Squirrels in NY: Project Proposal

Team BCHZ - Nick Bayer, Richard Cui, Laura Han, Anna Zhang

```
library(tidyverse)
library(tidymodels)
library(httr)
```

```
squirrels <- read_csv("data/squirrel_data.csv")
head(squirrels)
```

```
# A tibble: 6 x 31
      X     Y `Unique Squirrel ID` Hectare Shift     Date Hectare Squirrel Num~1
  <dbl> <dbl> <chr>                <chr>   <chr>    <dbl>                  <dbl>
1 -74.0  40.8 37F-PM-1014-03       37F     PM    10142018                      3
2 -74.0  40.8 21B-AM-1019-04       21B     AM    10192018                      4
3 -74.0  40.8 11B-PM-1014-08       11B     PM    10142018                      8
4 -74.0  40.8 32E-PM-1017-14       32E     PM    10172018                     14
5 -74.0  40.8 13E-AM-1017-05       13E     AM    10172018                      5
6 -74.0  40.8 11H-AM-1010-03       11H     AM    10102018                      3
# i abbreviated name: 1: `Hectare Squirrel Number`
# i 24 more variables: Age <chr>, `Primary Fur Color` <chr>,
#   `Highlight Fur Color` <chr>,
#   `Combination of Primary and Highlight Color` <chr>, `Color notes` <chr>,
#   Location <chr>, `Above Ground Sighter Measurement` <chr>,
#   `Specific Location` <chr>, Running <lgl>, Chasing <lgl>, Climbing <lgl>,
#   Eating <lgl>, Foraging <lgl>, `Other Activities` <chr>, Kuks <lgl>, ...
```

```
library(tidyverse)
squirrels |> group_by(`Runs from`) |> summarize(n())
```

```
# A tibble: 2 x 2
  `Runs from` `n()`
```

```
  <lgl>       <int>
1 FALSE        2345
2 TRUE          678
```

Introduction

As a result of the continuous human development, animals are inevitably interacting with humans more often. However, this form of interaction has mostly shown to be a disturbance to animals [1]. Animals see humans as a threat, so it is no surprise that they would treat the presence of humans the same way they would when they face other predators. Nevertheless, recent studies show that the squirrels actually act differently, as characterized by a phenomenon called synurbization, or the process of becoming urbanized [2].

In an effort to investigate these two competing theories, and to better understand the dynamic between squirrels and humans, we carry out this research project to explore what factors affect whether a squirrel is indifferent to human presence. From there, we would like to deduce whether the squirrels' attitude to humans are caused by human presence or other factors such as their species.

We hypothesize that the age category, location, distance above ground when spotted, number of activities that the squirrel was doing, sound that the squirrel makes, and whether squirrel is disturbed by human activities (as measured by features like approaching or running away from humans and tail signs) could have a relationship with the attitude of the squirrel (whether indifferent or not).

Data description

We are sourcing our data set from the TidyTuesday project on GitHub. Their data originally came from The 2018 Squirrel Census, a project based on the sightings of squirrels in Central Park, New York City.

In October of 2018, the Squirrel Census Team and a group of over 300 volunteers collected the data based on squirrel sightings around Central Park.

