Alligator Food Choice

Florida Alligator Food Choice - Yiquan Xiao, Zihan Wang, 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 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. And the motivation for this study arises from the necessity to enhance our understanding of alligator behavior and its implications for ecosystem health.

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. 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. 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. **Independent variables include Lake of Capture, Gender, and Size**. **Lake of Capture (L)** is a nominal variable that identifies the lake where each alligator was captured. It can be one of *Hancock*, *Oklawaha*, *Trafford*, and *George*. **Gender (G)** is another nominal variable indicating the gender of the alligator. It is a binary variable and has two categories: *Male*, *Female*. **Size (S)** is a binary ordinal variable that classifies alligators based on their length (in meters). The categories are: ≤ 2.3, > 2.3. **There is only one dependent variable Primary Food Type (F)**, which represents the primary food type, in volume, found in a n alligator's stomach. It is a nominal polytomous 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

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 L9 to be a variable that counts the number of specific conditions (Lake

= Hancock, Gender = Male, Size $>$ 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 L1, ⋯, L16 and construct corresponding one-way frequency tables. However, 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 p-values of GOF tests for fish/invertebrate/reptile L9 data are far less than 0.05, suggesting 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

The blue tiles from the association plot (F, L, G) 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.

The blue tiles from the association plot (F, L, S) 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.

The blue tiles from the association plot (F, G, S) 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.

Although the association measures across different background variables sometimes appear similar (we will see this in section 2.2.1 and section 2.2.6), the association plots from our analysis clearly reveal distinct patterns across these variables, underscoring the necessity of stratified analysis. Also, since controlling 2 background variables makes our data sparse, so we decide to control 1 background variable each time.

2.2.1 Association between Food Type and Lake, stratified by 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 does not include 0, which led to the same conclusion. For male alligators, Contingency Coefficient (0.488) and Cramer's V (0.323) 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 small, not negligible. 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.

2.2.2 Association between Food Type and Lake, stratified by 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 (≤ 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 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 small but not negligible 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.

2.2.3 Association between Food Type and Gender, stratified 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 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 Contingency Coefficient and Cramer’s V does not include 0. Notice that for Hancock/George alligators, the CI for Goodman Kruskal’s Lambda includes 0, which seems to contradict with the conclusion drawn from Goodman Kruskal’s Lambda. However, this can be due to the reason that Gender is indeed associated with Primary Food Type, but is not effective in predicting it when used alone. We will also see similar things in section 2.2.4 and 2.2.5. For Oklawaha/Trafford alligators, the relatively high Contingency Coefficient and Cramer's V indicate a moderate association between gender and primary food type. A relatively high Goodman Kruskal’s Lambda value suggests that knowing the gender provides a moderate predictive power over the primary food type.

2.2.4 Association between Food Type and Gender, stratified by 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 for each association measure does not include 0. For small alligators, the CI for Goodman Kruskal’s Lambda includes 0. But for the same reason mentioned in section 2.2.3, we believe this does not really contradict the previous conclusion. 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.

2.2.5 Association between Food Type and Size, stratified 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 of each association measure does not include 0, which led to the same conclusion, except for the CI of Goodman Kruskal’s Lambda for Lake George. But for the same reason mentioned in section 2.2.3, we believe this does not really contradict the previous conclusion. For alligators from all lakes, the high Contingency Coefficient and Cramer's V indicate a relatively 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 (0.125), 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

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 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

