Free Bird (Data): Using community sourced Bird-Cam data to document bird feeding patterns

redacted

10/26/2020

Executive Summary

In this study, we define a processing method to analyze data derived from live-streamed "Bird Cams" in Panama. Data is reported in 1 minute intervals, and user observation disagreements are reconciled using weighted averages based on viewer activity. We present detailed graphical displays of bird feeding behavior over time, and formalize a behavior model for Clay-colored Thrush birds using generalized linear regression models with a Poisson error distribution.

Introduction

The natural habitat and behavioral patterns of birds in the wild can be difficult to observe. Recently, researchers have installed "Bird Cams", discrete and continuously recording cameras, near the sites of bird feeders. The Bird Cams featured in this study capture footage of frequently re-stocked feeders in Panama [1].

Bird Cams offer a wealth of data on the frequency with which different species of birds visit the feeder. However, recording bird sightings from Bird Cams is a considerable "bird-en" on investigators. In order to process the hours of footage from the Bird Cams, researchers have turned to "citizen science" to aid in data collection. In order to collect data, Bird Cam footage is live-streamed to a publicly available website. Users of the site watch the Bird Cam and click icons to indicate when different bird species visit the feeder or when fruit is added to the feeder. While this data collection effort offers a unique opportunity to collect large amounts of information, there are also challenges associated with citizen science. For example, because multiple users can view the Bird Cams at the same time, multiple records are collected over popular viewing times. Thus, we must reconcile potentially different records from users that watched the same Bird Cams, but reported different sightings. In addition, we must account for the fact that some time intervals are missing data, due to users not logging-in during some portions of the live-stream. Finally, we must consider that Bird Cam is generally offline overnight, when it would be difficult to identify bird species in poorly-lit footage. As such, data used from Bird Cams has the additional caveat of only being reflective of daytime behavior [1].

The goal of this paper is to develop and utilize data processing strategies that allow researchers to make use of Bird Cam data. We also demonstrate the utility of this data by characterizing sightings of 6 bird species over time, using data from the Panama Bird Cam. This characterization is formalized for one specific species, the Clay-colored Thrush. For this species, we aim to determine the association between time of day, feeder refills, and the number of birds that visit the feeder.

Methods

Bird Cam Data and Data Processing

The data in this study comes from a Bird Cam placed in Panama, with footage recorded from February 11, 2020 to February 25, 2020. The original data included Bird Cam activity, with identifiers for each user and for the time of the observation. Recorded activities included markers for “start” and “stop” of data collection for each user’s session of data collection, an observed species of bird, or an observed fruit refill at the feeder. The bird species that were recorded in this study included Clay-colord Thrush, Crimson-backed Tanager, Gray-cowled Wood-Rail, Gray-headed Chachalaca, Rufous Motmot, and Thick-billed Euphonia.

As the first step in data processing, we generated 1 minute intervals between the first observation in the Bird Cam data and the final observation in the Bird Cam data using the `cut` function in `R` [2]. These 1 minute intervals were matched to observations in the database. Time intervals that did not have a corresponding observation were assumed to have no activity (i.e. no birds were sighted, fruit was not refilled, and no user started or stopped viewing the Bird Cam). Within each 1 minute interval, we summed the number of each type of activity that was reported per user. Because we were primarily interested in feeding times and bird sightings, we dropped markers of user “start” and “stop” times at this point in the analysis.