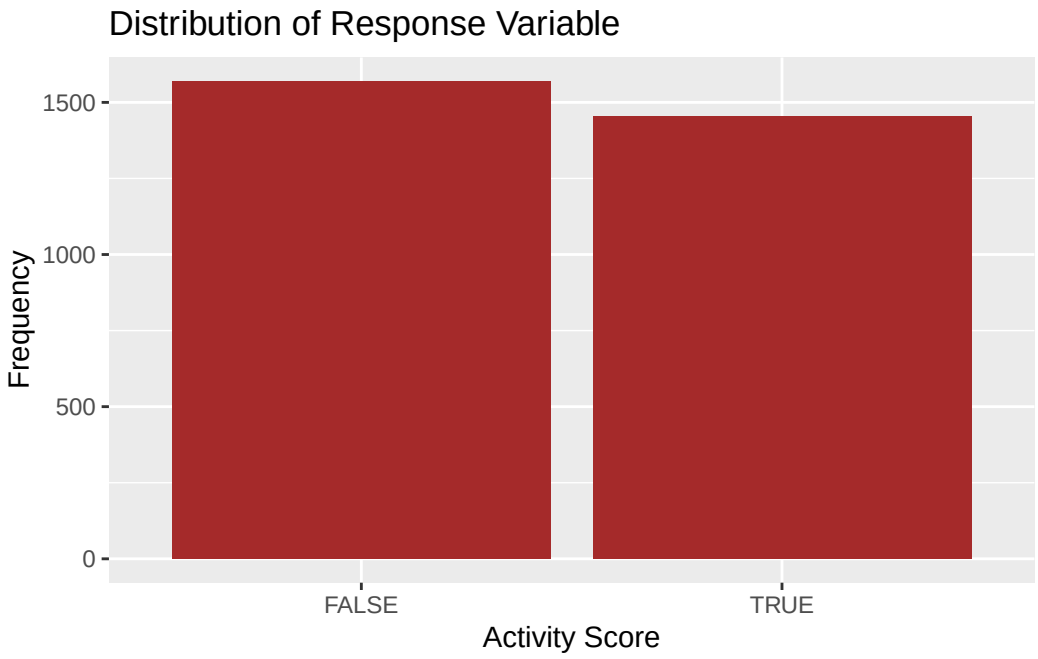
The Data gives a wide range of observations and characteristics. It first gives us the location, in both longitude and latitude, the hectare of the park the squirrel was located in, the date, and whether it was found in the AM or PM. It also assigns each squirrel a unique ID. It has information on whether the squirrel is an adult or a juvenile, its primary and highlight fur colors, and has a number for the sequence of sightings in one session. It also contains data on their exact location, how far they were from the ground, and the objects they were found on. There is data for the activities the squirrel was found doing, ranging from running to foraging, with a separate column for any activity that was not chosen to be a column. It gives data on the sounds the squirrel made and tail movements, if any. Finally, it has 4 columns for the

response, being either that the squirrel approached, was indifferent, ran away, or any other action.

Initial exploratory data analysis

There seems to be roughly equal number of squirrels that were and weren't indifferent to the researchers.

```
squirrels |>
  ggplot(aes(x = Indifferent)) +
  geom_bar(fill = "brown") +
  labs(title = "Distribution of Response Variable",
         x = "Activity Score",
         y = "Frequency")
```

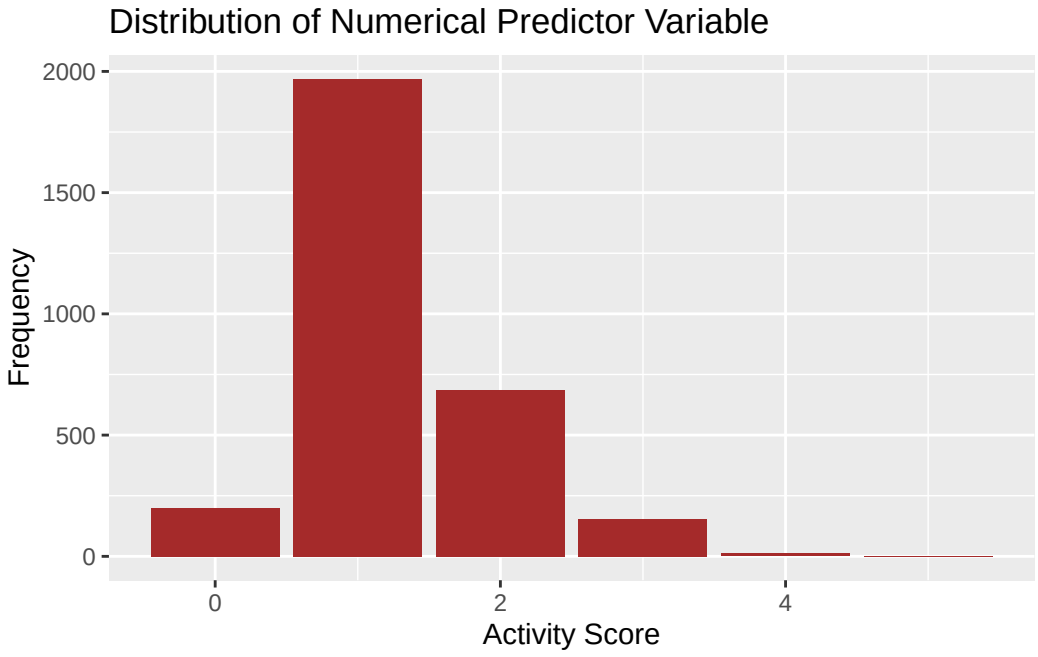


We create a numeric variable named Activity_Score that encapsulates the number of activities a squirrel is engaged in during the span of observation. The distribution of the Activity_Score variable is unimodal, slightly right skewed, and has a median of 1. This means that most squirrels in the data set were only engaged in 1 activity during the span of observation.

```
activities <- squirrels |>
  select(c(Running, Chasing, Climbing, Eating, Foraging)) |>
  mutate(Running = ifelse(Running == TRUE, 1, 0),
         Chasing = ifelse(Chasing == TRUE, 1, 0),
         Climbing = ifelse(Climbing == TRUE, 1, 0),
         Eating = ifelse(Eating == TRUE, 1, 0),
         Foraging = ifelse(Foraging == TRUE, 1, 0))

squirrels$Activity_Score <-
  rowSums(activities)
```

```
squirrels |>
  ggplot(aes(x = Activity_Score)) +
  geom_bar(fill = "brown") +
  labs(title = "Distribution of Numerical Predictor Variable",
         x = "Activity Score",
         y = "Frequency")
```