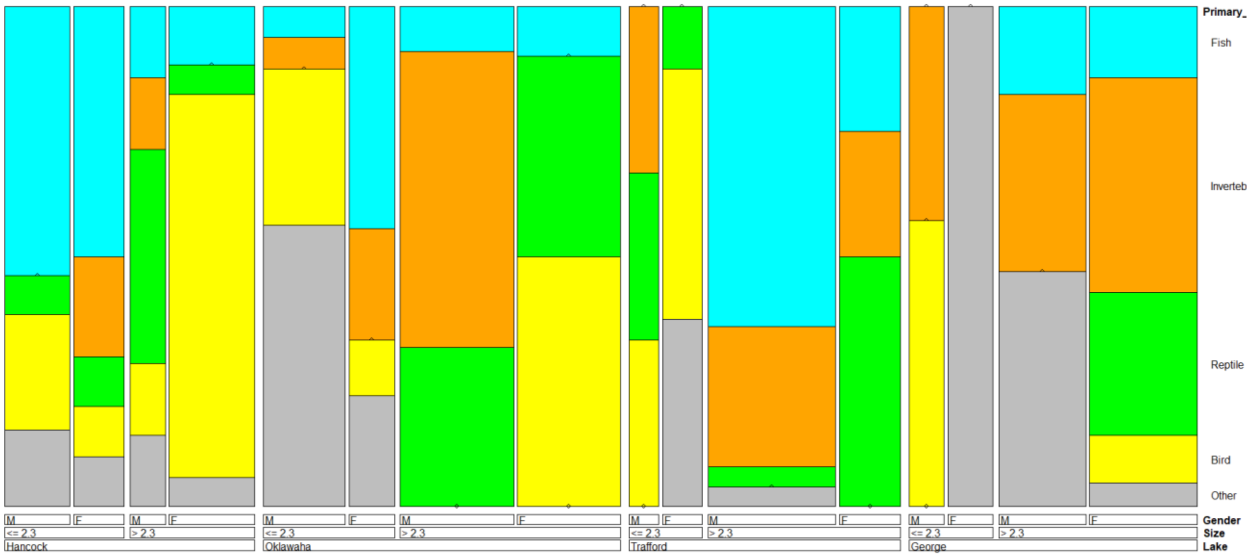

Figure 1: Doubeldecker Plot, Fish (Cyan), Invertebrate (Orange), Reptile (Grean), Bird (Yellow), Other (Grey)

The Doubledecker plot (shown in Figure 1) 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 genders and sizes. For males with a size smaller than 2.3 meters, they have a varied diet. For males with a larger size, they have a much stronger preference for reptiles than smaller males. For females with a size smaller than 2.3 meters, they have a similar varied diet while also consuming the invertebrate. Larger females, on the other hand, had a strong focus on birds as a food source. So in general, alligators in Hancock Lake have some gender differences in food choice but in general still diverse across males and females. Size, however, influences quite a lot, larger-sized ones, despite gender, would have a relatively more favorite toward some certain foods than smaller-sized alligators.

In contrast, in the other three lakes, not all food types are represented within certain categories. In Lake Ocklawaha, 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. Additionally, gender influences the preference quite significantly as well. Small-sized males' favorite food choices are other types and birds, whereas for small-sized females, they prefer fish the most. Large-sized males prefer invertebrates and reptiles, whereas large-sized females prefer reptiles and birds. 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. Size also leads to a great difference as well. It is not hard to see that small-sized alligators in this Lake don't consume fish at all, whereas large-sized ones, particularly males, have a strong preference for fish. It is also worth noting that small-sized alligators are caught much less proportionally comparing to large-sized ones. In Lake George, there is a significant lack of detailed data on the food choices of small-sized alligators. In this lake, the small-sized alligators had an even smaller proportion comparing to Lake Trafford. However, the data shows that large-sized alligators exhibit a much greater variation in their diet compared to those in other lakes. Here, for large-sized alligators, gender influences the preferences for food as well. Males don't consume reptiles and birds, whereas the female had a generally similar proportion for all listed food types.

2.4 Generalized Mosaic Matrices

In order to explore different types of association structures, we produced multiple mosaic matrices under mutual independence, joint independence, and conditional independence, respectively. For mosaic plots in these matrices, we found that only one of them is "clean" enough. Although this mosaic plot still contains several blue/red tiles, it is much cleaner than the others. This mosaic plot corresponds to ($L \perp G|F, S$), which implies that Lake and Gender are independent given all other variables.

