Final Project Proposal - Bird Feeder Count

Statement Of Goals

- Understand How Predictors Relate: Investigate how the presence of animals (squirrels, cats, and dogs) relates to housing density. Determine if there is a connection between these predictors and whether the current model already accounts for these relationships or if adjustments are needed to capture these effects.

- Summarize and Visualize Raw Data: Include in the description of the raw data more details with numbers and visualizations. The overall number of observations for each category which are squirrels, cats, dogs, combinations of animals, and none, as well as the distribution of these data across various housing density levels should be included.

- Study how animal presence and housing density together influence bird counts. For example, analyze how the average bird count changes with density for different combinations of animal presence. Use these insights to strengthen the understanding of the predictors and their effects.

- Have a modeling approach to ensure it provides clear and meaningful interpretations. It also provides more clarity to the numbers and patterns observed in light of easy-to-understand results that relate directly to the key research questions concerning bird behavior, housing density, and animal presence.

Data Collection Criteria

1. **Animal Presence**:
    - Squirrels: Observers record the presence of squirrels if they take food from feeders at least three times per week.
    - Cats/Dogs: Cats or dogs are considered "present" if they are active within 30 meters of feeders for at least 30 minutes, three days per week.
      These specific thresholds ensure consistent and reliable recording of animal presence across observers.
2. **Bird Count:**
    - The bird count is not a cumulative total of every bird seen but rather the maximum number of individuals of a species observed at the same time at the feeder. This simplifies manual tallying and ensures data integrity.
    - Observers are encouraged to select two days per week where they have the most time to observe throughout the day. This structured approach improves observation consistency.
3. **Season and Year:**
    - The FeederWatch season runs annually from November 1st to April 30th.

- For example, the 2024 data corresponds to the period from November 2023 to April 2024. Since the current season has ended, the 2024 data is complete and reliable.

Data Description and Graphs

The dataset originates from the Cornell Lab's Project FeederWatch for the years 2021 to 2024. This large-scale citizen science effort involves participants observing and recording annual bird counts at supplemental feeding stations across North America. The dataset captures key variables that allow for the study of bird activity in relation to environmental and contextual factors, particularly housing density and animal presence

The main variables in the dataset include Bird Count, which represents the total annual count of birds observed at each feeder site. Housing Density is a categorical variable that is modified to show the level of urbanization in the following way: 1 = Rural, low density, 2 = Suburban, moderate density, 3 = Slight Urban, higher density, and 4 = Urban core, highest density. Animal Presence includes binary indicators (1 = present, 0 = absent) for squirrels, cats, and dogs, along with multiple combinations of these animals under the column animal_type (no cat, no dog, dog, etc). The Year refers to years the data was taken which is from 2021 through 2024.

To walkthrough how our dataset looks like

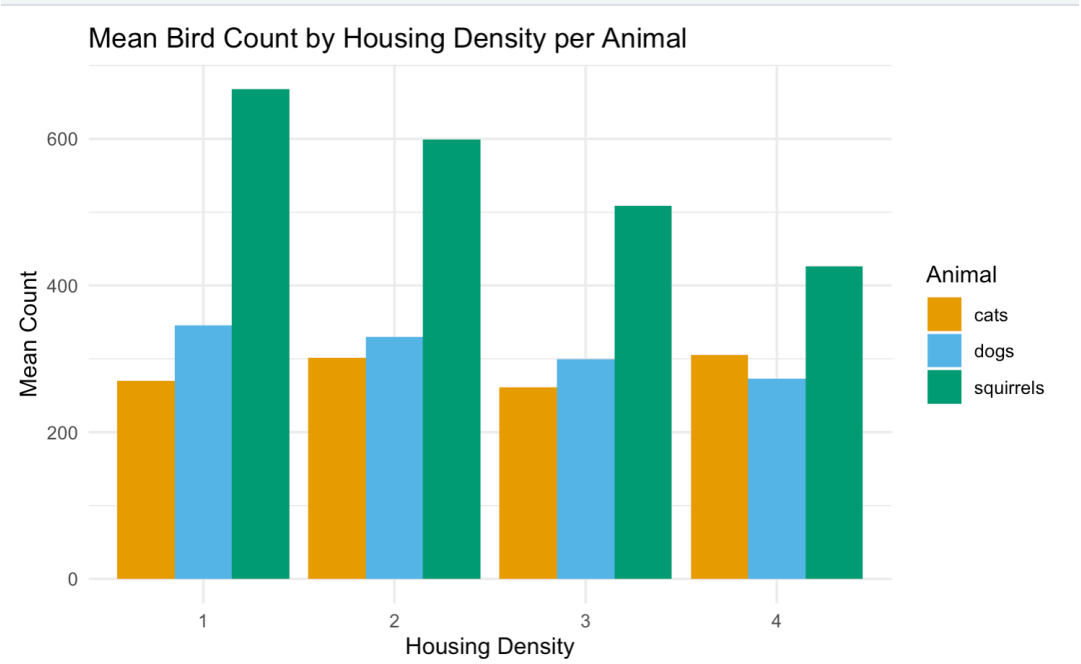
	LOC_ID <chr>	PROJ_PERIOD_ID <chr>	count <int>	squirrels <int>	cats <int>	dogs <int>	huma... <int>	housing_density <fctr>	animal_type <fctr>
1	L100032	PFW_2021	1145	1	0	0	1	2	squirrel

