# Alligator Food Choice

Florida Alligator Food Choice - Zihan Wang, Yiquan Xiao, Yusen Wang

## 1 Introduction

## 1.1 Background and Literature Review

Understanding the dietary habits of alligators is pivotal for ecological balance and conservation strategies. Alligators, as apex predators, play an essential role in their ecosystems, influencing prey populations and nutrient distribution. Their diet, influenced by a range of factors including habitat, biological traits, and prey availability, provides insights into their ecological roles and adaptability to environmental changes. This project aims to investigate the dietary patterns of alligators across different Florida lakes, exploring how factors such as location, gender, and size impact their primary food choices.

The motivation for this study arises from the necessity to enhance our understanding of alligator behavior and its implications for ecosystem health. Knowledge of how various factors affect alligator food choices can inform conservation efforts, guide the management of alligator populations, and help predict the effects of environmental shifts on these key predators.

The dietary patterns and conditions of American alligators (Alligator mississippiensis) have been the focus of extensive research due to their pivotal role in maintaining the ecological balance within their habitats. Studies by Delany et al. [2] and Delany and Abercrombie [1] have laid the groundwork in understanding the complex interplay between alligator size, gender, and their preferred prey, unveiling how these factors dictate the diet composition across various Floridian lakes. These foundational works not only highlight the dietary shift from invertebrates to larger vertebrates as alligators grow but also emphasize the significance of habitat characteristics in influencing these patterns.

Further contributions by Rice [4] provide a closer examination of the dietary preferences and physical conditions of adult alligators in central Florida, identifying a strong preference for fish and noting considerable variances in dietary habits across different lake environments. This research underscores the importance of prey availability and habitat diversity in shaping the feeding behavior and overall health of alligators. On the other hand, the study by Platt et al. [3] focuses on the lesser-studied juvenile stage of alligators, exploring how the unique estuarine ecosystems of the Upper Lake Pontchartrain influence the diet of young alligators. Their findings reveal a reliance on crustaceans and small fish, shedding light on the early dietary adaptations that enable juvenile alligators to thrive in their specific habitats.

## 1.2 Data Description

The study of alligator food choices involves several variables classified into independent and dependent categories, with distinctions between continuous, ordinal, and nominal types. Below is a detailed description of these variables:

### 1.2.1 Independent Variables

- **Lake of Capture (L):** This nominal variable identifies the lake where each alligator was captured. The categories are as follows: *Hancock, Oklawaha, Trafford, George*
- **Gender (G):** Another nominal variable indicating the gender of the alligator. It is a binary variable and has two categories: *Male, Female*.
- **Size (S):** This is a binary ordinal variable that classifies alligators based on their length (in meters). The categories are: $\leq 2.3$, $> 2.3$

### 1.2.2 Dependent Variable

- **Primary Food Type (F)**: The primary food type, in volume, found in an alligator's stomach is a nominal polytomous dependent variable with five categories: *Fish, Invertebrate, Reptile, Bird, Other*

## 1.3 Research Questions

Our study aims to explore the dietary habits of alligators, focusing on the primary food type found in their stomachs. We classify these alligators based on the lake of capture, gender, and size. The research questions are structured to unravel the complex interactions between these variables and their impact on the alligator's primary food choice. To be specific, we are interested in:

- Investigate the associations between the primary food type of alligators and each independent variables (lake of capture, gender, and size) with conditioning on the other remaining independent variables.
- Explore the associations between the independent variables (lake of capture, gender, and size) themselves
- Analyze how the combination of lake of capture, gender, and size, along with the potential interactions between these factors, affects the primary food choice of alligators.

## 2. Preliminary Data Analysis

### 2.1 Marginal Distribution of Primary Food Choice

| Size | | <= 2.3 | > 2.3 |
|---|---|---|---|
| Lake | Gender | | |
| Hancock | Male | L1 | L9 |
| | Female | L5 | L13 |
| Oklawaha | Male | L2 | L10 |
| | Female | L6 | L14 |
| Trafford | Male | L3 | L11 |
| | Female | L7 | L15 |
| George | Male | L4 | L12 |
| | Female | L8 | L16 |

*Table 1: Li Definition*

In this section, we will explore the marginal distribution of the primary food choice according to the conditions of independent variables. To do that, we define L1 to be a variable that counts the number of specific conditions (Lake = Hancock, Gender = Female, Size $\leq$ 2.3 meters) met by each alligator. Then, we can construct 5 one-way frequency tables depending on the categories of primary food choice. Using the same way, we can define L2, …, L16 and construct corresponding one-way frequency tables (shown in Table 1). Goodness-of-fit (GOF) tests require that the expected frequency in each category be sufficiently large. Here, only L9 has no 0 count in 5 of its corresponding one-way frequency tables. Therefore, we will only focus on L9 in order to have appropriate GOF tests. The barplots for each one-way table of L9 is shown in Figure 1.

| | | L9 tables (for each Primary Food Type) | | | | |
|---|---|---|---|---|---|---|
| Distribution | $G^2$ test | Fish | Invertebrate | Reptile | Bird | Other |
| Binomial | $G^2/df$ | 7.22 | 6.64 | 5.3 | 2.82 | 2.21 |
| | p-value | 7.35E-04 | 1.31E-03 | 5.00E-03 | 5.96E-02 | 1.10E-01 |
| Poisson | $G^2/df$ | 18.61 | 18.20 | 11.44 | 9.76 | 3.02 |
| | p-value | 8.26E-09 | 1.25E-08 | 1.07E-05 | 5.76E-05 | 4.86E-02 |
| nBinomial | $G^2/df$ | 37.4 | 36.58 | 22.98 | 19.62 | 6.13 |
| | p-value | 9.65E-10 | 1.47E-09 | 1.64E-06 | 9.43E-06 | 1.33E-02 |