After the data was split into 1 minute intervals, we aimed to reconcile time intervals that had observations from multiple users. To accomplish this, we first tabulated the number of unique users watching the Bird Cam per 1 minute interval. If there was only one user in a given time interval, we accepted all data that was recorded by that user. If multiple unique users were recording observations simultaneously, we next checked to see if their observations “matched”. “Matched” observations were defined as 1 minute intervals where the observed activity was equal across all users with recorded data. For example, a “match” would occur when two unique users both observed 2 Rufous Motmots during the 1 minute time interval from 9:50 am to 9:51 am on February 20, 2020. An example of a “mismatch” might occur if two unique users reported seeing two Crimson-backed Tanagers during a 1 minute time interval, but another unique user only reported one Crimson-backed Tanager in that interval. In the case of mismatches, we took a weighted average of the observation reports between users, where the weights were determined based on user activity on the Bird Cam site during the entire 2 week period of data collection. To determine these weights, the number of observations that each user contributed to the data collection effort was tallied. In mismatch cases, a user’s weight was determined as the proportion of total observations they contributed, divided by the total of observations contributed by all users present during the mismatch time interval. Thus, if two users in a mismatch contributed 1000 and 10 observations to the Bird Cam database, their weights for that mismatch would be 0.990099 and 0.00990099, respectively. If these users recorded sightings of 2 or 3 birds of a unique species during the contested interval, the final weighted observation for that interval would be $(2 * 0.990099) + (3 * 0.00990099) = 2.009901$. We utilized this weighting scheme for each observed mismatch, and report only the weighted average of each activity per 1 minute interval for the remainder of the study.

At the conclusion of this data processing effort, we obtained a long-form data set with one observation of the number of birds sighted per 1 minute time, per species. Additionally, this data set contained observations of the minute interval in which fruit was added to the feeder. Fruit additions were characterized as a binary variable.

This method of dividing the data into 1 minute intervals and weighting observations by user activity can be utilized to address some of the limitations of Bird Cam data. This method used an informed process to reconcile multiple observations between users and effectively “fills in” time intervals that did not include observations. In addition, because we retained the true time of data recording, subsequent analyses can ignore time intervals that are not of interest (like overnight periods). Before utilizing this data processing method, however, the assumption that empty 1 minute intervals should be recorded as 0 activity observations should be verified by mapping the span over which users were active on the Bird Cam site. If the majority of time intervals did have user interaction, this assumption should be valid. Disagreement between user records over identical time spans were reconciled so that higher value was placed on observations by active users. By weighting in this way, we are making the assumption that active users also tend to be accurate data recorders. This assumption seems realistic and allows for informed reconciliation of records, but the assumption cannot be explicitly checked. In addition, this weighting scheme biases our activity data upwards, because users that have more observations will have higher weight. Users who frequently record observations on the site users will likely also have the higher number of observations in a “mismatch” case, and consequently the higher number of observations will be upweighted.

Characterizing Bird Behavior

Our next goal was to use the processed Bird Cam data to determine when the different bird species arrive at the feeder, and how their behavior is influenced by the timing of feeder refills. First, we visualized the distribution of bird sightings over time by plotting each minute interval of observations, separated by bird species. The timing of feeder refills was also marked on these plots so that the potential association between feeder refills and bird sightings could be visualized.

While the plots provided a general display of bird behavior patterns, we sought to formalize the pattern of bird sightings over time and the impact of feeder refills using statistical modeling. We focused exclusively the Clay-colored Thrush species for this analysis because it was the most commonly observed bird. Before the modeling process, we also reduced our data to only include the hours 6:00 am and 9:00 pm. The remaining hours were ignored because the Bird Cam was generally not recording in the overnight hours and any data present in that interval would have been the result of "filling in" missing 1 minute intervals with 0 activity observations in the first data processing step. We also omitted any days from our analysis that did not include sightings of Clay-colored Thrushes or of feeder refills.