Each column in the dataset represents specific information about bird counts at feeder sites. *LOC_ID* is the unique identifier for the observation site. *PROJ_PERIOD_ID* refers to the FeederWatch season (e.g., PFW_2021). *count* indicates the maximum number of birds observed at the same time. *squirrels, cats, and dogs* are binary indicators (1 = present, 0 = absent) for animal activity near the feeder. *humans* are calculated similarly (1 = present, 0 = absent) but are out of focus for this research. *housing_density* represents the urbanization level, with 1 as rural and higher numbers indicating increasing density. Finally, *animal_type* categorizes the site based on the presence of specific animals, such as squirrels in this case.

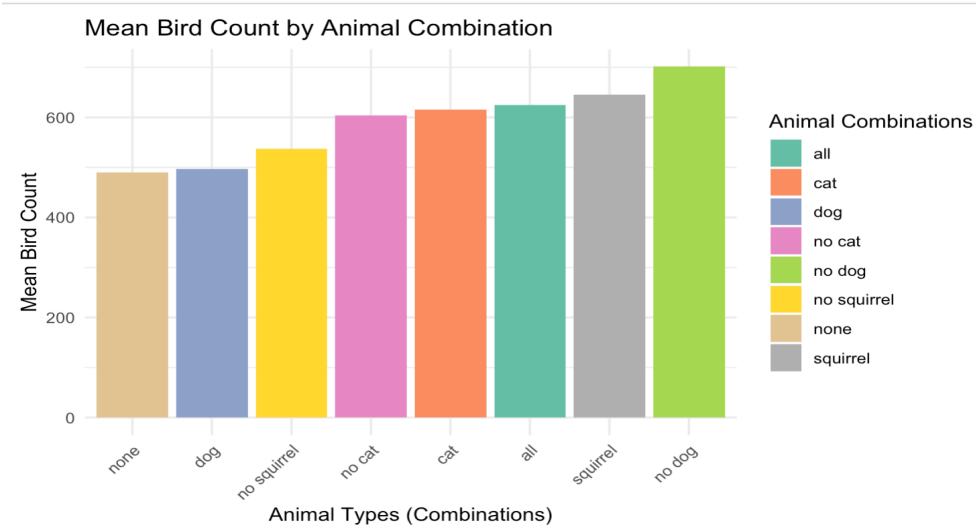
The data highlights the uneven distribution of observations across animal types. Summarized totals for each group are as follows:

- Squirrels: 39,871 observations (most frequent)
- Dogs: 23,929 observations
- Cats: 19,496 observations
- None: 2,533 observations (least frequent)
- All Animals Present: 9,613 observations
- Total: 47693

These totals reveal that sites with squirrels are significantly more common, aligning with their observed impact on bird counts. Sites without animals are rare, which may reflect reporting biases or the prevalence of wildlife in feeder environments. The raw data, while extensive, is unevenly distributed across housing densities and animal combinations. For instance, rural and suburban areas (densities 1 and 2 with 11521 and 12711 entries respectively) contain more observations compared to urban cores (density 4 with just 3748 entries respectively), and certain animal combinations occur more frequently than others. So we only account for the mean count by housing density per animal.