*Table 2: Likelihood Ratio Test for L9*

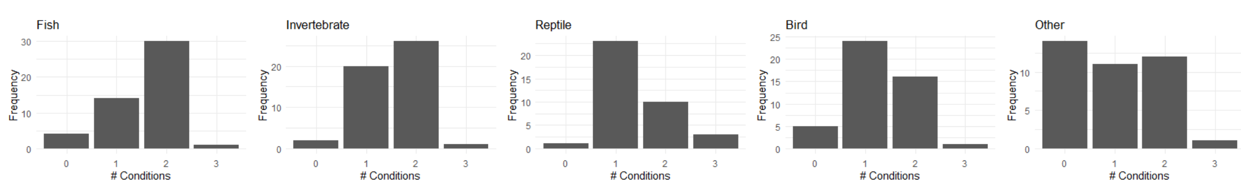


*Figure 1: Barplots for each one-way frequency tables of L9*

The p-values of GOF tests (shown in Table 2) for fish/invertebrate/reptile L9 data are far less than 0.05, suggesting that indicating that none of the tested distributions (Binomial, Poisson, Negative Binomial) suitably models the observed data. In contrast, for bird/other L9 data, p-values of GOF tests for Binomial distribution are larger than 0.05, indicating that Binomial distribution is a reasonable fit for these observed data.

The conclusions drawn from hanging rootograms and binomialness plots are the same as those drawn from the GOF test: In the hanging rootogram for "Other" and "Bird" L9 data, the bars do not deviate significantly from the expected values, as indicated by the close proximity of the base line to the end of the bars. Also, in the binomialness plot for “Other” and “Bird” L9 data, the red line in the binomialness plot is within the confidence intervals, suggesting that the logit of the binomial probabilities changes linearly with the number of specific conditions met by alligators. However, for “Fish”, “Invertebrate”, and “Reptile” L9 data, we can see blue and red bars deviate a lot from the expected values in the hanging rootogram. And in their binomialness plots, we can observe that the points and their confidence interval are far from the red line.

## 2.2 Stratified Analysis

In this section, we will do the stratified analysis to our data. Controlling 2 background variables makes our data sparse, so we decide to control 1 background variable each time. The corresponding association plots are shown as Figure 2, 3, 4.

### 2.2.1 Association between Food Type and Lake, stratified by Gender

| | Measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | ContCoef | ContCoefLB | ContCoefUB | CramerV | CramerVLB | CramerVUB | Lambda | LambdaLB | LambdaUB |
| Male | 0.488 | 0.449 | 0.610 | 0.323 | 0.290 | 0.444 | 0.173 | 0.097 | 0.321 |
| Female | 0.489 | 0.429 | 0.612 | 0.324 | 0.274 | 0.447 | 0.237 | 0.138 | 0.356 |

*Table 3: Association Measures (Lake by Gender)*

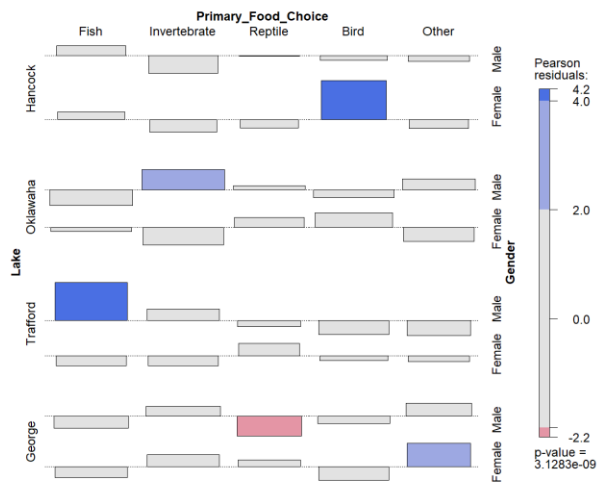


*Figure 2: Association Plot (Lake, Primary Food Choice, Gender)*

The Pearson's Chi-squared $X^2$ test, Likelihood Ratio $G^2$ test, and general test from CMHtest indicate a significant association between lake of capture and primary food type among male/female alligators, with p-values ($X^2$: 4.04e-4/8.60e-4, $G^2$: 2.90e-5/5.65e-4, general: 4.54e-4/9.63e-4) far below the typical alpha level of 0.05. For both males and females, the confidence interval for each association measure (shown in Table 3) does not include 0, which led to the same conclusion.

For male alligators, Contingency Coefficient and Cramer's V both indicate a moderate association between the lake of capture and primary food type. Goodman Kruskal’s Lambda suggests that knowing the lake of capture can improve the prediction of the primary food type by about 17.33%, which, while modest, is not small. For female alligators, the measures of association are very similar to those for males, indicating a moderate association between lake and primary food type among females. Goodman Kruskal’s Lambda for females is higher (23.68%) than for males, suggesting that lake of capture is a slightly better predictor of primary food type for female alligators than for male alligators.

From the association plot (shown in Figure 2), we can see that the p-value (3.13e-9) is much less than 0.05, which indicates that the overall association between lake and primary food type, stratified by gender, is statistically significant. The blue tiles suggest that male alligators from Trafford/Oklawaha are more likely to consume fish/invertebrate than expected, and female alligators from Hancock/George are more likely to consume bird/other than expected. While the red tiles suggest that male alligators from George are less likely to consume reptile than expected.