3 Modeling

3.1 Log-Linear Model

Based on the result from section 2.4, we will start with the model [FLS][FGS]. The small p-value of the goodness-of-fit test (0.00068) indicates that this model does not fit the data well. Pattern of the blue/red tiles in the mosaic display still exist among variables whose association are not considered.

Current model [FLS][FGS] has already shown that there is association between response and each predictor. 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, 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 suggest that this model is a good fit for our data.

The final model suggests that there is an association between Lake and Gender, and this association does not depend on other variables. This also suggests that given Primary Food Type and Size, Lake and Gender will be independent of each other. However, the above associations do not hold in several cases represented by the indicators in the model. For example, FO_LH_GF_SG indicates that large-sized female alligators in Hancock are more likely to eat "Other" type of food.

3.2 Logistic Regression Models for Binary Response

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.

For each defined binary response variable, we will fit 3 different models: main effect model (M1), model that allow all 2-way interactions (M2), model that allow 3-way interaction (M3). Then we compare and choose the best model from them according to deviance test. For brevity, we will not repeat this process for each subsection. When interpreting models, the reference groups mentioned in this section refer to Hancock male alligators with size less or equal to 2.3 meters. Also, for brevity, the following abbreviations will be used when we write the model: *LakeOklawaha* is denoted by *LO*, *LakeTrafford* is denoted by *LT*, *LakeGeorge* is denoted by *LG*, *GenderFemale* is denoted by *GF*, *Size> 2.3* is denoted by *SG*. For interaction terms, colons will be used as a separator. For example, *LakeOklawaha:GenderFemale* will be denoted by *LO:GF*, and *LakeOklawaha:GenderFemale:Size> 2.3* will be denoted by *LO:GF:SG*, etc.

3.2.1 Fish as Response

The deviance test between M1 and M2 rejects null hypothesis with p-value 2.96e-06, suggesting that the more complex model (M2) is more suitable for the data. The deviance test between M2 and M3 accepts null hypothesis with p-value 0.778, suggesting that the simpler model (M2) is more suitable for the data. The significance test for its coefficients shows that among interaction terms, only the interaction between Lake and Size are highly significant. Therefore, we fit the model that only allows interaction between Lake and Size (M4), and compare it with M2. The deviance test between M2 and M4 accepts null hypothesis with p-value 0.061, suggesting that the simpler model (M4) is more suitable for the data. The large p-value (0.8879) of GOF test for M4 indicates that this model fits the data well. Therefore, M4 will be our final model, which is: logit(π) = 0.181 – 1.493 * *LO* - 18.627 * *LT* - 18.629 * *LG* - 1.978 * *SG* - 0.216 * *GF* + 1.136 * *LO:SG* + 20.548 * *LT:SG* + 18.867 * *LG:SG* where π = P(Fish = 1 | Lake, Gender, Size).

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

In the Gender effect plot of M4, the flat line indicates that male and female alligators are equally likely to have fish as their primary food choice. And in the Lake*Size effect plot, for alligators with size > 2.3 meters, we observe that such alligators in Trafford are more likely to consume Fish compared to alligators in Hancock, Oklawaha, and George. And Hancock, Oklawaha, and George alligators are equally likely to have fish as their primary food choice. However, for alligators with size <= 2.3 meters, Hancock alligators are more likely to consume fish than Oklawaha alligators, who is more likely than Trafford alligators, who in turn is more likely than George alligators.

After fitting our final model, we are interested in the outliers and potential influential cases. Based on diagnostic plots and corresponding identified outliers, we see that many cases have large studentized residuals, which means that they are outliers in terms of y (Fish). For example, some alligators are unusual in that they are large-sized Hancock female who have fish as their primary food choice. And some alligators are unusual in that they are large-sized Oklawaha male who does not have fish as their primary food choice. We also observe that none of the cases have large hat values and none of the cases have a large Cook's distance, suggesting that even though these cases are identified as outliers, they are not influential.