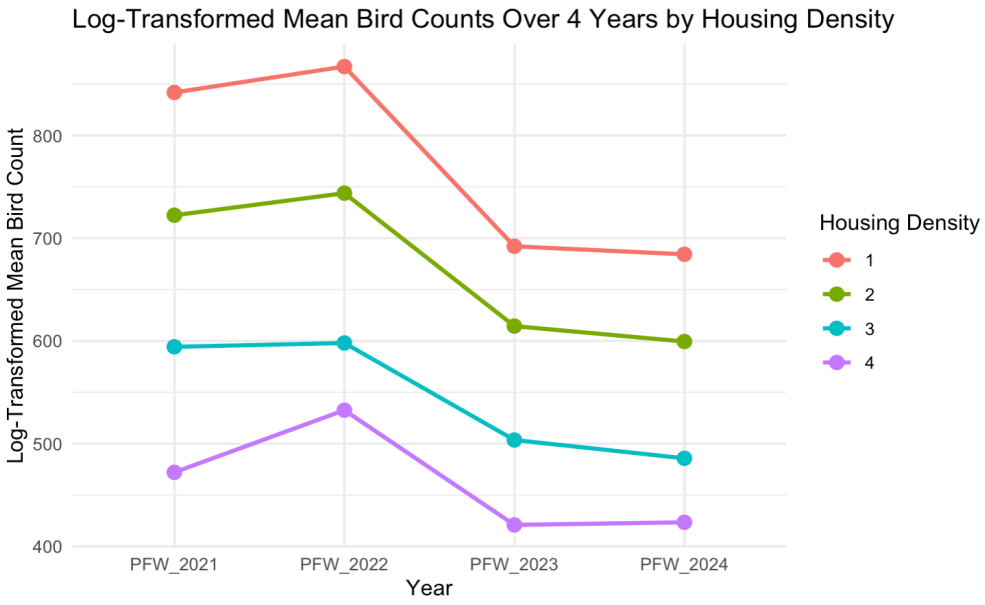
This graph shows how animal presence differs across housing densities and its relationship with mean bird counts. The presence of squirrels is most notable in rural (density 1) and suburban (density 2) areas, where mean bird counts are significantly higher (~600 birds). In contrast, urban areas (density 4) show a decline in bird counts, though squirrels still maintain a positive impact compared to cats and dogs. Dogs and cats are distributed more evenly across housing densities, but their impact on bird counts is lower than that of squirrels. Specifically, rural and suburban areas exhibit higher bird activity when squirrels are present, whereas urban cores, despite the presence of animals, report the lowest mean counts. This distribution reflects the differing habitat quality and disturbance levels across urbanization categories.



The graph shows the mean bird count for various combinations of animal presence, including squirrels, cats, dogs, and their coexistence. This analysis aligns with the project's goal of understanding how animal presence impacts bird visits to feeders. The highest mean bird counts are observed at sites where no dogs are present (~2,700), suggesting that dogs may disturb bird activity near feeders. Sites with all animals present (squirrels, cats, and dogs) also report high bird counts (~2,500), indicating that while dogs may disrupt feeders, the presence of squirrels and cats can signal richer habitats that attract more birds.

Squirrels alone are associated with consistently high bird counts (~2,400), reinforcing their role as indicators of favorable environments. In contrast, sites with no squirrels show lower counts (~2,100), highlighting their positive influence on bird presence. Similarly, sites with cats alone average around 2,300 birds, suggesting that cats, like squirrels, may indicate productive habitats. Sites with dogs alone display lower bird counts (~1,800), supporting the idea that dogs deter bird activity. Finally, sites with no animals present report the lowest mean bird counts (~1,700), which may reflect less wildlife activity or lower habitat quality.The findings reinforce that animals such as squirrels and cats positively correlate with higher bird counts, while dogs may act as a deterrent. Furthermore, the coexistence of animals highlights richer habitats, which aligns with the hypothesis that areas with multiple species are more ecologically diverse and supportive of higher bird activity.