### 2.2.2 Association between Food Type and Lake, stratified by Size

| | Measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Size | ContCoef | ContCoefLB | ContCoefUB | CramerV | CramerVLB | CramerVUB | Lambda | LambdaLB | LambdaUB |
| $\leq 2.3$ | 0.550 | 0.498 | 0.674 | 0.381 | 0.331 | 0.527 | 0.189 | 0.104 | 0.386 |
| $> 2.3$ | 0.594 | 0.531 | 0.680 | 0.427 | 0.362 | 0.536 | 0.303 | 0.186 | 0.418 |

*Table 4: Association Measures (Lake by Size)*

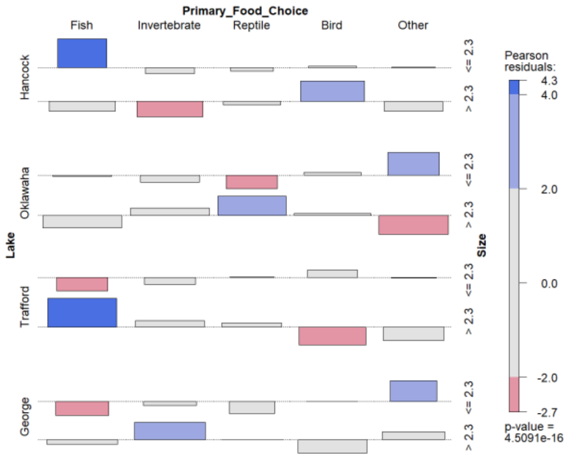

*Figure 3: Association Plot (Lake, Primary Food Choice, Size)*

$X^2$ test, $G^2$ test, and general test from CMHtest indicate a significant association between lake of capture and primary food type among small ($\leq 2.3$) / large ($> 2.3$) alligators, with p-values ($X^2$: 6.98e-4/1.52e-11, $G^2$: 1.21e-5/9.44e-12, general: 8.17e-4/1.93e-11) far below the typical alpha level of 0.05. For both small and large alligators, the CI for each association measure (shown in Table 4) does not include 0, which led to the same conclusion.
For small alligators, the Contingency Coefficient is 0.55 and Cramer's V is 0.381, indicating a moderate association between lake of capture and primary food type. Goodman Kruskal's Lambda is 0.1887, suggesting that knowing the lake of capture provides a moderate predictive power over the primary food type. The association measures for large alligator are larger than those for small alligators, indicating a stronger association than in smaller alligators.

From the association plot (shown in Figure 3), we can see that the p-value (4.51e-16) is much less than 0.05, which indicates that the overall association between lake and primary food type, stratified by size, is statistically significant. The blue tiles suggest that small alligators from Hancock/Oklawaha/George are more likely to consume fish/other/other than expected, and large alligators from Hancock/Oklawaha/Trafford/George are more likely to consume bird/reptile/fish/invertebrate than expected. While the red tiles suggest that small alligators from Oklawaha/Trafford/George are less likely to consume reptiles/fish/fish than expected, and large alligators from Hancock/Oklawaha/Trafford are less likely to consume invertebrate/other/birds than expected.

### 2.2.3 Association between Food Type and Gender, stratified by Lake

| | Measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Lake | ContCoef | ContCoefLB | ContCoefUB | CramerV | CramerVLB | CramerVUB | Lambda | LambdaLB | LambdaUB |
| H | 0.334 | 0.216 | 0.546 | 0.355 | 0.221 | 0.652 | 0.200 | 0.000 | 0.550 |
| O | 0.437 | 0.319 | 0.572 | 0.486 | 0.337 | 0.697 | 0.345 | 0.115 | 0.621 |
| T | 0.456 | 0.329 | 0.611 | 0.513 | 0.348 | 0.772 | 0.400 | 0.111 | 0.696 |
| G | 0.323 | 0.250 | 0.490 | 0.342 | 0.258 | 0.563 | 0.083 | 0.000 | 0.423 |

*Table 5: Association Measures (Gender by Lake)*

$X^2$ test, $G^2$ test, and general test from CMHtest indicate that there is a significant association between gender and primary food type among Oklawaha/Trafford alligators (p-values: $X^2$: 3.23e-3/9.43e-3, $G^2$: 1.83e-3/7.67e-3, general: 3.59e-3/1.06e-2) but there is no association between them among Hancock/George alligators (p-values: $X^2$: 0.206/0.178, $G^2$: 0.190/0.073, general: 0.216/0.186). The CI for association measures (shown in Table 5) leads to the same conclusion: For Oklawaha/Trafford alligators, the CI for each association measure does not include 0. For Hancock/George alligators, the CI for Goodman Kruskal's Lambda includes 0. Notice that for Hancock/George alligators, the CI for Contingency Coefficient and Cramer's V does not include 0, which seems to contradict with the conclusion drawn from Goodman Kruskal's Lambda. However, considering that Goodman Kruskal's Lambda treat one variable as predictor and the other as response variable while Contingency Coefficient and Cramer's V are symmetric, we believe Goodman Kruskal's Lambda is more appropriate here. We will also see similar things happen in later section.

For Oklawaha/Trafford alligators, the high Contingency Coefficient and Cramer's V indicate a strong association between gender and primary food type. High Goodman Kruskal's Lambda value suggests that knowing the gender provides a strong predictive power over the primary food type.

