Team members: Maya Wheeler, Mikayla Comptom, Ashley Pritchard
Project Name: Squirrels In My Park (SIMP)
Repository Link: https://github.com/pritchardashley/Data-Science-SIMP

Group Information
- Mikayla Compton, mikayla.compton@utah.edu, u1538218
- Ashley Pritchard u1256025@utah.edu, U1256025
- Maya Wheeler, maya.wheeler@hsc.utah.edu, u0999875

Topic: Understanding the correlation between physical squirrel attributes and location on behavior

Background and Motivation

Almost everywhere we go, campus, a park, or even a big city, there will be squirrels. There are over 280 species of squirrels around the world (1) and come in a variety of colors: Black, brown, grey, red, and even containing purple, orange, and maroon (2) Research has found that squirrels have personalities and similar behaviors to humans (3). There has also been some research into the patterns of dominance in squirrels with older squirrels being more aggressive than younger squirrels and similar patterns for larger and smaller squirrels (4). There is a lot more to learn about squirrels and how and why they behave the ways that they do.

A lot of our interest on this topic stems from our personal observations, specifically with the squirrels on the University of Utah campus. We as a group concluded that the level of human interactions, and food availability may have impacted their behaviors in certain situations. The squirrels on the University of Utah campus even have their own portion of the University of Utah website dedicated to them (University of Utah Website) as well as an annual "Squirrel Fest" held by the Natural History Museum (Squirrel Fest). Although the data set of squirrels that we will be using is not based in Utah, we feel as though we can learn a lot of information, through the application of skills learned in class, to gain a better understanding of the squirrels in Utah and around the world.

Project Objectives

The primary question that we want to try to answer within our project is if there is a strong correlation between the physical characteristics and location of the squirrels in central park and their observed behaviors. The specific characteristics that we are focusing on are fur color, highlighted fur color, age, and location of the squirrels. In a part of this we want to see if the squirrels that show fear towards interactions with humans have different characteristics than those squirrels that do not. Our null hypothesis is that the physical characteristics have no effect on behavioral observations

and our alternative hypothesis is that the physical characteristics have an effect on the observed behaviors.

Our learning objective is to utilize what we have learned in class to perform an accurate analysis of our data. Potential skills we hope to apply and understand more deeply include but are not limited to: R-squared, P-value, OLS Regression, descriptive statistics, categorical data, dummy values, data clean, and data visualization. Another thing we want to look into is creating a map of the squirrels in Central Park that shows our results for better visualization. Depending on time we may explore KMeans clustering for fur color and location, as well as using machine learning to predict squirrel behaviors.

Data Description and Acquisition

Our data comes from the 2018 NYC Squirrel Census. This was a citizen science project that had volunteers observing squirrels in Central Park in New York City. The data is in two CSV files, as well as a word document with data definitions. The first file has our primary data, which includes 3022 observations. Each observation has 7 columns identifying the location, time, date, etc for the observation. There are 23 different columns with boolean or nominal values for different attributes of the observed squirrel, such as color (grey, cinnamon or black), what the squirrel was doing(eating, foraging, etc.), and behaviors it displayed, including vocalizations or interactions with humans. The second file is a CSV containing the latitude and longitude of the hectares that the park was broken into for the observations. Both of these datasets are very complete- with the majority of the missing data being under "other" qualitative observations.

- Squirrel census data:
    - https://data.cityofnewyork.us/Environment/2018-Central-Park-Squirrel-Census-Squirrel-Data/vfnx-vebw/about_data
- Hectare data:
    - https://data.cityofnewyork.us/dataset/2018-Central-Park-Squirrel-Census-Hectare-Grid/qad5-y26n/about_data

Ethical Considerations

The group that would be the most affected by our analysis would be the squirrel community within central park. Depending on what we find in our analysis it could encourage people to interact more with squirrels. This, in turn, could lead to the endangerment of the squirrels or an increase of stress onto them. If people are influenced to interact with the squirrels it could also lead to the transmission of disease between the squirrels and humans. With a possible influx of people interacting with squirrels this would also mean a direct increase of people to central park. This could

affect other aspects of the park outside of the squirrel population like people straying off paths or stepping on plants.

Data Cleaning and Processing.

The census data is fairly clean, so we are not expecting substantial clean up there. The biggest piece of processing that we are expecting will be joining the census data to the hectare data and making sure the location data in the census table aligns with the data in the hectare table.

Exploratory Analysis.

With the data being nominal we plan to use bar charts to get a sense of the general distribution of our data. Working with geographical data is new for us, so we are open to different methods, but based on our initial research we are going to start with the folium library in Python to map up the distribution of observations

Analysis Methodology.

We are planning to test for the correlation (and strength of that correlation) between physical traits and behaviors, using the chi squared test of independence and Carmer's V test. We will also test the correlation between the location in the park, and squirrel behavior, which we will analyze with an ANOVA test.

Project Schedule.
Important Project Dates:
~~Feb 11 – Meet with group~~
- Decide on project topic
~~Feb 21 – Announce Team/Title~~
~~Feb 25 – Meet with group~~
- split up proposal task
    - Ashley = Intro/background and schedule
    - Maya = Data cleaning and processing and data description?
    - Mikayla = Project objectives and Ethical considerations
    - Group = anything else that I don't remember
~~March 4 – Meet with group~~
- review proposal
- Complete and revisions decided upon
- Figure out how we will all work with the data (Github repo? Emailing files?)
- How often do we want to meet? Days/time that work best?
March 7 – Project Proposal
- Ideally submit early
March 11 – Meet with group

- Clean data
- https://utah.zoom.us/j/96194506954 Passcode: simp

March 18 – Meet with group
- Initial analysis
- https://utah.zoom.us/j/96194506954 Passcode: simp

March 20 – Peer Feedback (in-class)
- zoom or facetime Mikayla, if not make sure that detailed notes are taken!
- Schedule a group meeting to go over peer feedback as a team

March 21 – Project Milestone Report
- Assignment on canvas, seems similar to proposal but with additional updates
- Group Feedback Form - All submit

March 25 – Meet with group
- Final analysis
- https://utah.zoom.us/j/96194506954 Passcode: simp

April 1 – Meet with group
- Visualizations
- https://utah.zoom.us/j/96194506954 Passcode: simp

March 31- April 4 – Professor/TA feedback (by appointment)
- Email about a week in advance to schedule an appointment if there is not a sign up sheet provided
- Meet with group afterwards to discuss feedback

April 8 – Group meeting
- Film presentation
- https://utah.zoom.us/j/96194506954 Passcode: simp

April 18 – Final Project

April 22 – Project Awards

Sources

1. https://www.iaismuseum.org/wp-content/uploads/2020/05/Lets-Learn-About-Squirrels.pdf
2. https://www.wildlifeonline.me.uk/animals/article/squirrel-fur-colour-characteristics#:~:text=Among%20the%20Greys%2C%20silvery%20(frosted,grey%20but%20the%20feet%20orange.
3. https://www.ucdavis.edu/curiosity/news/personality-matters-even-squirrels-0
4. https://www.wildlifeonline.me.uk/animals/article/squirrel-behaviour-dominance-aggression-territoriality
5. https://attheu.utah.edu/students/animals-of-the-u-fox-a-squirrel/
6. https://nhmu.utah.edu/citizen-science/utah-fox-squirrels/survey