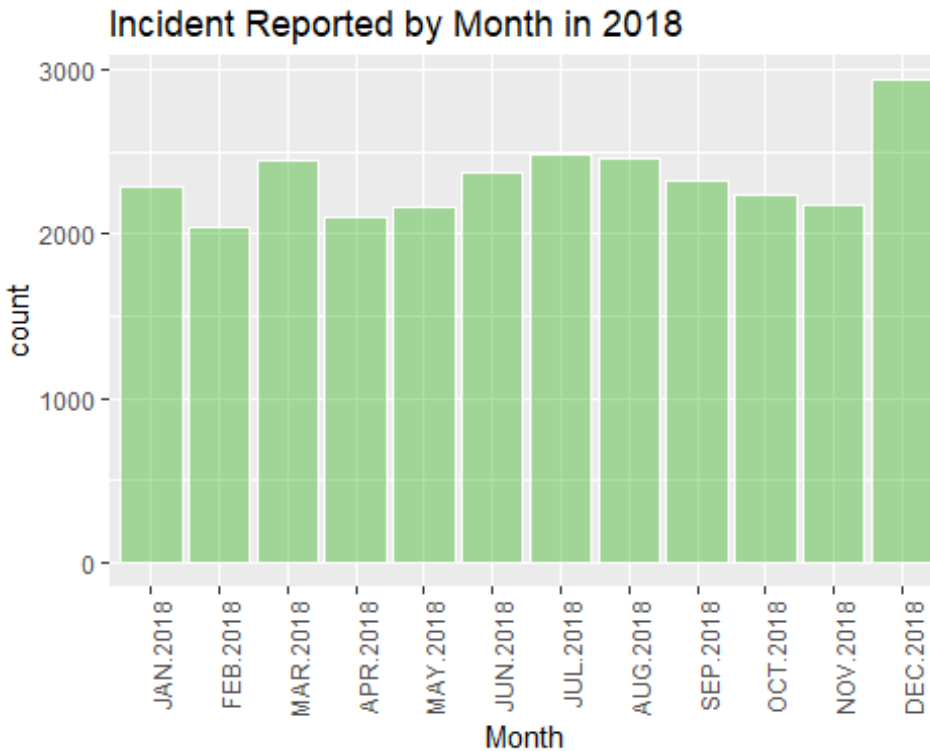
```
# Plot months with a histogram for best visual (green = all dogs, blue = pit
bulls)
ggplot(dat_org_DOG, aes(x = Month)) +
  geom_histogram(stat = "count", binwidth = "0.1", color = "white", fill =
rgb(0.2,0.7,0.1,0.4)) +
  labs(title = "Incident Reported by Month in 2018") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  geom_histogram(data = dat_org__DOG_PITBULL, stat = "count", binwidth =
"0.1", color = "white", fill = rgb(0.1,0.5,1,0.4)) +
  labs(title = "Incidents Reported by Month in 2018\n green = all dogs, blue
= pit bulls") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Incidents Reported by Month in 2018
green = all dogs, blue = pit bulls
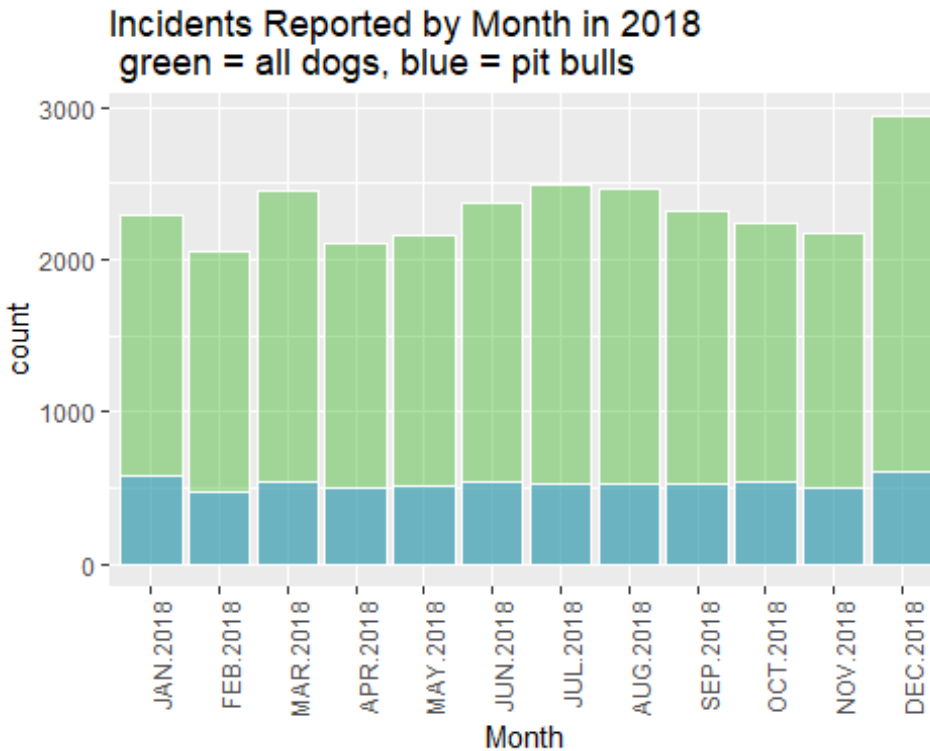
ANSWER: From the data, we can see that the shelter reported more dog-related incidents in December than any other month. Aside from that, it is evident that the number of dog-related incidents rises slightly during summer months. Comparing this to Pit Bulls which is highlighted in blue, the quantity of incidents per month is fairly even throughout the year, so there is no indication that Pit Bulls are more likely to be reported during any given month.

## Q6: Are the majority of the animals taken in healthy or sick?