Mean Bird Counts by Year and Housing Density

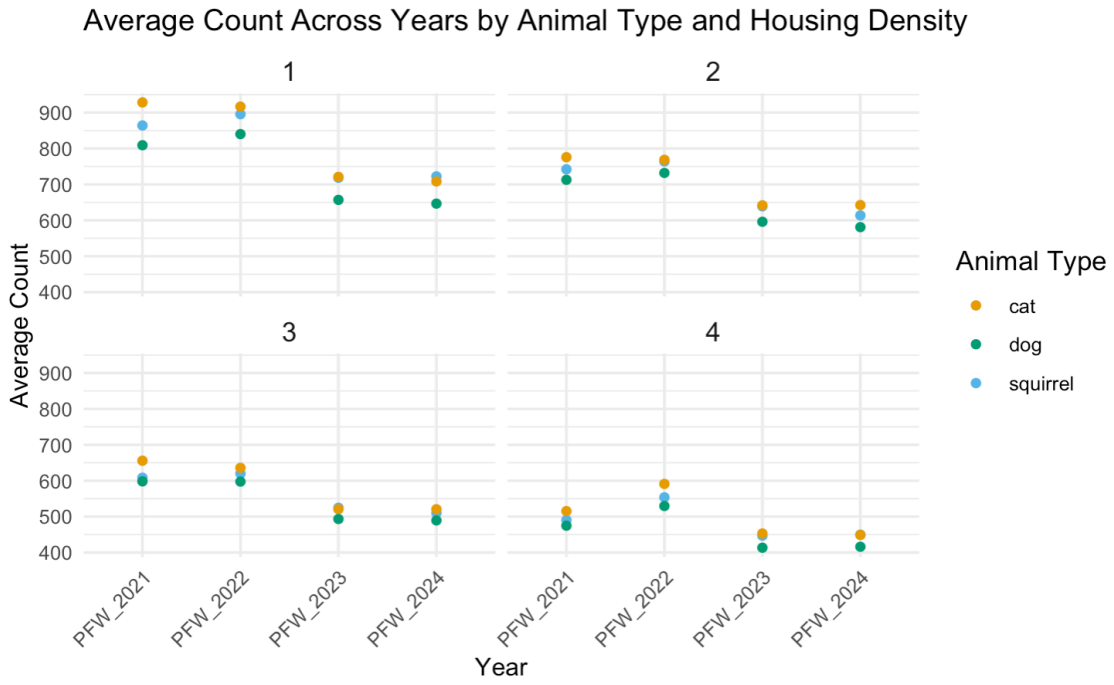


The first graph shows the trend in mean bird counts over the years (2021–2024) across four levels of housing density: rural, rural/suburban, suburban, and urban.

- Rural areas consistently exhibit the highest bird counts, averaging above 800 birds in the earlier years (2021 and 2022). However, there is a clear decline from 2023 onward, where the mean count stabilizes around 700 birds. This indicates a downward trend despite rural areas generally supporting higher bird activity.
- Rural/suburban areas follow a similar trend but at slightly lower levels. Bird counts remain between 700 and 600 across the years but display a gradual decline after 2022.
- Suburban areas maintain moderate bird counts with averages near 600 birds in 2021 and 2022, which then drop to approximately 500 birds by 2024. Suburban areas appear to show a consistent decrease in bird activity over time.
- Urban areas report the lowest mean bird counts, starting below 500 birds in 2021. The counts show a slight peak in 2022, followed by a drop to below 400 birds in 2023 and 2024, indicating limited bird presence in high-density urban environments.

  The clear trend across all housing densities is a general decline in bird counts over time, with rural and rural/suburban areas consistently outperforming suburban and urban areas. This reinforces the observation that lower housing density supports higher bird activity, while urbanization negatively impacts bird presence.

Average Bird Count by Housing Density, Year, and Animal Presence:



The graph shows average bird counts by year, housing density, and the presence of cats, dogs, and squirrels across four housing density panels: rural, rural/suburban, suburban, and urban.In rural areas, bird counts are consistently the highest, with cats and squirrels associated with the highest counts (approaching 900), while dogs show slightly lower averages, particularly in 2023 and 2024. Rural/suburban areas follow a similar pattern but at lower levels, with a gradual decline in counts over time.

In suburban areas, bird counts range between 500 and 600, with squirrels and cats maintaining higher averages compared to dogs. A noticeable decline occurs from 2022 onward, reflecting a less supportive environment for birds. Urban areas have the lowest bird counts (below 500) across all years and animal groups, though squirrels and cats still contribute to relatively higher counts compared to dogs.