3.2.2 Invertebrate as Response

The deviance test between M1 and M2 accepts null hypothesis with p-value 0.3118, suggesting that the simpler model (M1) is more suitable for the data. The deviance test between M1 and M3 rejects null hypothesis with p-value 0.00011, suggesting that the more complex model (M3) is more suitable for the data. The large p-value (0.9333) of GOF test for M3 indicates that this model fits the data well. Therefore, M3 will be our final model, which is: logit(π) = - 19.566 + 16.858 * *LO* + 18.873 * *LT* + 19.278 * *LG* + 18.180 * *GF* + 17.774 * *SG* - 16.724 * *LO*:*GF* - 37.053 * *LT*:*GF* - 37.458 * *LG*:*GF* - 14.699 * *LO*:*SG* - 18.026 * *LT*:*SG* - 18.093 * *LG*:*SG* - 35.954 * *GF*:*SG* + 14.565 * *LO*:*GF*:*SG* + 54.673 * *LT*:*GF*:*SG* + 55.551 * *LG*:*GF*:*SG* where π = P(Invertebrate = 1 | Lake, Gender, Size).

We can then interpret the estimated coefficients similar to section 3.2.1. For example, the estimated coefficient for the term *LO*, *GF*, *SG*, *LO:GF*, *LO:SG*, *GF:SG*, *LO:GF:SG* are 16.858, 18.180, 17.774, -16.724, -14.699, -35.954, 14.565 respectively, which means that Oklawaha female alligators with size greater than 2.3 meters are exp(16.858 +

18.180 + 17.774 - 16.724 - 14.699 - 35.954 + 14.565) = 1 times as likely to have invertebrate as their primary food choice compared to the reference group. Other estimated coefficients can be interpreted using the same way.

Based on the Lake*Gender*Size effect plot, for alligators with size <= 2.3, we could find that males will more preferred to take invertebrate as their primary food choice in Lake Trafford and George than females and less preferred in Lake Hancock and Oklawaha. However, for alligators with size > 2.3, males would more preferred to take invertebrate in Lake Hancock, Oklawaha than females, equally likely to take invertebrate as females in Lake Trafford and less preferred to take Invertebrate than females in Lake George.

After fitting our final model, we are interested in the outliers and potential influential cases. Based on diagnostic plots and corresponding identified outliers, we see that few cases have large studentized residuals, which means that they are outliers in terms of y (Invertebrate). Some alligators are unusual in that they are small-sized Oklawaha male who have invertebrate as their primary food choice. And some alligators are unusual in that they are large-sized Hancock male who have invertebrate as their primary food choice. We also observe that small-sized Trafford male alligators have large hat values, which means that they are outliers in terms of x (Lake, Gender, Size). However, none of the cases above have a large Cook's distance, suggesting that even though these cases are identified as outliers, they are not influential.

3.2.3 Reptile as Response

The deviance test between M1 and M2 rejects null hypothesis with p-value 0.0011, suggesting that the more complex model (M2) is more suitable for the data. The deviance test between M2 and M3 accepts null hypothesis with p-value 0.0636, suggesting that the simpler model (M2) is more suitable for the data. The significance test for its coefficients shows that among interaction terms, only the interaction between Lake and Gender, and interaction between Lake and Size are highly significant. Therefore, we fit the model that only allows these interactions (M4), and compare it with M2. The deviance test between M2 and M4 accepts null hypothesis with p-value 0.2184, suggesting that the simpler model (M4) is more suitable for the data. The large p-value (0.9955) of GOF test for M4 indicates that this model fits the data well. Therefore, M4 will be our final model, which is: logit(π) = - 1.907 - 17.797 * _LO_ - 0.539 * _LT_ - 35.024 * _LG_ - 1.507 * _GF_ + 1.204 * _SG_ + 1.864 * _LO_:_GF_ + 3.176 * _LT_:_GF_ + 19.673 * _LG_:_GF_ + 17.737 * _LO_:_SG_ - 0.947 * _LT_:_SG_ + 16.644 * _LG_:_SG_ where π = P(Reptile = 1 | Lake, Gender, Size).

