Group_16_Project

Group_16

```
library(tidyverse)
library(moderndive)
library(gapminder)
library(sjPlot)
library(stats)
library(jtools)
library(dplyr)
library(knitr)
library(magrittr)
```

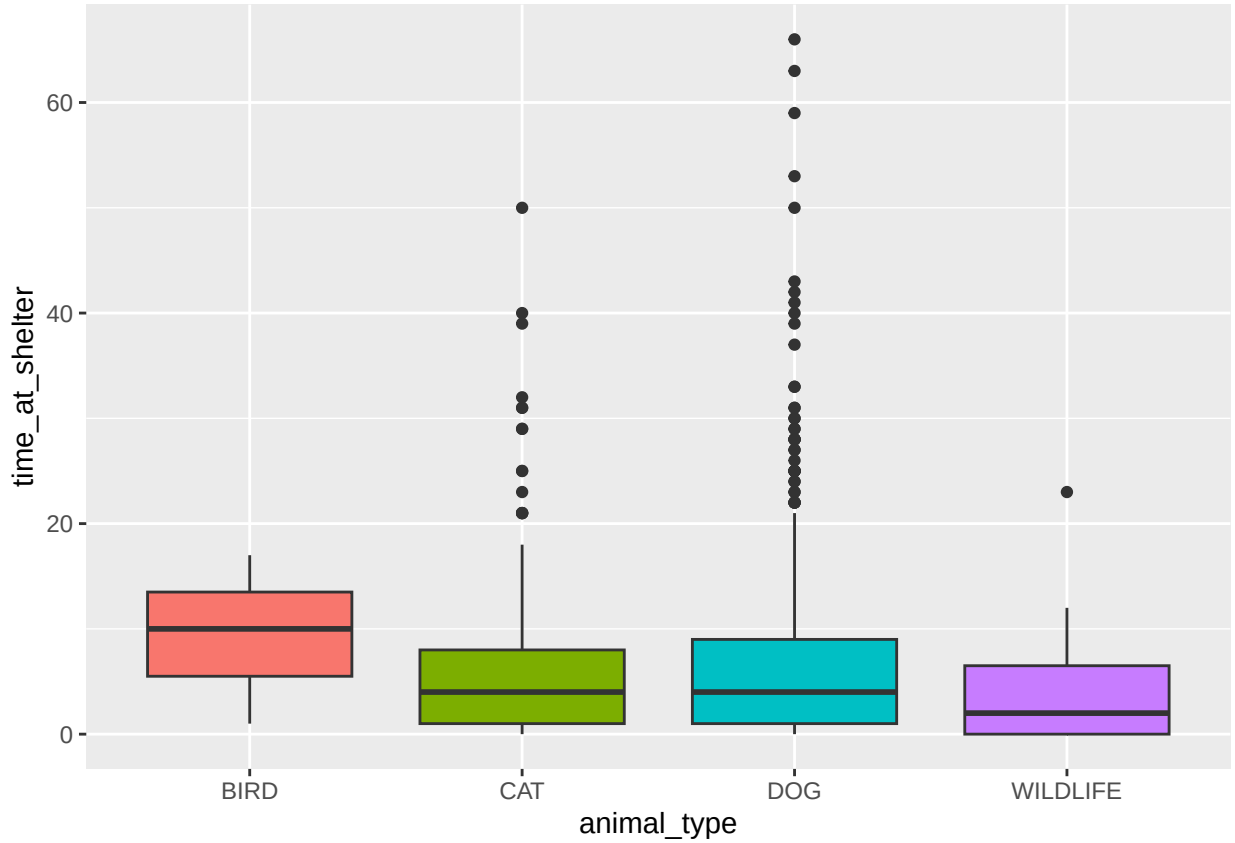
Introduction

Exploratory Data Analysis

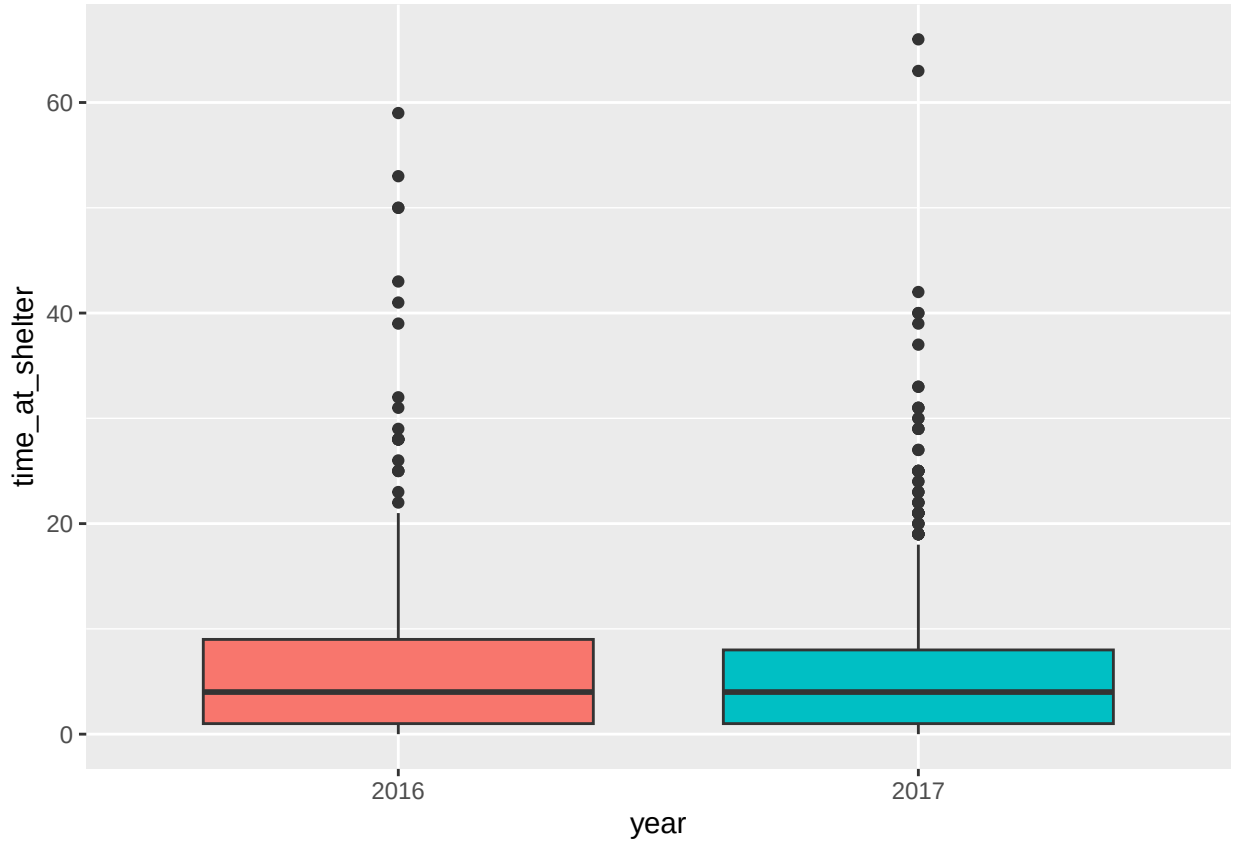
```
# create a training data set and transfer the character
# explanatory variable and integer variable into
# factor variable
df = animal%>%
  mutate_if(is.character, as.factor)
df$year=as.factor(df$year)