We fit models predicting the number of Clay-colored Thrush sightings per minute from time and feeder refill indicators. Models were fit by day, because we were interested in average daytime behavior of the birds and wanted to ensure our method could be utilized to compare bird behavior by day, for further research endeavors. Each daily model fit was pooled using Rubin's Rules, to produce an average daily behavior model [3, 4]. Because our outcome was a count over time, we utilized generalized linear models with a Poisson outcome and a log link [5]. In the event that observation data that was reconciled from mismatch cases and the reconciliation yielded a non-integer count, we rounded the recorded observation to the nearest whole number. No offset term was included in the model because all observations were taken over the same 1 minute intervals and we wished to retain the expected count interpretation of the model. The predictors in our model were time (minutes since 6:00 am), time squared, and an indicator of feeder refill within the past 60 minutes. A squared time term was included in the model due to the observed parabolic shape of bird sightings by day. We incorporated data on feeder refills through an indicator variable. If the feeder had been refilled within the previous 60 minutes, this indicator variable was "1". If the feeder was not refilled within the previous 60 minutes, this indicator variable was "0". Such an indicator was selected for these models because bird behavior appears to be clustered near feeding times. The interval of 60 minutes seems biologically plausible given our graphs of bird behavior, and the time window balances the strength of our highly temporal data with the necessity of a moderate amount of events for model fitting. Previous models utilized smaller, and more specific feeding windows, but suffered from complete separation issues because there were sometimes too few feeding cases [8].

In this report we present the pooled model fit across days, predicting Clay-colored Thrush sightings by minute, with anindicator of food refill within 60 minutes. We present this model to provide one example of how citizen science Bird Cam data can be used to analyze bird behavior.

Results

The original data set consisted of 12,019 observations. After our data processing steps, we recorded 4,169 1 minute intervals with bird or feeder refill activity. The number of 1 minute intervals with activity for each of the bird species and for refill activity is presented in **Table 1**. The most common siting was the Clay-colored Thrush, and the most rare bird to visit the feeder was the Rufous Motmot. Fruit was added to the feeder 73 time during the two weeks of data collection.

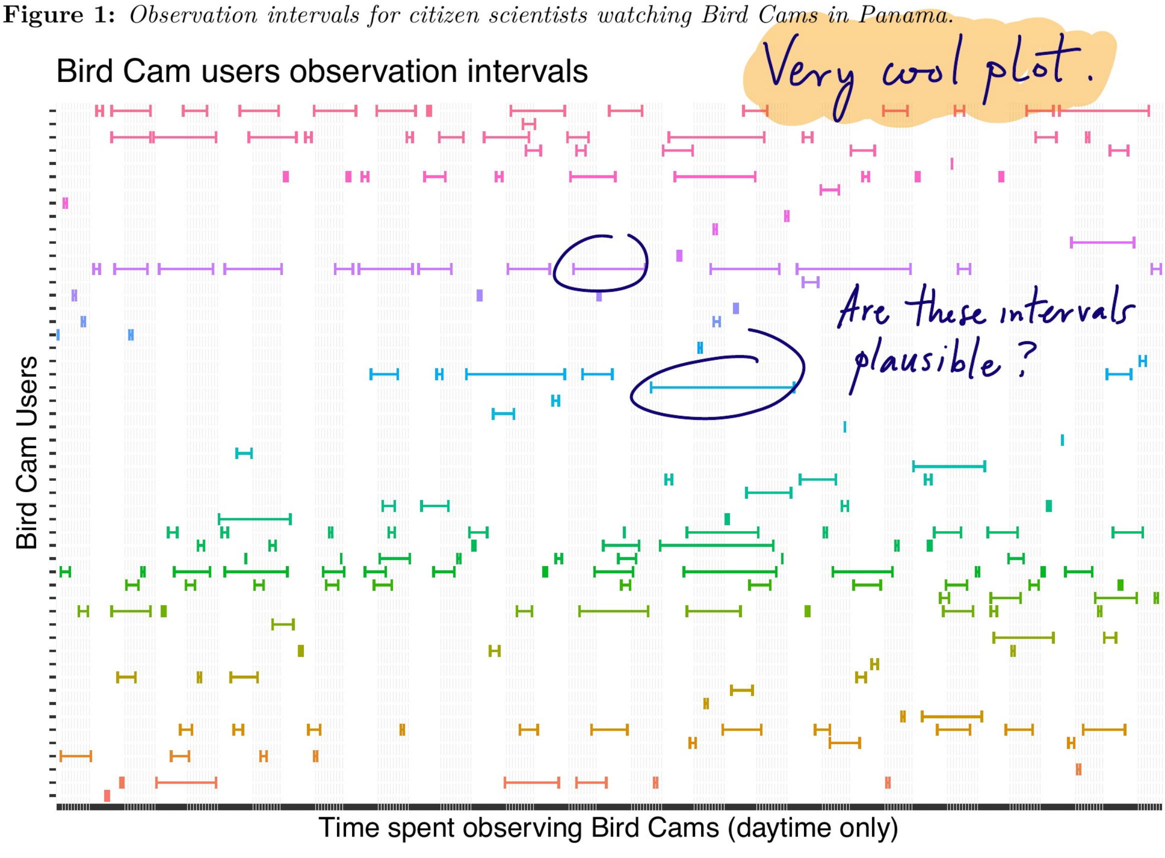
Table 1: *Number of time intervals (1 minute each) with observed activity. Observed activity could be bird species sighting, or observation of the bird feeder being refilled with fruit.*