Overall, squirrels and cats consistently align with higher bird counts across all densities, while dogs are associated with lower counts. The graph highlights a clear trend of decreasing bird activity with increasing housing density, with rural areas supporting the most birds and urban areas the least.

Model Selection:

Model, AIC, and Deviance Results:

Our model: glm.nb(count ~ PROJ_PERIOD_ID + housing_density * squirrels + housing_density * cats + housing_density * dogs)

We fit the model count ~ PROJ_PERIOD_ID + housing_density * squirrels + housing_density * cats + housing_density * dogs using different GLM methods to analyze bird counts.The Poisson model has an AIC of 23986645.44 and a deviance of 23620709.72. The Quasi-Poisson model does not provide an AIC but has the same deviance as Poisson (23620709.72). *The Negative Binomial model shows a much lower AIC of 706491.25 and a deviance of 54915.55*. For interaction effects, the AIC with housing density and animal interaction is 706491.25. The AIC with period and housing density interaction is slightly higher at 706531.47. The AIC with no interactions stands at 706520.77. We played around with multiple variables to get optimal AIC.

We find that the negative binomial model outperforms the other two significantly. We also find that a model containing an interaction between housing density and the animal types outperform models with no/different interactions.

Let's look at the exponentiation of coefficients to interpret them.

```
##                             (Intercept)                  PROJ_PERIOD_IDPFW_2022
##                             719.9026832                               1.0237191
##                  PROJ_PERIOD_IDPFW_2023                  PROJ_PERIOD_IDPFW_2024
##                               0.8438000                               0.8252321
##           housing_densityrural/suburban                housing_densitysuburban
##                               0.8352105                               0.5923967
##                     housing_densityurban                              squirrels
##                               0.4542240                               1.2061956
##                                    cats                                    dogs
##                               1.1027796                               0.9249746
## housing_densityrural/suburban:squirrels       housing_densitysuburban:squirrels
##                               1.0107001                               1.1493809
##          housing_densityurban:squirrels      housing_densityrural/suburban:cats
##                               1.2263926                               1.0114195
##            housing_densitysuburban:cats              housing_densityurban:cats
##                               1.0184158                               1.1134433
##      housing_densityrural/suburban:dogs            housing_densitysuburban:dogs
##                               1.0260320                               1.0526456
##               housing_densityurban:dogs
##                               1.0067213
```

Our coefficients are relative to a base site: a bird feeder in a rural area in 2021 with no animals present. Year-wise, the model shows a decline in bird counts post-2022, which aligns with the patterns observed in the data graphs. For the housing density and animal interactions, we can illustrate their effects using specific examples. The intercept is 719.9, representing the predicted bird count at the base site. For a feeder classified as suburban, the coefficient for housing_densitysuburban is 0.592, meaning the bird count would be multiplied by a number less than 1, resulting in a decline. If squirrels are observed at this site, we multiply by the interaction coefficient housing_densitysuburban:squirrels (1.01), which has a minor positive effect, and by the squirrels coefficient (1.21), providing a larger boost to the bird count.

This process can be repeated for any combination of density, animal presence, and year. Among the coefficients, the factors with the strongest negative effect on bird counts are the