df$month=as.factor(df$month)
```

Plot to get an initial impression of the data

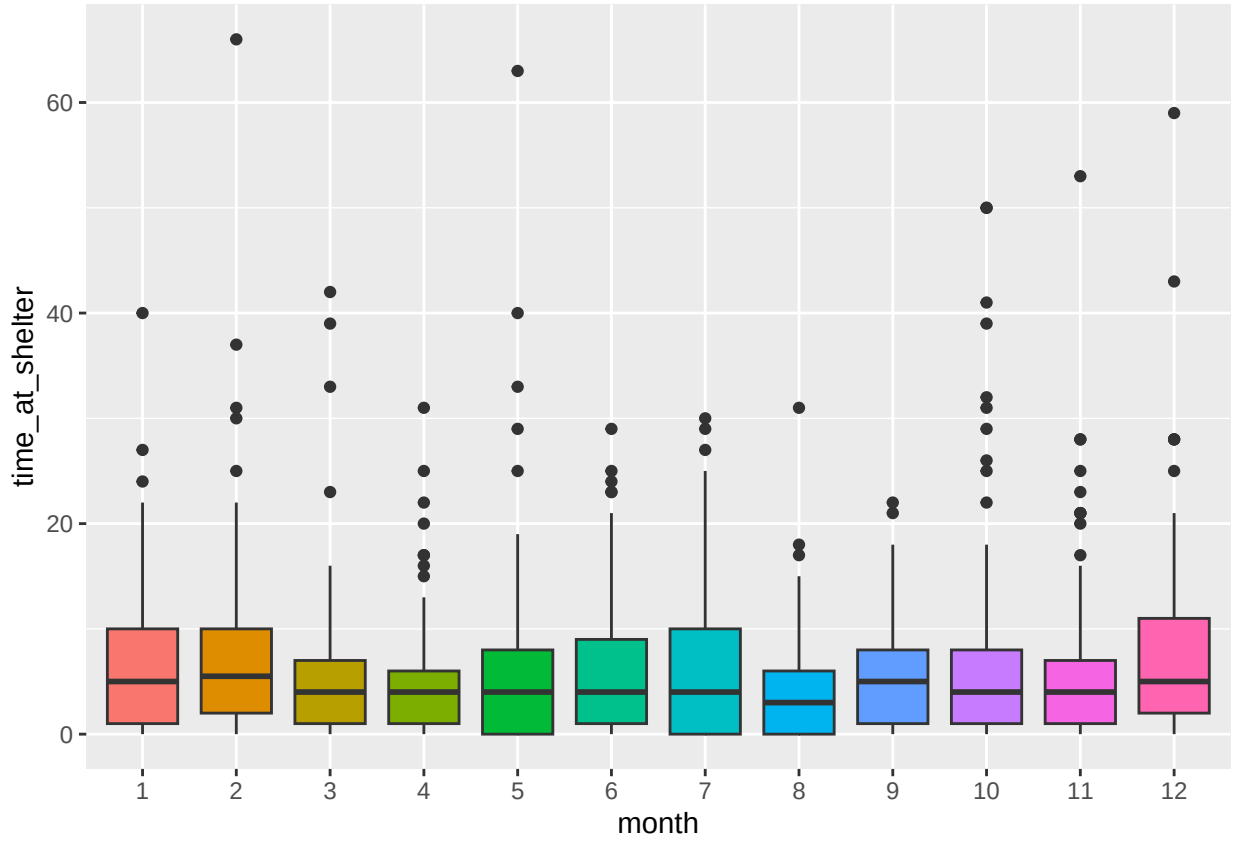
```
#Plot of animal_type against outcome_type
ggplot(data = df, aes(x = animal_type, y =time_at_shelter , fill = animal_type)) +
geom_boxplot() +
labs(x = "animal_type", y = "time_at_shelter") +
theme(legend.position = "none")
```



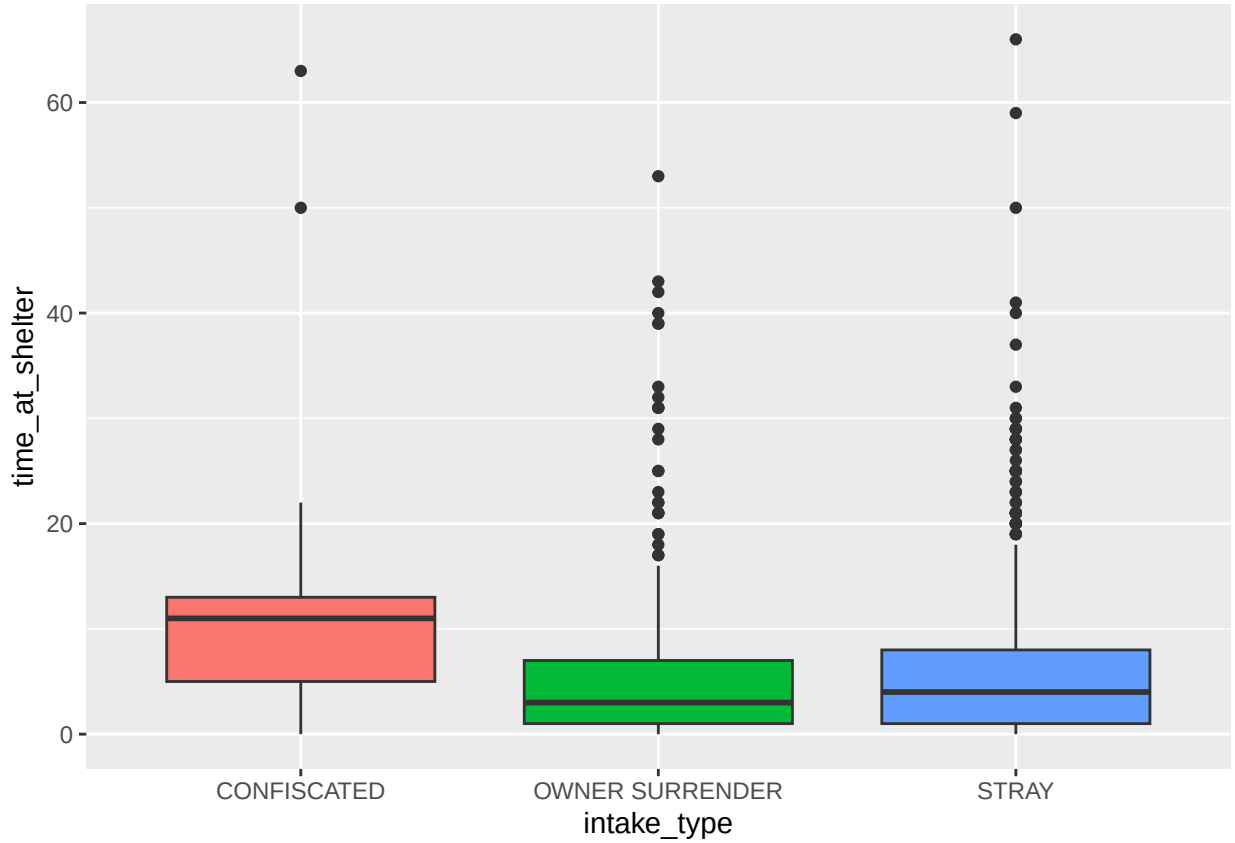

```
#Plot of year against outcome_type
ggplot(data = df, aes(x = year, y = time_at_shelter, fill = year)) +
  geom_boxplot() +