Observed Activity	Number of Time Intervals with Activity
Clay-colored Thrush	2274
Crimson-backed Tanager	638
Gray-cowled Wood-Rail	138
Gray-headed Chachalaca	386
Rufous Motmot	52
Thick-billed Euphonia	608
Fruit was added	73

After conducting our data processing, we wanted to check the validity of our assumption that daytime 1 minute time intervals without observation data could be reasonably estimated as having 0 activity. To verify this assumption, we plotted spans of user activity over time in **Figure 1**. Each row (and color) of data corresponds to a unique Bird Cams user, and intervals where the user was active are marked by colored bars. Because there is a great deal of overlap between user observation intervals, we can be confident that times without observations are truly indicative of 0 activity. Since users were watching the Bird Cam frequently, we can be confident that most activity was recorded.

While there was a great deal of overlapping use of the Bird Cam, there was not a considerable amount of disagreement between users. Of the 6,468 recorded observations, only 668 were mismatched between users, representing just 10% of the total data collection effort. This low rate of disagreement between users suggests that Bird Cams are a viable method of data collection for studying bird behavior.

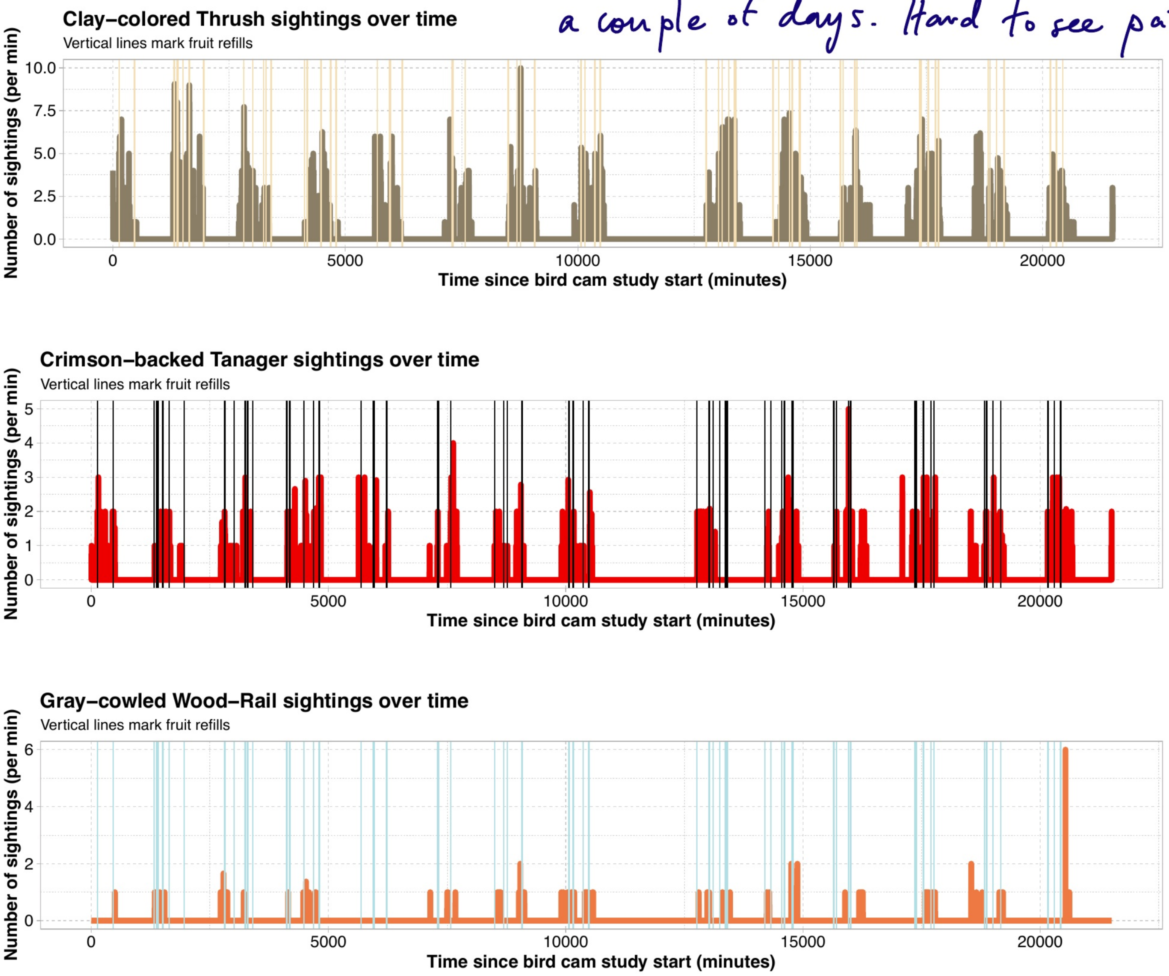
Figure 1: *Observation intervals for citizen scientists watching Bird Cams in Panama.*