### 2.2.4 Association between Food Type and Gender, stratified by Size

| | Measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Size | ContCoef | ContCoefLB | ContCoefUB | CramerV | CramerVLB | CramerVUB | Lambda | LambdaLB | LambdaUB |
| $\leq 2.3$ | 0.222 | 0.131 | 0.448 | 0.228 | 0.132 | 0.501 | 0.139 | 0.000 | 0.412 |
| $> 2.3$ | 0.474 | 0.394 | 0.552 | 0.538 | 0.429 | 0.661 | 0.486 | 0.276 | 0.600 |

*Table 6: Association Measures (Gender by Size)*

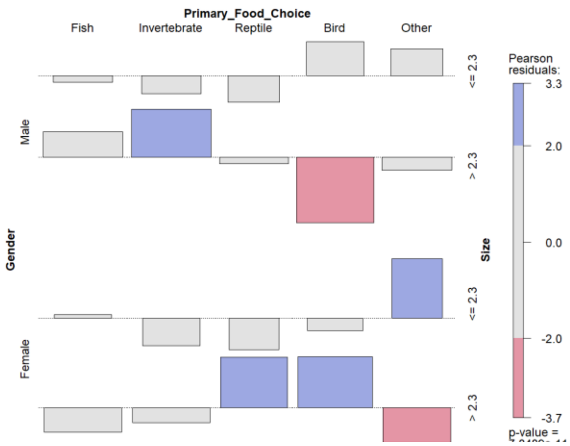


*Figure 4: Association Plot (Gender, Primary Food Choice, Size)*

$X^2$ test, $G^2$ test, and general test from CMHtest indicate that there is a significant association between gender and primary food type among large alligators (p-values: $X^2$: 2.85e-8, $G^2$: 1.46e-9, general: 3.27e-8) but there is no association between them among small alligators (p-values: $X^2$: 0.401, $G^2$: 0.391, general: 0.408). The confidence intervals for association measures lead to the same conclusion: For large alligators, the CI (shown in Table 6) for each association measure does not include 0. For small alligators, the CI for Goodman Kruskal's Lambda includes 0. For large alligators, the high Contingency Coefficient and Cramer's V indicate a moderate association between gender and primary food type. High Goodman Kruskal's Lambda value suggests that knowing the gender provides a moderate predictive power over the primary food type.

From the association plot (shown in Figure 4), we can see that the p-value (7.85e-11) is much less than 0.05, which indicates that the overall association between gender and primary food type, stratified by size, is statistically significant. The blue tiles suggest that large male alligators are more likely to consume invertebrate than expected, and small/large female alligators are more likely to consume other/(reptile&bird) than expected. While the red tiles suggest that large male/female alligators are less likely to consume bird/other than expected.

### 2.2.5 Association between Food Type and Size, stratified by Lake

| | Measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Lake | ContCoef | ContCoefLB | ContCoefUB | CramerV | CramerVLB | CramerVUB | Lambda | LambdaLB | LambdaUB |
| H | 0.453 | 0.314 | 0.611 | 0.508 | 0.331 | 0.772 | 0.478 | 0.158 | 0.727 |
| O | 0.563 | 0.502 | 0.634 | 0.681 | 0.581 | 0.821 | 0.480 | 0.304 | 0.720 |
| T | 0.598 | 0.526 | 0.673 | 0.745 | 0.618 | 0.910 | 0.571 | 0.285 | 0.846 |
| G | 0.454 | 0.329 | 0.586 | 0.510 | 0.349 | 0.723 | 0.125 | 0.000 | 0.563 |

*Table 7: Association Measures (Sizer by Lake)*

$X^2$ test, $G^2$ test, general and rmeans tests from CMHtest indicate that there is a significant association between size and primary food type among alligators captured from Hancock/Oklawaha/Trafford/George, with p-values ($X^2$: 1.63e-2/3.00e-6/1.07e-5/7.19e-3, $G^2$: 1.20e-2/5.38e-8/1.57e-6/2.10e-3, general: 1.83e-2/3.72e-6/1.39e-05/8.06e-3, rmeans: 7.45e-3/8.74e-3/1.62e-06/1.98e-3) far below the typical alpha level of 0.05. For alligators from all lakes, the CI for each association measure (shown in Table 7) does not include 0, which led to the same conclusion, except for Lake G that it lower bound already reach zero, contradict to the conclusion.

For alligators from all lakes, the high Contingency Coefficient and Cramer's V indicate a strong association between gender and primary food type. For alligators from Hancock/Oklawaha/Trafford, high Goodman Kruskal's Lambda value suggests that knowing the gender provides a strong predictive power over the primary food type. However, for alligators from George, the Goodman Kruskal's Lambda is much lower even with zero, which suggests that size is a much poorer predictor of primary food type for George alligators than it is for Hancock/Oklawaha/Trafford alligators.

### 2.2.6 Association between Food Type and Size, stratified by Gender

| | Measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | ContCoef | ContCoefLB | ContCoefUB | CramerV | CramerVLB | CramerVUB | Lambda | LambdaLB | LambdaUB |
| Male | 0.460 | 0.367 | 0.556 | 0.518 | 0.395 | 0.670 | 0.333 | 0.195 | 0.544 |
| Female | 0.492 | 0.401 | 0.581 | 0.566 | 0.438 | 0.714 | 0.361 | 0.200 | 0.575 |

*Table 8: Association Measures (Size by Gender)*

$X^2$ test, $G^2$ test, general and rmeans tests from CMHtest indicate that there is a significant association between size and primary food type among male/female alligators, with p-values ($X^2$: 4.18e-6/7.74e-7, $G^2$: 1.85e-6/4.23e-7, general: 4.74e-6/3.72e-6, rmeans: 7.45e-3/2.76e-4) far below the typical alpha level of 0.05. For both male and female alligators, the CI for each association measure (shown in Table 8) does not include 0, which led to the same conclusion.
For both male and female alligators, the high Contingency Coefficient and Cramer's V indicate a moderate association between size and primary food type. The Goodman Kruskal's Lambda value suggests that knowing the size provides a moderate predictive power over the primary food type.