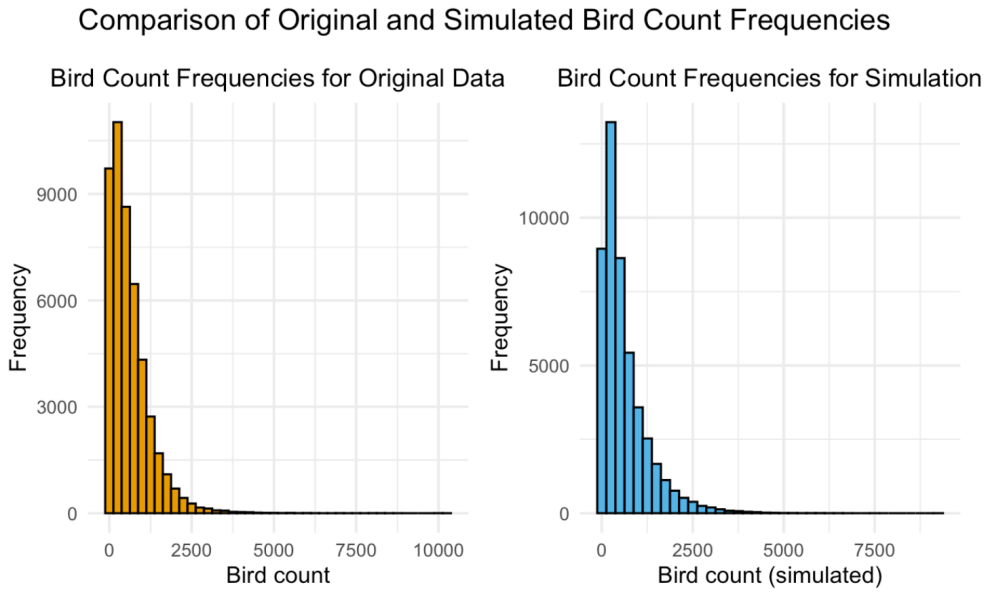
feeder's location in urban areas (0.454) and suburban areas (0.592), reflecting the impact of higher housing density. On the other hand, squirrels consistently have a positive effect across all densities, with the greatest boost observed in urban areas (1.23), where they appear to counterbalance some of the negative effects of urbanization

For example a feeder in 2023, located in a suburban area with cats present, the predicted bird count starts with the intercept of 719.9. It is then adjusted by multiplying with the year coefficient for 2023 (0.843), followed by the housing_densitysuburban coefficient (0.59), the cats coefficient (1.10), and the housing_densitysuburban:cats interaction (1.01).

Combining these factors, the predicted bird count is:

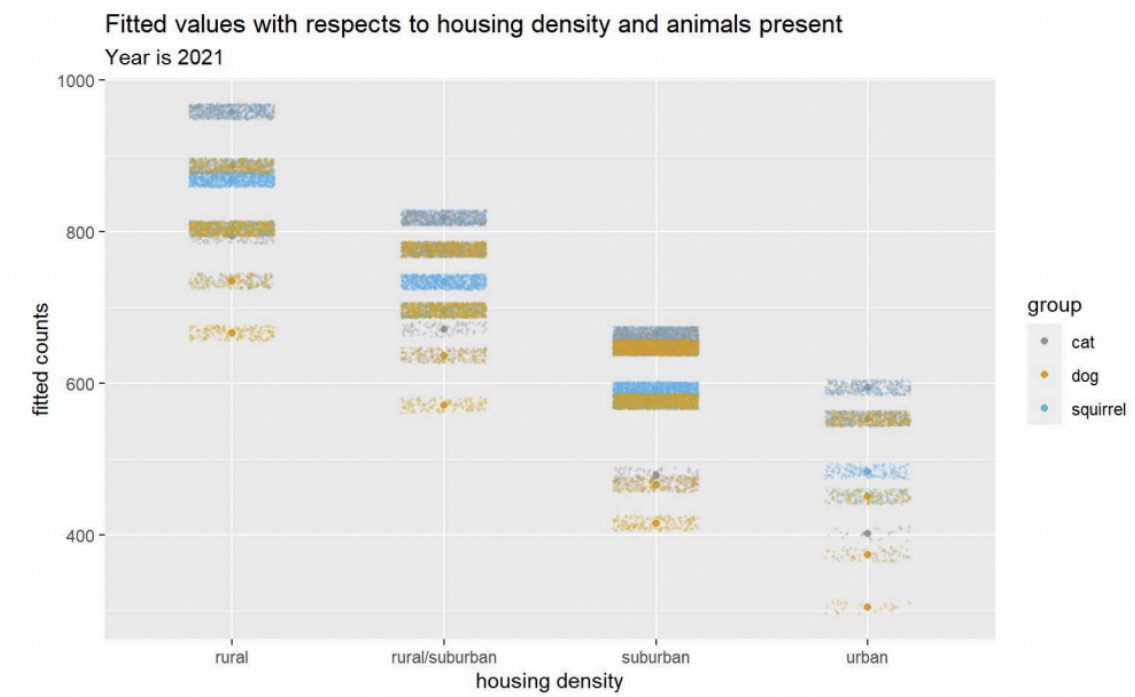
719.9×0.843×0.59×1.10×1.01≈393.37. Thus, for this case, the bird count is approximately **393**.

