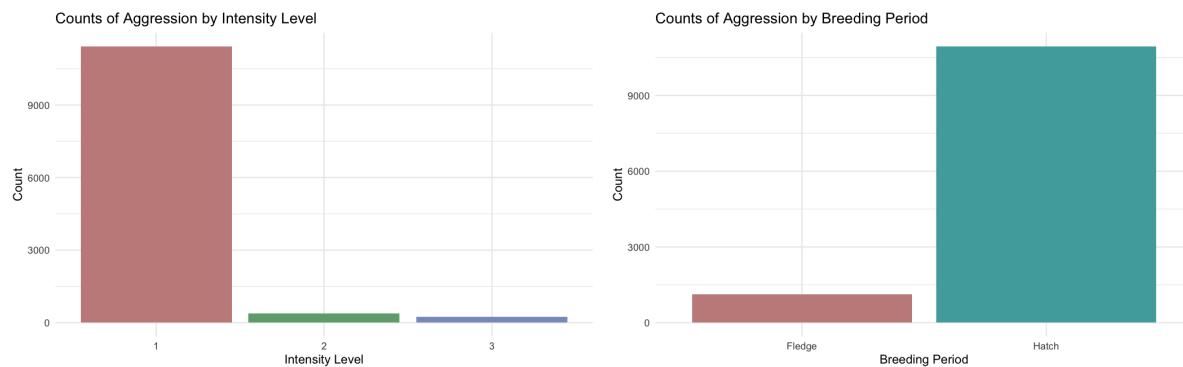


The dataset is a subsetted dataset consisting of 70% of the real dataset and only morning observations. There are 12,056 counts of aggression spanning 13 noncontinuous days from June 15, 2022 to July 7, 2022. The first 10 days observed in the data were during the hatching period, while the final three, July 6, 7, and 11, were during the fledgeling phase after most eggs have hatched. The variables included in the subsetted dataset are **date**, **breeding_period**, **observation** (stratified time of day), **stint_start**, **stint_end**, **nest** (the aggressive bird's nest of origin), **nearest_nest** (the nearest nest to the observation, if the aggressive bird is not near its nest of origin), **on_off** (how close the aggressive bird is to the nest), **sitting_standing**, code (intensity of observation, from 1-3), **description**, **duration**, **participant**, **file_name**, **comments**, **segment**, and **standing_on_camera**. Some of these variables will be removed as they are irrelevant to the questions we seek to answer or we do not have access to the details they provide.

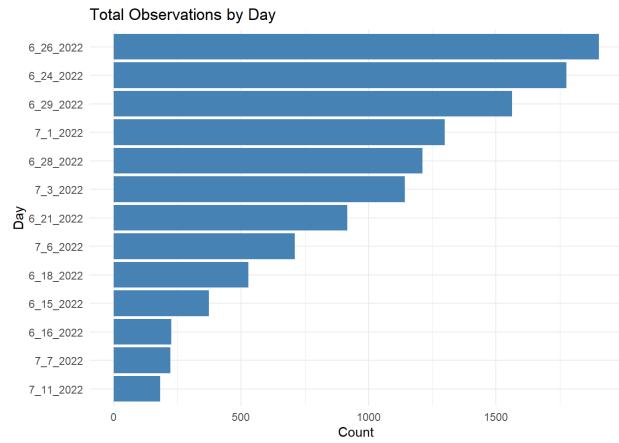
We cleaned the dataset by removing irrelevant variables and rows with missing values for specific variables. The variables removed were **standing_on_camera**, **comments**, and **file_name**. These variables don't contribute at all to our analysis or are entirely missing (**standing_on_camera**). We also removed rows that had missing values for the variable **duration**.

In our initial analysis of the dataset, we made several quick discoveries. First, the vast majority of aggressive incidents were registered at the lowest level of intensity, which includes things such as kip calls, attack calls, fear calls, or alarms. Events at level 2 aggression include walking or flying towards other terns. Finally, violent conflict including pecks, ground fights, and dive bombing are considered level 3 aggression. Further observations include the majority of events occurring near the tern's nest of origin, and the vast majority of observations occurring during the hatching period.

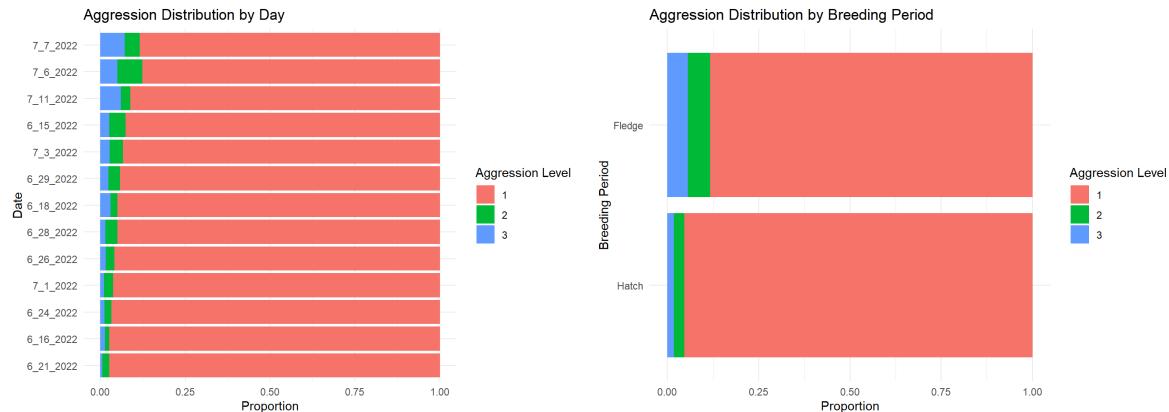


To better understand the data, we counted the number of observations recorded in each of the 13 days. There was a large disparity between the day with the most observations, June 26, 2022 with a count of 1903 observations, and the day with the least observations, July 11, 2022, with only 182 observations. Six days had a count of over 1000 observations, while the other seven had fewer than 1000. During the final three days in our sample, the fledgeling phase, each day had an

average of 371 observations, while the average observations for the hatching phase was 1094, nearly three times as much.



This reveals a disparity between the number of observations during the height of the hatching period in late June and early July, and the beginning of the hatching phase and the fledgling phase. In order to understand how aggression factors into these observations, we observed the proportion of incidents on each level per day. The following bar chart shows this proportion, ordered from the day with the highest mean aggression to the day with the lowest mean aggression.



Interestingly, the three days with the highest mean aggression are those during the fledgeling period after most of the chicks have significantly matured. It is important to note that although these days have high mean aggression, the total number of aggressive incidents is still much higher during the peak of the hatching phase. This simply means the aggressive events that occurred during those days were more likely to be of higher intensity. There may be a small observation bias towards higher intensity violence on days with fewer observations. The net correlation between mean violence and number of daily occurrences is -0.51, which appears to be explained by differences in aggression during the hatching and fledgling phases. The

correlation is only -0.16 among days during the hatching period, and 0.24 during the fledgling period.

The resulting dataset includes 18 variables and 12050 observations. 5 of which are numeric and 13 are categorical. n, time start, time end, time, and duration are numeric variables while date, breeding_period, observation, camera, nest, nearest_nest, type, on_off, sitting_standing, code, description, participant, and segment are categorical variables.