
Analysis and Visualization of 2018 Central Park Squirrel Census - Squirrel Data

Guray Ozgur

Matrikelnummer 6041382

guray.ozgur@student.uni-tuebingen.de

Moritz Vannahme

Matrikelnummer 5968774

moritz.vannahme@student.uni-tuebingen.de

Abstract

In 2018 there has been a census of the population of squirrels in the New York Central Park. In this paper, we analyze and visualize this publicly available data to investigate patterns in squirrel color and behaviour.

For multiple years now, a census has been made of the New York Central Park squirrel population. The data is collected with the help of volunteers, who name themselves "[Squirrel Sighters](#)" and made available to the public on their [webpage](#). In this paper, we will work with the data from the 2018 Central Park Squirrel Census. It was collected by 323 volunteers and comprises a total of 3023 squirrel sightings. These do not necessarily represent unique squirrels and the official estimate is instead 2373, for a more precise estimate however, tagging would likely be required. For our project, we have chosen to look into a variety of distributions potentially present in the data and visualized a selection of them.

1 Squirrel Distributions

The first approach we chose was to visualize the squirrels on a map of Central Park. Our initial step was to simply add a dot for each squirrel at its latitude and longitude coordinates. A later stage with dots coloured according to squirrel color can be seen in Fig. 1. It makes obvious the prevalence of the Eastern grey squirrel (*Sciurus Carolinensis*). Other maps showed whether the squirrel was sighted in the morning or the afternoon, the date of the sighting or whether it was a juvenile or an adult. Notable information that can be gleaned from this is that squirrels particular seem to favour the area south of Vista Rock. It is one of the two forested areas of the park but unlike the northern areal intercut with many paths. This may make it more attractive to the squirrels either due to left-over food or may make it easier for squirrel sighters to reach and see squirrels. Another important observation can be that even though the most populated area is the forest near Vista Rock, it is not as populated by black squirrels as it is by gray and cinnamon ones.

We also created a heat-map from the same data, by not using the coordinate information but hectare information, which can be seen in Fig. 2. Despite being stripped of the identifying information of the geography of Central Park, it lets us guess at not only the shape of the park but also the altitude even though the altitude information is not given at all in the data. The presence of the squirrels forms a kind of topographical map of the park.



Figure 1: Squirrel-sightings in Central Park - Colored in Squirrel Color

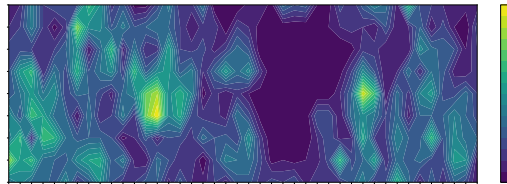


Figure 2: Heat-map of squirrel sightings in Central Park

2 Fur Color

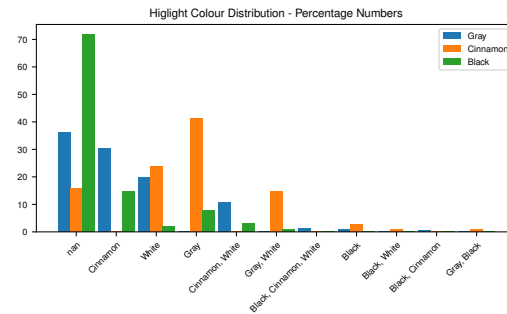


Figure 3: Distribution of Highlight Colors over Primary Colors in Percent

Next, we took a closer look at the different colors of the squirrels. Each squirrel has one of three primary colors (Gray, Cinnamon, Black) or None if the recorder failed to identify the color. Each squirrel can then have any combination of those 3 colors or White as a highlight color (including having no highlight color at all). The split of highlight colors over the three primary colors is shown in Fig 3. None was omitted as the only highlight color recorded was None. Even after taking into consideration that primary and highlight color can not match, it is evident that the highlight colors follow different distribution depending on the primary color. Cinnamon squirrels show the highest variety in highlight color, with white, gray and cinnamon all being prevalent. Cinnamon is especially the only primary color with a significant level of tricolored squirrels. Black squirrels are primarily mono-colored with some few individuals having cinnamon or gray highlights.