  labs(x = "year", y = "time_at_shelter") +
  theme(legend.position = "none")
```



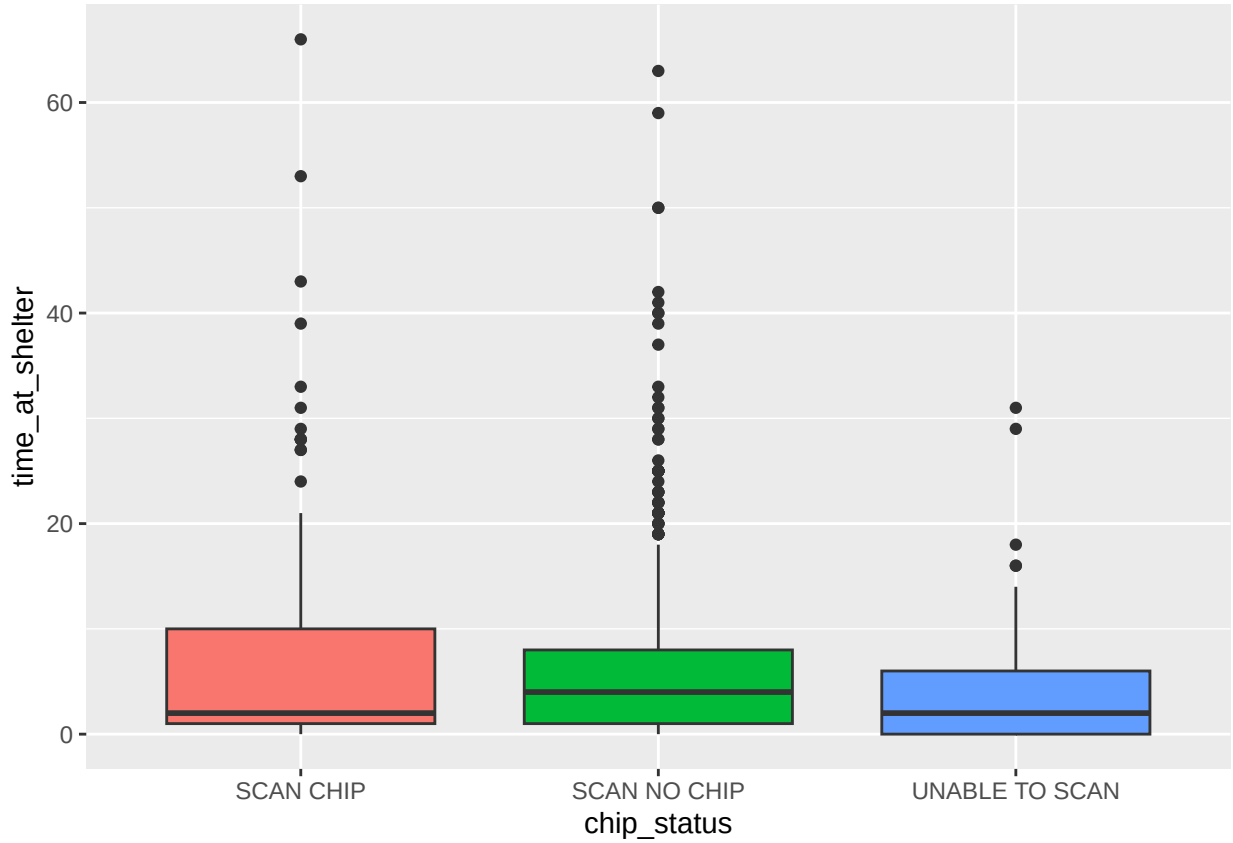

```
#Plot of month against time_at_shelter
ggplot(data = df, aes(x = month, y = time_at_shelter, fill = month)) +
  geom_boxplot() +
  labs(x = "month", y = "time_at_shelter") +
  theme(legend.position = "none")
```




```
#Plot of intake_type against time_at_shelter
ggplot(data = df, aes(x = intake_type, y = time_at_shelter, fill = intake_type)) +
  geom_boxplot() +
  labs(x = "intake_type", y = "time_at_shelter") +
  theme(legend.position = "none")
```




```
#Plot of chip_status against time_at_shelter
ggplot(data = df, aes(x = chip_status, y = time_at_shelter, fill = chip_status)) +
  geom_boxplot() +
  labs(x = "chip_status", y = "time_at_shelter") +
  theme(legend.position = "none")
```




```
#Modelling
model = glm(time_at_shelter ~ animal_type + month + year + intake_type + chip_status, family="poisson",
#Optimize the model
#Summarize the first model
model %>%
  summary()
```

```
Call:
glm(formula = time_at_shelter ~ animal_type + month + year +
    intake_type + chip_status, family = "poisson", data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9857  -2.6038  -0.7551   0.8060  12.7653

Coefficients: (1 not defined because of singularities)
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              2.75306    0.19690  13.982  < 2e-16 ***
animal_typeCAT          -0.18315    0.19396  -0.944 0.345017
animal_typeDOG          -0.23305    0.19251  -1.211 0.226055
animal_typeWILDLIFE     -0.38615    0.22984  -1.680 0.092945 .