## 2.3 Doubledecker Plot

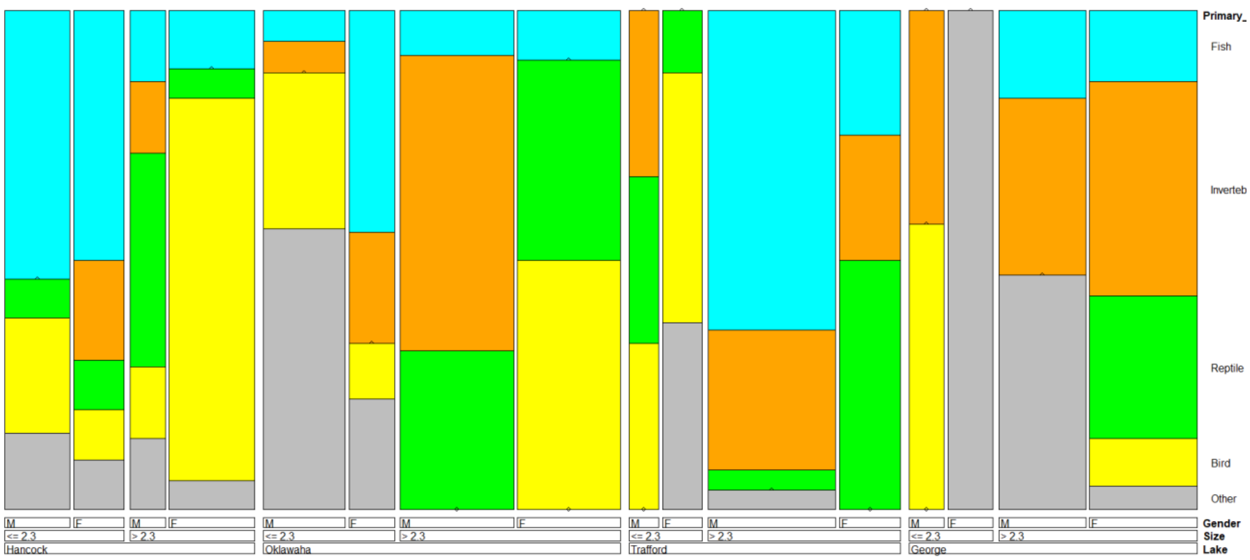


*Figure 5: Doubeldecker Plot, Fish (Cyan), Invertebrate (Orange), Reptile (Grean), Bird (Yellow), Other (Grey)*

The Doubledecker plot (shown in Figure 5) clearly shows that the primary food choices of alligators significantly vary depending on the lake in which they reside. In Lake Hancock, alligators have a more varied diet across different types. In contrast, in the other three lakes, not all food types are represented within certain categories. In Lake Oklawaha, the food preferences of alligators significantly vary by size. Larger alligators primarily consume invertebrates, reptiles, and birds, while smaller alligators, particularly males, show a strong preference for other food sources. In Lake Trafford, there is also a notable difference in food choices based on the size of the alligator. Smaller alligators tend to prefer food sources such as birds and others, while larger alligators favor invertebrates and fish. Additionally, in this lake, the difference in food preferences between male and female alligators is more pronounced, regardless of size. In Lake George, there is a significant lack of detailed data on the food choices of small-sized alligators. However, the data shows that large-sized alligators exhibit a much greater variation in their diet compared to those in other lakes.

We also analyzed the doubledecker plot for Lake, Size, and Gender separately. From the marginal categories, we observed that the distribution of food choices among alligators in Lake Oklawaha is the closest to uniform compared to other lakes. In Lake Hancock, birds and fish are the primary food sources. In Lake Trafford, fish dominate the diet, while in Lake George, other types of food are most prevalent. Focusing solely on the marginal data for gender, female alligators show a preference for birds and other types of food sources, whereas male alligators favor fish and invertebrates more. Upon examining the marginal data for size, as previously noted, large-sized alligators tend to have a more diverse range of food choices, while small-sized alligators show a marked preference for birds and other types of food.

# 3 Modeling

## 3.1 Log-Linear Models

We started with the baseline log-linear model [F][LGS]. The small p-value of the goodness-of-fit test (< 2.2e-16) and the blue/red tiles shown in the corresponding mosaic display indicate that this model does not fit the data well. From the mosaic display, we see that the primary food type of alligators varies among lakes, suggesting that we need to consider the association between Primary Food Type and Lake.

Then we fit the log-linear model [FL][LGS]. The small p-value of the GOF test and unclean mosaic display show that this model does not fit the data well. From the mosaic display, we see that the primary food type of alligators varies between size of alligators. Since we cannot see obvious pattern from it, we try [FL][FS][LGS] and [FLS][LGS]. Both of them have a small -value for the GOF test and an unclean mosaic display, which means that none of them are acceptable models. The deviance test of these two models has p-value = 8.346e-10 < 0.05, indicating that the more complex model, [FLS][LGS], is more suitable for the data.

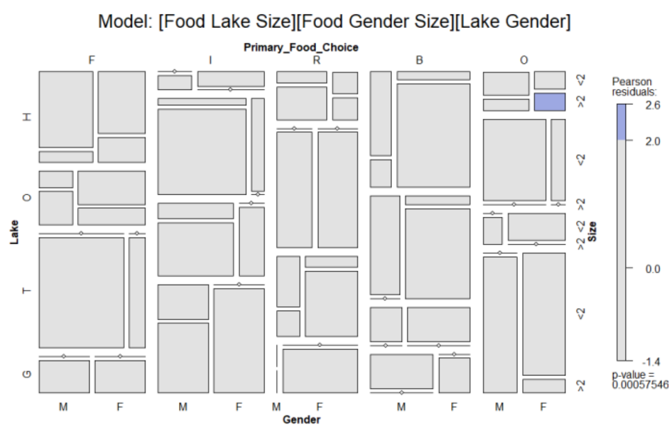


*Figure 6: Mosaic display for [FLS][FGS][LG]*

From the mosaic display of [FLS][LGS], we see that the primary food type of alligators varies between gender of alligators. We still hard to see obvious pattern from it, so we try [FLS][FG][LGS], [FLS][FLG][LGS], and [FLS][FGS][LGS] to see which one best describe the data. All of them have a small -value for the GOF test and an unclean mosaic display, which means that none of them are acceptable models. The deviance test of [FLS][LGS] and [FLS][FG][LGS] has p-value = 0.0139 < 0.05, confirming that the association between Primary Food Type and Gender should be considered. And the deviance tests between [FLS][FG][LGS], [FLS][FLG][LGS], and [FLS][FGS][LGS] indicate that [FLS][FGS][LGS] is the most appropriate log-linear model among them.