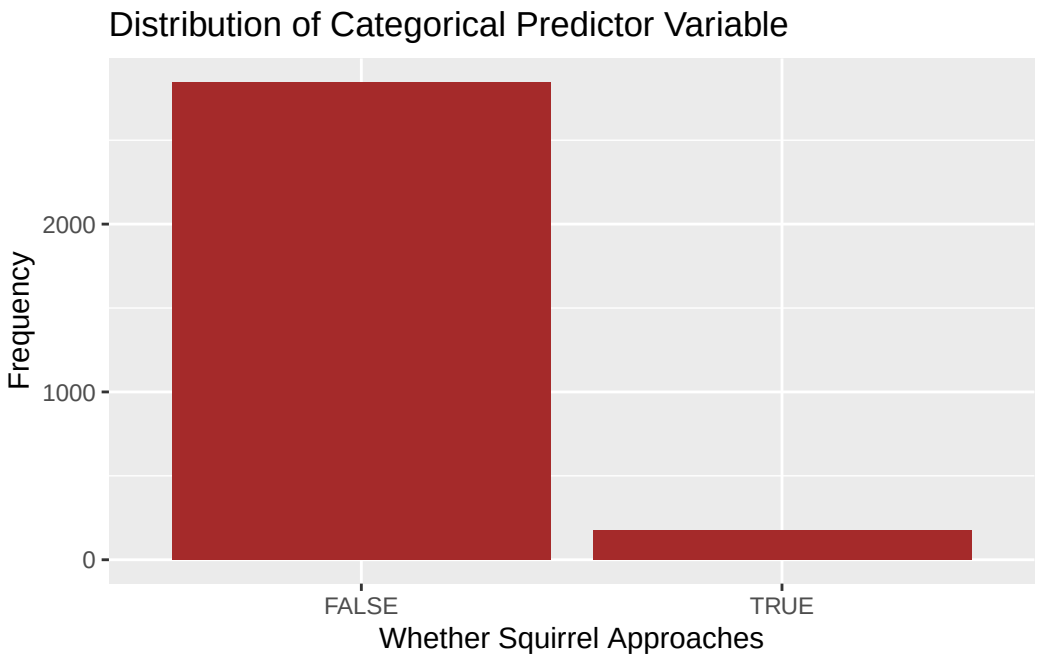

```
summary(squirrels$Activity_Score)
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
  0.000   1.000   1.000   1.278   2.000   5.000
```

There were far less squirrels that approached the researchers than squirrels that didn't. This could show that they are not as synurbanized as we thought.

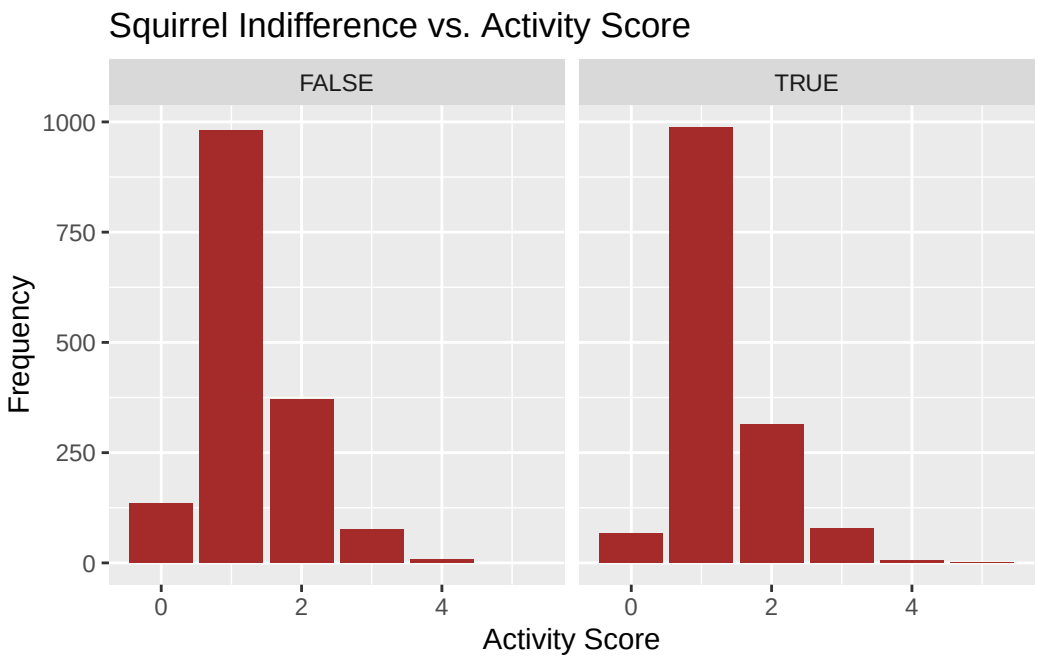
```
squirrels |>
  ggplot(aes(x = Approaches)) +
  geom_bar(fill = "brown") +
  labs(title = "Distribution of Categorical Predictor Variable",
         x = "Whether Squirrel Approaches",
         y = "Frequency")
```