month2                   0.14546    0.05533   2.629 0.008562 **
month3                  -0.25477    0.05692  -4.476 7.61e-06 ***
month4                  -0.28620    0.05657  -5.059 4.21e-07 ***
month5                  -0.11780    0.05181  -2.273 0.022996 *
```

```
month6                      -0.08372    0.04991  -1.677 0.093482 .
month7                      -0.14828    0.05051  -2.936 0.003327 **
month8                      -0.50955    0.05871  -8.679  < 2e-16 ***
month9                      -0.20499    0.05596  -3.663 0.000249 ***
month10                      0.05627    0.05167   1.089 0.276127
month11                     -0.05734    0.05436  -1.055 0.291483
month12                      0.16794    0.05157   3.257 0.001127 **
year2017                         NA         NA      NA       NA
intake_typeOWNER SURRENDER -0.77494    0.04076 -19.012  < 2e-16 ***
intake_typeSTRAY           -0.58951    0.03765 -15.658  < 2e-16 ***
chip_statusSCAN NO CHIP    -0.01469    0.02805  -0.524 0.600525
chip_statusUNABLE TO SCAN  -0.27589    0.06781  -4.068 4.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 10551.2  on 1449  degrees of freedom
Residual deviance:  9957.4  on 1431  degrees of freedom
AIC: 14017

Number of Fisher Scoring iterations: 6
```

```
#Raise the second model by the summary table
#Drop the year factor since it is only none value
model2 = glm(time_at_shelter ~ animal_type + month + intake_type + chip_status, family="poisson", data=
#Summarize the second model
model2 %>%
  summary()
```

```
Call:
glm(formula = time_at_shelter ~ animal_type + month + intake_type +
    chip_status, family = "poisson", data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9857  -2.6038  -0.7551   0.8060  12.7653

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                2.75306    0.19690  13.982  < 2e-16 ***
animal_typeCAT            -0.18315    0.19396  -0.944 0.345017
animal_typeDOG            -0.23305    0.19251  -1.211 0.226055
animal_typeWILDLIFE       -0.38615    0.22984  -1.680 0.092945 .
month2                     0.14546    0.05533   2.629 0.008562 **
month3                    -0.25477    0.05692  -4.476 7.61e-06 ***
month4                    -0.28620    0.05657  -5.059 4.21e-07 ***
month5                    -0.11780    0.05181  -2.273 0.022996 *
month6                    -0.08372    0.04991  -1.677 0.093482 .
month7                    -0.14828    0.05051  -2.936 0.003327 **
month8                    -0.50955    0.05871  -8.679  < 2e-16 ***
month9                    -0.20499    0.05596  -3.663 0.000249 ***
month10                    0.05627    0.05167   1.089 0.276127
```

```
month11                     -0.05734    0.05436  -1.055 0.291483
month12                      0.16794    0.05157   3.257 0.001127 **
intake_typeOWNER SURRENDER -0.77494    0.04076 -19.012  < 2e-16 ***
intake_typeSTRAY           -0.58951    0.03765 -15.658  < 2e-16 ***
chip_statusSCAN NO CHIP    -0.01469    0.02805  -0.524 0.600525
chip_statusUNABLE TO SCAN  -0.27589    0.06781  -4.068 4.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 10551.2  on 1449  degrees of freedom
Residual deviance:  9957.4  on 1431  degrees of freedom
AIC: 14017

Number of Fisher Scoring iterations: 6
```