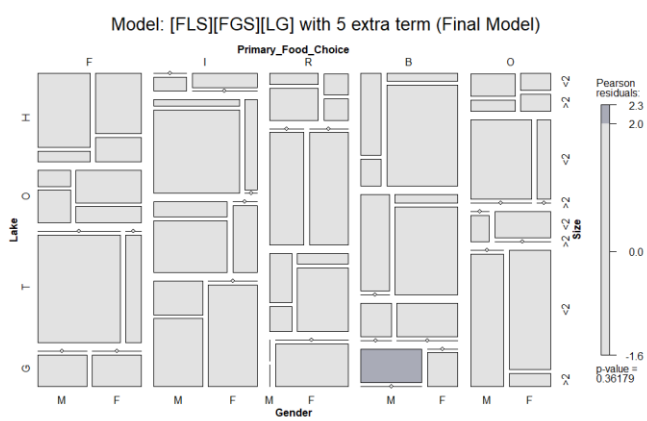


*Figure 7: Mosaic display for [FLS][FGS][LG] + FO_LH_GF_SG + FI_LH_GF_SL + FF_LH_GF_SG + FI_LO_GF_SL + FF_LO_GF_SG*

Next, we want to see if the association between the predictor variables depends on the response variable. To do that, we compare [FLS][FGS], [FLS][FGS][LG], and [FLS][FGS][LGS]. The deviance tests show that [FLS][FGS][LG] is the best among them. There is a blue tile (Primary Food Choice = "Other", Lake = "Hancock", Gender = "Female", Size = "> 2.3") in the mosaic display of this model (shown in Figure 6), so we add an indicator (denoted as FO_LH_GF_SG) and fit the model [FLS][FGS][LG] + FO_LH_GF_SG to capture corresponding association. The p-value of the GOF test for it is 0.0004 < 0.05, indicating that this model is still inappropriate for the data. From its mosaic display, we see another two blue tiles. So, we add two indicators

FI_LH_GF_SL and FF_LH_GF_SG, and then fit the model [FLS][FGS][LG] + FO_LH_GF_SG + FI_LH_GF_SL + FF_LH_GF_SG. The GOF test for this model suggests that it is still inappropriate (p-value = 0.01). We add two new indicators FI_LO_GF_SL and FF_LO_GF_SG to the data, each corresponding to a newly appeared blue tile in the mosaic display. Finally, we fit the model [FLS][FGS][LG] + FO_LH_GF_SG + FI_LH_GF_SL + FF_LH_GF_SG + FI_LO_GF_SL + FF_LO_GF_SG. The large p-value of the GOF test (0.3214) and the clean mosaic display (shown in Figure 7) suggest that this model is a good fit for our data.

## 3.2 Logistic Regression Models

Since the dependent variable, Primary Food Choice, has 5 levels, we will fit 5 separate logistic regression models, each for one of primary food types. Also, as predictors are all categorical, the relationship between the dependent variable and one of the independent variables is linear when other independent variables are held constant.

### 3.2.1 Fish as Response

We first fit the main effect model Fish ~ Lake + Gender + Size and compare it with the model that allows all interaction terms (model ①). The result deviance test rejects the null hypothesis with p-value = 2.401e-05, which means that we should consider the interaction between predictors. The large p-value from model ① GOF test (0.9202) suggests that model ① fit the data well. The hypothesis tests on each term of this complex model show that only Lake and the two-way interaction term Lake:Size are highly significant. So we fit Fish ~ Lake + Gender + Size + Lake:Size (model ②). The large p-value from model ② GOF test (0.8879) suggests that model ② also fit the data well. Finally, we compare model ① and model ② with deviance test, which shows that the simple model (model ②) is suitable for the data (p-value = 0.1833).

We can then interpret the estimated coefficients in terms of odds ratio based on this model. For example, the estimated coefficient for the term "LakeOklawaha", "Size> 2.3", and "LakeOklawaha:Size> 2.3" are -1.493, -1.978, and 1.136 respectively, which means that Oklawaha alligators with size greater than 2.3 meters are exp(-1.493-1.978+1.136) = 0.096 times more likely to have fish as their primary food choice compared to those Hancock alligators with size less or equal to 2.3 meters. Other estimated coefficients can be interpreted using the same way.

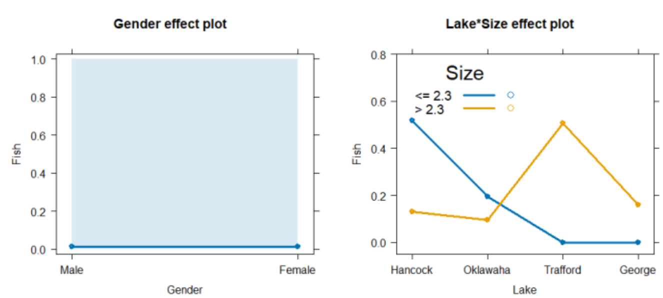


*Figure 8: effect plots for model Fish ~ Lake + Gender + Size + Lake:Size*

The effect plots of model ② is shown in Figure 8. From the Gender effect plot, we can see that male and female alligators are equally likely to have fish as their primary food choice, which suggests that Gender might not have a significant impact on Fish. From the Lake:Size effect plot, we can see that the curves cross each other, which suggests that we do need to consider the two-way interaction between Lake and Size.