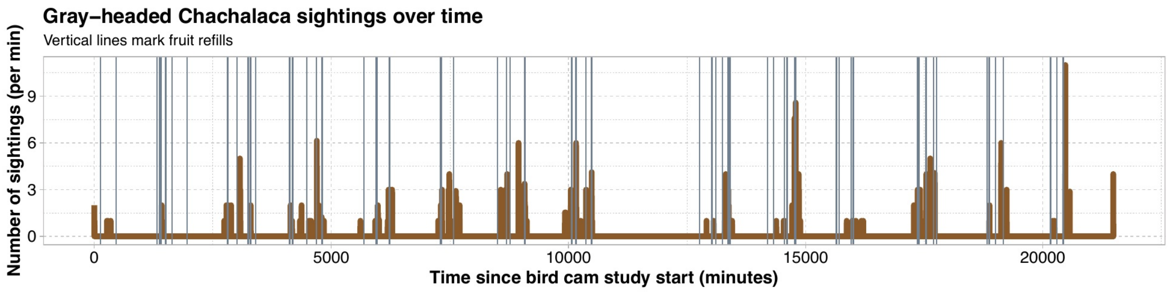


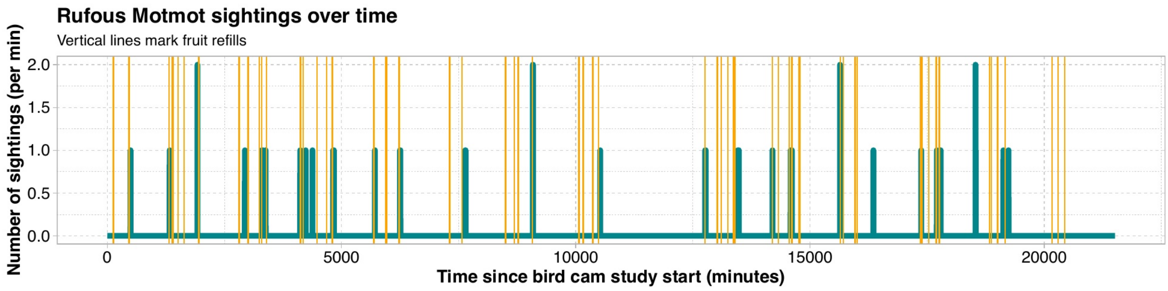
After cleaning the data and verifying our assumptions, we sought to analyze bird behavior patterns over time. **Figure 2** displays the number of bird sightings per minute since study start, in separate plots for each bird species. Vertical lines indicate times when the feeder was replenished. Data are clustered in daylight hours, with 0 observations during nighttime hours because data was not collected overnight. Overnight hours were still included in the plot, however, to facilitate distinction and comparison across days of observation. A distinctive gap is present near the middle of the plot, as no observations were recorded on February 18, 2020. Bird observations appear relatively stable across days, though there is a distinctive uptick in observations of Gray-cowled Wood-Rail and Gray-headed Chachalaca species in the final days of data collection. Clay-colored Thrush, Gray-headed Chachalaca, and Thick-billed Euphonia species generally had the greatest number of sightings in a given minute interval. This likely means that when these bird species visit the feeder, they come in groups. The remaining birds species are often observed in singles or pairs when they visit the feeder. These plots also allow us to get a sense of how feeder refills and time of day