Now lets plot our simulated data and the actual data:



The shapes of the original and simulated bird count frequencies are similar, both displaying a right-skewed distribution with the majority of counts concentrated at lower values. In both histograms, the frequency of bird counts decreases as the count increases, with very few observations exceeding 2000. This indicates that the simulation successfully captures the key distributional properties of the original data, including the long tail at higher bird counts. The alignment in shape suggests that the simulated data reflects the patterns observed in the original dataset well.

Now we know the modeling looks better, and let's check the fitted values now in the year 2021.



The main indication of this graph is how bird counts are divided across the different options for animal presence. Each bar represents a collection of different colored points, with the top bar including blue and red (birds and squirrels), the next containing all three colors (squirrels, cats, and dogs), and so on. This breakdown aligns with the trend observed in the earlier graph. At every level of housing density, the same pattern emerges for bird counts, ranked from highest to lowest: birds and squirrels, all three animals, just squirrels, dogs and squirrels, just cats, cats and dogs, and finally just dogs. This consistent trend highlights the influence of animal presence on bird counts across all housing density levels.

We can put our coefficients and visuals together to draw some answers to our research question.

First, we find that there was a harsh drop in bird count from 2022 to 2023, but other than that, general trends stay the same across all years (by almost remaining constant after that. Second, let's talk about animals, squirrels being present at a site are a positive indication of bird count. Their coefficient of 1.2 as well as a positive coefficient for every housing density shows that if squirrels are present, recorded bird count goes up. A safe assumption is that squirrel presence indicates a sufficient amount of food at a feeder, which in turn leads to a high bird count.

Cat presence also has a positive effect on bird count. Their coefficient of 1.1 and similar positive coefficient at every housing density shows that if cats are present, recorded bird count goes up. It's likely this is because a feeder with more birds is more likely to have a cat present, as cats are fond of birds!

Dogs actually seem to have a negative effect on recorded bird count. Their coefficient of 0.925 brings down the predicted bird count, and none of the coefficients of the housing_density * Dog interactions are large enough to cover this difference. The question is then, why do cats have a positive effect but dogs don't? Perhaps dogs aren't drawn to high bird count feeders like cats may be, but also, dogs are simply louder and more hyper than cats. Something interesting worth noting is the best performing sites have coexistence between these mammalsI suppose if an environment has the capacity for various animal types it likely has room for solid bird feeders as well!

Third would be the housing density. This one is quite simple to see and understand ; the more urban you get, the less birds you should expect to count. Many birds enjoy eating in more peaceful environments, which, due to the presence of humans and other animals, is harder to attain as you get more urban. Finally interactions ! Looking at coefficients, we can see all animal types have an increasingly positive effect on bird count as you get more urban. This is likely due to the same reasons as animal coexistence; not all urban/suburban homes have adequate space for a bird feeder, but if other animals are present the home is more likely to be suitable.

Conclusion

Our analysis highlights clear relationships between bird counts, housing density, animal presence, and yearly trends, supported by both the data and model results. Bird counts have shown a decline post-2022, with coefficients for 2023 (0.84) and 2024 (0.83) indicating lower counts compared to the base year, 2021. This trend aligns with patterns observed in the data, suggesting external factors or habitat changes over time.

Housing density remains a key factor, with bird counts decreasing as density increases. Rural areas maintain the highest counts, while urban areas show the lowest counts (housing_densityurban: 0.45). However, the presence of animals significantly influences these trends. Squirrels consistently have the strongest positive impact (1.21), particularly in urban areas (1.23), where they partially mitigate the negative effects of housing density. Cats also contribute positively (1.10), while dogs show a negative effect (0.92), likely due to disturbance near feeders.

Interactions between housing density and animal presence reveal nuanced effects. For example, in suburban areas, the presence of squirrels (1.15) provides a small boost despite the overall decline in bird counts caused by density (0.59). Similarly, cats in suburban and urban areas further increase counts through their respective interaction terms.

In summary, bird counts are influenced by a combination of time, housing density, and animal presence. While higher housing density reduces counts, the presence of squirrels and cats consistently increases them, with squirrels showing the most significant positive impact. Interactions further demonstrate how animals mitigate declines in bird counts, particularly in suburban and urban areas, underscoring the importance of species coexistence in supporting bird populations.