3 Squirrel Behaviour

The Central Park squirrels spend the majority of their visible time occupied with foraging, which makes up nearly 50% of all sightings. They can also frequently be seen eating. Nearly 50% are also indifferent to humans. In 20% of cases however, squirrels instead ran away. Only about 5% of sightings involved squirrels approaching humans, suggesting a "live and let live" attitude but nonetheless a baseline of wariness. Vocal signalling was only present in about 5% of cases, but communication with the tail to make the squirrel seem more imposing was present in ~16% of sightings. This may however be recorders' bias as positively identifying tail motion is easier than identifying squirrel noises.

4 Behavioural Correlation - Internal

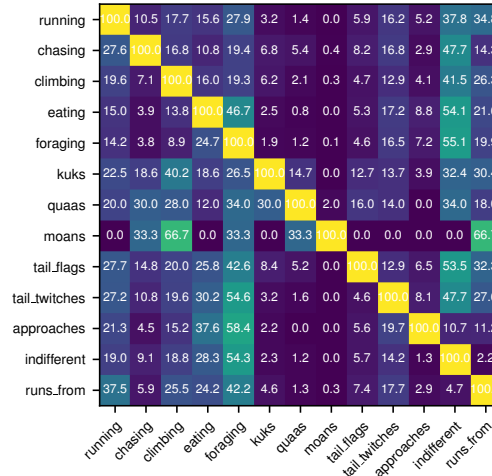


Figure 4: The conditional probabilities of behaviours given as $P(\text{Column}|\text{Row})$

We investigated the correlation between different behaviours of squirrels. The results can be seen in Fig. 4. It should be noted that due to being nominal data, the information given is in the form of conditional probability, not covariance. While we can see what appears like strong relations especially in the case of "foraging" and "indifferent", these should be treated with caution as they each make up about 50% of all recorded squirrels already. Consequently the strong relation can actually be seen in those examples where the odds are disproportionately low. This is the case for intuitively mutually exclusive behaviours like indifferent-approaches-runs_from or foraging-climbing-chasing. Kukking and quaaing also predict against indifferent. As these are both danger signalling noises this is consistent with expectations. Moaning should be disregarded due to only showing up 3 times across all sightings.

5 Behavioural Correlation - Other Factors

We also checked the conditional probabilities of behaviours depending on the color, the time of day and the height above ground. Neither color nor time of day had any strong effect on squirrel behaviour. Height above ground unsurprisingly indicated for climbing and against foraging, running or eating.

6 Logistic Regression

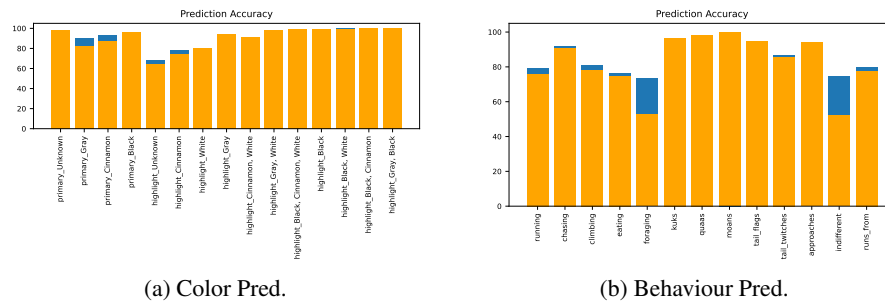


Figure 5: Accuracy of our logistic regression in blue. Yellow is the accuracy achieved by choosing the most common answer (True/False) for each category

In addition to the above analysis, we created a logistic regression model of different squirrel qualities. The accuracies are shown in Fig. 5. We used a Training/Test split of 75/25. We see that a logistic regression is able to achieve some success in predicting the primary color from the highlight colors and significant success in determining whether a squirrel was eating and whether it was indifferent to humans from its other behaviours. We also used the Themis-ML, developed by Bantilan [2017] to calculate mean difference scores for color and age with regard to foraging. We found that black squirrels were more likely to be foraging while juveniles were less likely. Perhaps black squirrels are more curious and less wary towards humans and consequently approach them, while juvenile squirrels are either cared for by their parents or simply more wary of the unknown.

7 Summary

On the whole, while it is an amusing and fascinating piece of data, the primary insight this census offers into the behaviour of squirrels lies in the initial breakdown of behaviours rather than in deeper patterns within the data. All our data and processing code can be found at:

https://github.com/hutcun/2022-02-NYC_Squirrels.git

References

Niels Bantilan. Themis-ml: A fairness-aware machine learning interface for end-to-end discrimination discovery and mitigation. *CoRR*, abs/1710.06921, 2017. URL <http://arxiv.org/abs/1710.06921>.