```
# Collapse the data file's factors into 'Healthy' and 'Sick'
levels(dat_org$Intake.Condition) <- list("HEALTHY" = c("HEALTHY", "NORMAL",
""), "SICK" = c("TREATABLE MANAGEABLE CONTAGIOUS", "TREATABLE MANAGEABLE NON-
CONTAGIOUS", "TREATABLE REHABILITABLE CONTAGIOUS", "TREATABLE REHABILITABLE
NON-CONTAGIOUS", "UNHEALTHY UNTREATABLE CONTAGIOUS
", "UNHEALTHY UNTREATABLE NON-CONTAGIOUS"))

# Collapse factors for the subsetted data file for Pit Bull only into
'Healthy' and 'Sick'
levels(dat_org__DOG_PITBULL$Intake.Condition) <- list("HEALTHY" =
c("HEALTHY", "NORMAL", ""), "SICK" = c("TREATABLE MANAGEABLE CONTAGIOUS",
"TREATABLE MANAGEABLE NON-CONTAGIOUS", "TREATABLE REHABILITABLE CONTAGIOUS",
"TREATABLE REHABILITABLE NON-CONTAGIOUS", "UNHEALTHY UNTREATABLE CONTAGIOUS
", "UNHEALTHY UNTREATABLE NON-CONTAGIOUS"))

# Pie chart of animal intake condition. Use column labeled 'Intake.Condition'
```

```
pie(table(dat_org$Intake.Condition), main = "INTAKE CONDITION of DOGS in
2018", col = rainbow(6), labels = paste(c("HEALTHY","SICK"), " ",
round(prop.table(table(dat_org$Intake.Condition))*100,1), "%", sep = "", "
(", table(dat_org$Intake.Condition),  ") "), cex = 0.8)
```


INTAKE CONDITION of DOGS in 2018
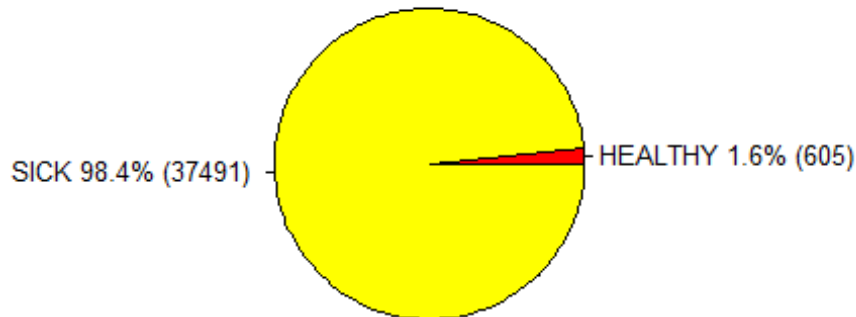
```
# Pie chart of Pit Bull intake condition. Use column labeled
'Intake.Condition'
pie(table(dat_org__DOG_PITBULL$Intake.Condition), main = "INTAKE CONDITION of
PIT BULLS in 2018", col = rainbow(6), labels = paste(c("HEALTHY","SICK"), "
", round(prop.table(table(dat_org__DOG_PITBULL$Intake.Condition))*100,1),
"%", sep = "", " (", table(dat_org__DOG_PITBULL$Intake.Condition),  ") "),
cex = 0.8)
```

INTAKE CONDITION of PIT BULLS in 2018

SICK 99.1% (6257) — | — HEALTHY 0.9% (58)
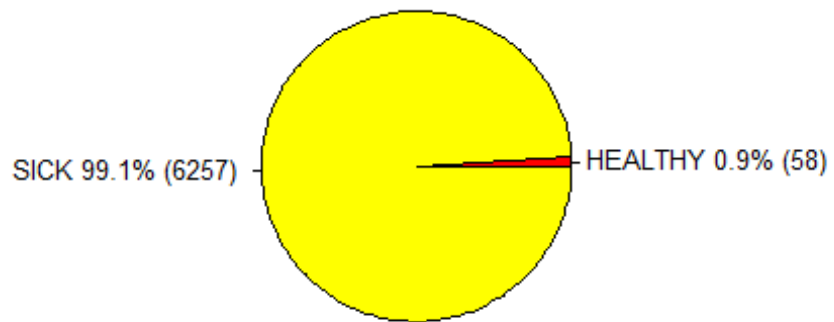
ANSWER: These pie charts reveal that nearly every dog-related incident reported concerns a dog that is sick. The same is true for Pit Bulls which, for all of 2018, over 99% were sick.