impact bird observations. Some species, like the Clay-colored Thrush, tend to be present at the feeder during most daylight hours with only small lag between feeder refills and observations. Other species, like the Rufous Motmot, appear to only visit the feeder at the very beginning or very end of each daylight cycle, and a considerable lag is observed from the final feeder refill of the day to an evening sighting.

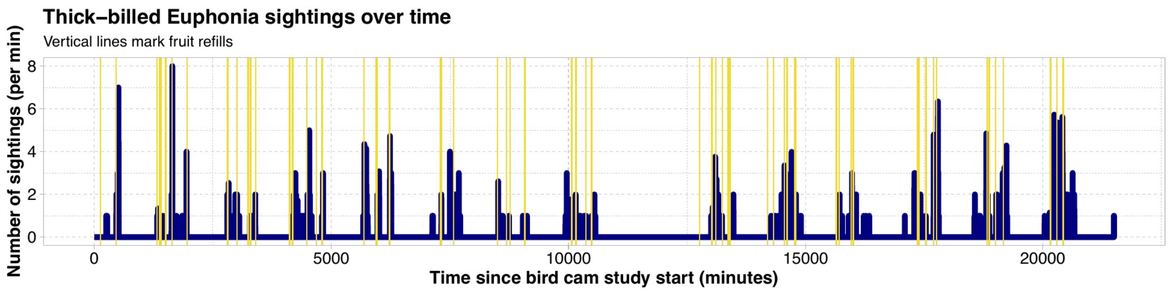
Figure 2: *Number of bird sightings per minute of the Bird Cam study, since study start, displayed by bird species. Vertical lines indicate times when the feeder was refilled with fruit. Plots are designed to match the coloring of the bird species they represent.*

Might be better to zoom in on a couple of days. Hard to see patterns here.









Finally, we present the results of our analysis of Clay-colored Thrush birds over time. After applying Rubin's Rules to pool model results across all days of analysis, we present the model results in **Table 2**. On any given data, for each 1 minute time, we expect the log-number of Clay-colored Thrush birds observed per minute on the Bird Cam to increase by 0.05, on average. A feeder refill that occurred within 60 minutes prior is associated with a 0.88 increase in the log-number of Clay-colored Thrush birds observed on the Bird Cam per minute.