The distribution of Activity_Score for squirrels that were indifferent vs. not indifferent is roughly the same shape. Therefore, activity score and indifference do not seem to be correlated.

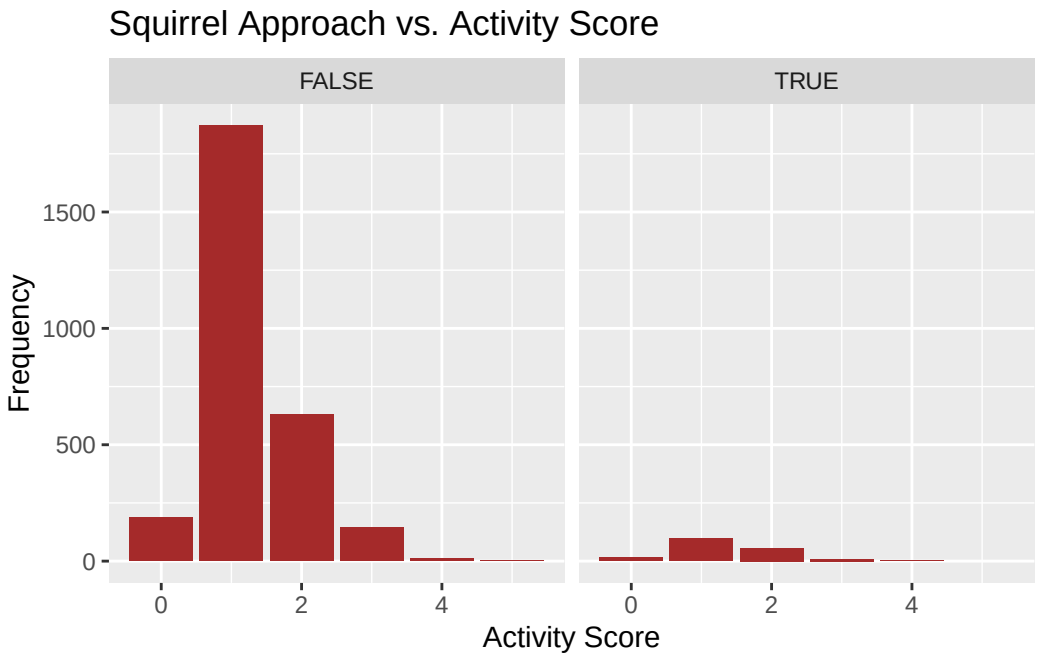
```
squirrels |>
  ggplot(aes(x = Activity_Score)) +
  geom_bar(fill = "brown") +
  facet_wrap(~ Indifferent) +
  labs(title = "Squirrel Indifference vs. Activity Score",
```

```
       x = "Activity Score",
       y = "Frequency")
```



Since there were so few squirrels that approached the researchers, it is hard to tell whether the graphs are very different in shape. The difference may be something that we explore further.

```
squirrels |>
  ggplot(aes(x = Activity_Score)) +
  geom_bar(fill = "brown") +
  facet_wrap(~ Approaches) +
  labs(title = "Squirrel Approach vs. Activity Score",
         x = "Activity Score",
         y = "Frequency")
```



An interaction that we would like to examine further is whether the presence of tail flags and/or twitches influences the likelihood that a squirrel is indifferent to the researchers.

Analysis approach

...

Data dictionary

The data dictionary can be found here [Update the link and remove this note!]