### 3.2.2 Invertebrate as Response

With a similar approach, we first fit the main effect model Invertebrate ~ Lake + Gender + Size and compare it with the model that allows all interaction terms (model ①). The result deviance test rejects the null hypothesis with p-value = 1.131e-4, which means that we should consider the interaction between predictors. The large p-value from model ① GOF test (0.9202) suggests that model ① fit the data well. The hypothesis tests on each term of this complex model show that the three-way interaction term Lake:Gender:Size are highly significant. So we fit Invertebrate ~ Lake + Gender + Size + Lake:Gender:Size (model ②). However, there are many NA and super large/small values for estimated coefficients in model ②. This may be because some level combinations are rare or unobserved, leading directly to NAs for those coefficients due to lack of data. So, we will not look at model ②.

For example, the estimated coefficient for the term "LakeGeorge", "GenderFemale", "Size> 2.3", "LakeGeorge: GenderFemale", "LakeGeorge:Size> 2.3", "GenderFemale:Size> 2.3", and "LakeGeorge:GenderFemale:Size> 2.3" are

19.278, 18.18, 17.774, -37.458, -18.093, -35.954 and 55.551 respectively, which means that George female alligators with size greater than 2.3 meters are exp(19.278 + 18.18 + 17.774 - 37.458 - 18.093 - 35.954 + 55.551) = exp(19.278) times more likely to have invertebrate as their primary food choice compared to those Hancock male alligators with size less or equal to 2.3 meters.

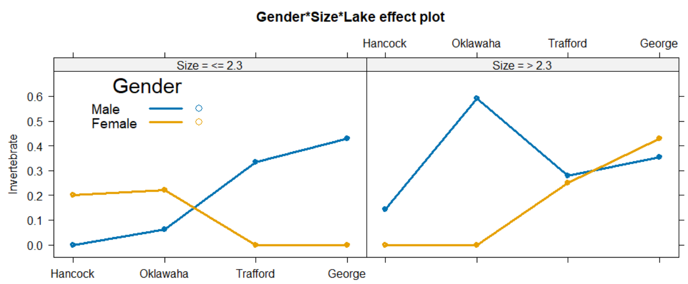


Figure 9: effect plots for model Invertebrate ~ Lake + Gender + Size + Lake:Gender:Size

The effect plots of model ① are shown in Figure 9. From the effect plot, we can see signs suggesting that all three factors would affect the alligators' interest in eating invertebrates as a primary food choice. No single factor would result in a general pattern. This means using the three-way interaction model is significant.

### 3.2.3 Reptile as Response

With a similar approach, we first fit the main effect model Reptile ~ Lake + Gender + Size and compare it with the model that allows all interaction terms (model ①). The result deviance test rejects the null hypothesis with p-value = 5.493e-4, which means that we should consider the interaction between predictors. The large p-value from model ① GOF test (0.9983) suggests that model ① fit the data well. The hypothesis tests on each term of this complex model show that 2 two-way interaction terms Lake:Gender and Lake:Size are highly significant. So, we fit Reptile ~ Lake * Gender + Lake * Size (model ②). The large p-value from model ② GOF test (0.9955) suggests that model ② also fit the data well. Finally, we compare model ① and model ② with deviance test, which shows that the simple model (model ②) is suitable for the data (p-value = 0.06653).

We can then interpret the estimated coefficients in terms of odds ratio based on this model. For example, the estimated coefficient for the term "LakeOklawaha", "GenderFemale", "Size> 2.3", "LakeOklawaha:GenderFemale", "LakeOklawaha:Size> 2.3" are -17.79665, -1.50691, 1.20429, 1.86358, 17.73709 respectively, which means that given Lake Oklawaha, female alligators with size greater than 2.3 meters are exp(-17.79665 -1.50691 + 1.20429 + 1.86358 + 17.73709) = 4.48797 times more likely to have reptiles as their primary food choice compared to those Hancock male alligators with size less or equal to 2.3 meters. Other estimated coefficients can be interpreted using the same way.

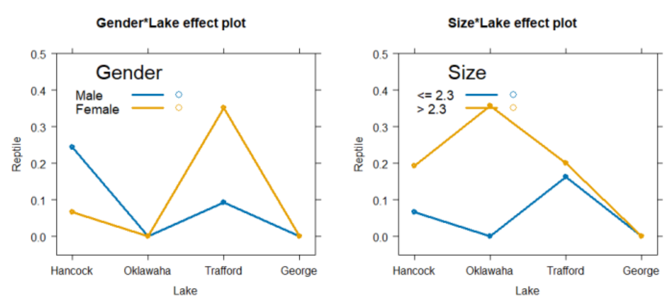


Figure 10: effect plots for model Reptile ~ Lake * Gender + Lake * Size

The effect plots of model ② are shown in Figure 10. The effect plots of model ② are shown in Figure 10. From the effect plot, we can see that there is a cross on Gender:Lake plot, which means that there is a interaction term between Lake and Gender. What's more, based on Lake:Size effect plot, what we could know is that larger size are more likely to choose reptile as the primary food choice since what ever the lake is, the larger size always has a higher probability choosing reptile. Still, we could notice that there may should exist an interaction term between Lake and Size since effect plot shows an same probabilty on Lake George.

### 3.2.4 Birds as Response

With a similar approach, we first fit the main effect model Bird ~ Lake + Gender + Size and compare it with the model that allows all interaction terms (model ①). The result deviance test rejects the null hypothesis with p-value = 1.625e-09, which means that we should consider the interaction between predictors. The large p-value from model ① GOF test (0.9997) suggests that model ① fit the data well. The hypothesis tests on each term of this complex model show that 2 two-way interaction terms Lake:Size and Gender:Size are highly significant. So, we fit Bird ~ Lake

* Size + Gender * Size (model ②). The large p-value from model ② GOF test (0.9994) suggests that model ② also fit the data well. Finally, we compare model ① and model ② with deviance test, which shows that the simple model (model ②) is suitable for the data (p-value = 0.2048).