Table 2: *Parameter estimates of pooled generalized linear model with Poisson outcome and log link. Outcome is number of Clay-colored Thrush birds observed in a 1 minute interval. Predictors are time in minutes since 6:00 am, time squared, and an indicator of the bird feeder refill within 60 minutes. Models were fit separately for each day of the study and model results were pooled using Rubin's Rules.*

	Estimate (log scale)	Standard Error	p-value
Minutes since 6am	0.04822	0.07922	0.56216
Minutes since 6am, squared	-0.00002	0.00002	0.36799
Feeder refill within an hour	0.88436	0.64973	0.18927

I would use sin and cos functions of time of day.
Maybe 3-6 harmonics.

Discussion

In this paper we presented a method for cleaning and utilizing data from Bird Cams, where data collection methods rely on citizen science. Our methods were strengthened by the use of short time intervals for data analysis. We were able to capture observations within 1 minute intervals, which is a plausible interval to assess agreement over between Bird Cam viewers. It is reasonable to assume that multiple viewers are reporting the same observations within 1 minute intervals. If we had used larger viewing intervals to reconcile disagreement in or generally collapse observations between Bird Cam users, it is possible that different user records were actually referencing different observations of the Bird Cam. Such nuance would be lost with longer analysis intervals. Moreover, the use of short-term intervals allows us to analyze bird feeding behaviors with granularity. While other work is limited to making general claims about bird behavior over multi-hour time frames (i.e. morning, midday, evening, etc.), we harness the rich data of the feeders to study behavior in greater detail [6].

We utilized Bird Cam data to create visual displays of each bird species' visits to the feeder over the 2-week study period. These patterns were analyzed in greater detail for the Clay-colored Thrush species. This detailed analysis utilized generalized linear mixed models with a Poisson distributed error model and a log link to capture the impact of time and feeder refills on the number of Clay-colored Thrush birds that visited the feeder per minute. Generalized linear models are a common choice for analyzing bird feeding behavior [6], but we departed from standard practice by fitting several day-specific models. This approach was taken, rather than "nesting" (Get it? This is a bird pun) time within specific days in a generalized linear mixed effect model, for instance, so that we could directly compare models for each day of study. While we did not utilize the advantage of this modeling strategy directly in this paper, our work opens up the possibility that future researchers can utilize our method to additionally explore the variation in bird feeding behavior by day. Time series analysis was another modeling strategy considered for this study, but this effort was abandoned due to computational inefficiencies [7].

In conclusion, we defined a data processing method to analyze data derived from live-streamed "Bird Cams" in Panama. We present detailed graphical displays of bird feeding behavior over time, and formalize a behavior model for Clay-colored Thrush birds. We hope that our methods and models can contribute to continued research utilizing data from Bird Cams.

Nice report and nice plots. The user interval plot was especially interesting.

It was smart to adjust for # of users.

It might be safer to model the data a binary,
1 = at least 1 bird is present
0 = no birds present.
This is common in ecology. While you lose the abundance info, you still get information about presence/absence.

References

[1] Mady R. "Bird Cams Lab". *In-class presentation.* (2020).
[2] The R Foundation. (2020). R Version 4.0.2. `https://www.r-project.org/`
[3] Heymans MW, Eekhout, I. "Rubin's Rules" *Applied Missing Data Analysis with SPSS and (R) Studio.* 2019. `https://bookdown.org/mwheymans/bookmi/rubins-rules.html`
[4] van Buuren, S. "Package 'mice' ". `https://cran.r-project.org/web/packages/mice/mice.pdf`
[5] Kent D. "A fish model for birds. (Poisson joke ... because Poisson means fish in French)". *iMessage conversation.* 22 Oct. 2020.
[6] Bonter DN, Zuckerberg B, Sedgwick CW, Hochacka WM. "Daily foraging patterns in free-living birds: Exploring the predation-starvation tradeoff". *Proceedings of the Royal Society B.* 20 March 2013.
[7] Hochstedler KA. "My computer crashed 3 times within the first hour of attempting a time series analysis". *Confessions.* 23 Oct. 2020.
[8] Salah S. "Why would there be a huge standard error for a simple logistic regression?". *ResearchGate.* 23 Oct. 2020.