We can then interpret the estimated coefficients in terms of odds ratio based on this model. For example, the estimated coefficient for the term "LakeOklawaha", "GenderFemale", "Size> 2.3", "LakeOklawaha:Size> 2.3" and "GenderFemale:Size>2.3" are 0.34457, -1.11330, -1.32353, -1.76825, and 4.94608 respectively, which means that given size greater than 2.3 meters, female alligators from Lake Oklawaha are exp(0.34457 - 1.11330 - 1.32353 - 1.76825 + 4.94608) = 2.96 times more likely to have birds as their primary food choice compared to those Hancock male alligators with size less or equal to 2.3 meters. Other estimated coefficients can be interpreted using the same way.

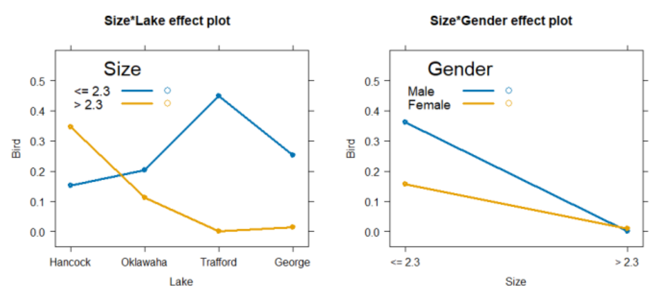


*Figure 10: effect plots for model Bird ~ Lake * Size + Gender * Size*

The effect plots of model ② are shown in Figure 11. From the effect plot, we can see that there is a cross on Lake:Size plot, which means that there is a interaction term between Lake and Size. What's more, based on Size:Gender plot, we could also find that male alligators are always more likely to choose bird as primary food choice than female. Still, we could also find that there is a cross on Size:Gender effect plot, which means that there should be an interaction term between them.

### 3.2.5 Others as Response

With a similar approach, we first fit the main effect model Other ~ Lake + Gender + Size and compare it with the model that allows all interaction terms (model ①). The result deviance test rejects the null hypothesis with p-value = 3.45e-07, which means that we should consider the interaction between predictors. The large p-value from model ① GOF test (1.0) suggests that model ① fit the data well. The hypothesis tests on each term of this complex model show that all interaction terms (both 2-way and 3-way) are highly significant. So we don't need to drop any terms.

We can then interpret the estimated coefficients in terms of odds ratio based on this model. For example, the estimated coefficient for the term "LakeGeorge", "GenderFemale", "Size> 2.3", "LakeGeorge: GenderFemale", "LakeGeorge:Size> 2.3", "GenderFemale:Size> 2.3", and "LakeGeorge:GenderFemale:Size> 2.3" are -17.861320, -0.492476, -0.087011, 39.624614, 19.535297, -0.488353, -41.521734 respectively, which means that given size greater than 2.3 meters, female alligators from Lake George are exp(-17.861320 - 0.492476 - 0.087011 + 39.624614 + 19.535297 - 0.488353 - 41.521734) = 0.275 times more likely to have others as their primary food choice compared to those Hancock male alligators with size less or equal to 2.3 meters. Other estimated coefficients can be interpreted using the same way.

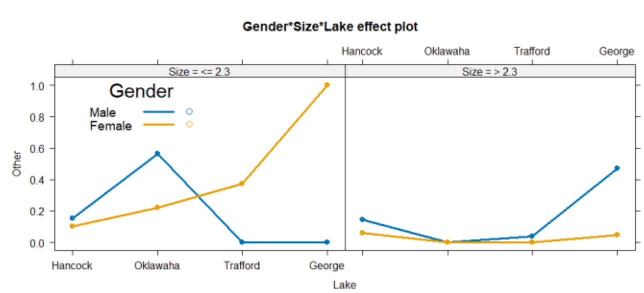


*Figure11: effect plots for model Other ~ Lake * Gender * Size*

The effect plots of model ② are shown in Figure 12. From the effect plot, we could find that for female, size <= 2.3 always has higher probability choosing other as the primary food choice than size > 2.3; for male, it changes on Trafford and George. If we consider the gender, we could find that male is more likely to choose other as primary food choice when size > 2.3 than female, it also varies at Trafford and George when size <= 2.3 that the female has a higher probability to choose other as the primary food choice. We could also find that there is a clear cross between male and female when size <= 2.3 which implies there should be multiple interaction terms between these three variables.

## 4 Peer Assessment

Zihan Wang is responsible for the code & analysis for section 3.1. Yiquan Xiao is responsible for the code & analysis for part of the section 3.2. Yusen Wang is responsible for remaining code & analysis for section 3.2 and writing of the report.

## References


[1] Michael F. Delany and C. L. Abercrombie. “American alligator food habits in northcentral Florida”. In: Journal of Wildlife Management 50 (1986), pp. 348–353. url: https://api.semanticscholar.org/CorpusID:87689453.

[2] Michael F. Delany, Stephen B. Linda, and Clinton T. Moore. “Diet and Condition of American Alligators in 4 Florida Lakes”. In: 1999. url: https://api.semanticscholar.org/CorpusID:87104107.

[3] Steven G. Platt, Christopher G. Brantley, and Robert W. Hastings. “Food Habits of Juvenile American Alligators in the Upper Lake Pontchartrain Estuary”. In: 1990. url: https://api.semanticscholar.org/CorpusID:54660192.

[4] Amanda Nicole Rice. “DIET AND CONDITION OF AMERICAN ALLIGATORS (Alligator mississippiensis) IN THREE CENTRAL FLORIDA LAKES”. In: 2004. url: https://api.semanticscholar.org/CorpusID:85680743.