We can then interpret the estimated coefficients similar to section 3.2.1. For example, the estimated coefficient for the term _LO_, _GF_, _SG_, _LO:GF_, _LO:SG_, are -17.797, -1.507, 1.204, 1.864, 17.737 respectively, which means that Oklawaha female alligators with size greater than 2.3 meters are exp(-17.797 - 1.507 + 1.204 + 1.864 + 17.737) = 4.486 times as likely to have reptile as their primary food choice compared to the reference group. Other estimated coefficients can be interpreted using the same way.

Based on the Lake*Gender effect plot, we could find that male alligators would more preferred to choose reptile as their primary food choice in Lake Hancock and less preferred in Lake Trafford. For the other two lakes, both male and female alligators do not prefer reptile as their primary food choice (with probability close to 0). Based on the Lake*Size effect plot, we could find that large-sized alligators would more preferred to choose reptile as their primary food choice than small-sized alligators in all lakes expect Lake George. And both large-sized and small-sized alligators in Lake George do not prefer reptile as their primary food choice (with probability close to 0).

After fitting our final model, we are interested in the outliers and potential influential cases. Based on diagnostic plots and corresponding identified outliers, we see that many cases have large studentized residuals, which means that they are outliers in terms of y (Reptile). For example, some alligators are unusual in that they are small-sized Hancock female who have reptile as their primary food choice. And some alligators are unusual in that they are small-sized Trafford male who have reptile as their primary food choice. We also observe that large-sized Hancock male alligators who have retile as primary food choice have large hat values, which means that they are outliers in terms of x (Lake, Gender, Size). However, none of the cases above have a large Cook's distance, suggesting that even though these cases are identified as outliers, they are not influential.

3.2.4 Bird as Response

The deviance test between M1 and M2 rejects null hypothesis with p-value 1.876e-10, suggesting that the more complex model (M2) is more suitable for the data. The deviance test between M2 and M3 accepts null hypothesis with p-value 0.5104, suggesting that the simpler model (M2) is more suitable for the data. The significance test for its coefficients shows that among interaction terms, only the interaction between Lake and Size, and interaction between Gender and Size are highly significant. Therefore, we fit the model that only allows these interactions (M4), and compare it with M2. The deviance test between M2 and M4 accepts null hypothesis with p-value 0.1035, suggesting that the simpler model (M4) is more suitable for the data. The large p-value (0.9994) of GOF test for M4 indicates that this model fits the data well. Therefore, M4 will be our final model, which is: logit(π) = - 1.171 + 0.345 * *LO* + 1.502 * *LT* + 0.623 * *LG* - 1.113 * *GF* - 1.324 * *SG* - 1.768 * *LO*:*SG* - 21.019 * *LT*:*SG* - 4.233 * *LG*:*SG* + 4.946 * *GF*:*SG* where π = P(Bird = 1 | Lake, Gender, Size).

We can then interpret the estimated coefficients similar to section 3.2.1. For example, the estimated coefficient for the term *LO*, *GF*, *SG*, *LO:SG*, *GF:SG* are 0.345, -1.113, -1.324, -1.768, 4.946 respectively, which means that Oklawaha female alligators with size greater than 2.3 meters are exp(0.345 - 1.113 - 1.324 - 1.768 + 4.946) = 2.962 times as likely to have bird as their primary food choice compared to the reference group. Other estimated coefficients can be interpreted using the same way.

Based on the Lake*Size Effect plot, we could find that large-sized alligators would more preferred to choose bird as their primary food choice in Lake Hancock but less preferred in other three lakes. Based on the Gender*Size effect plot, we could find that for small-sized alligators, males would more preferred to choose bird as their primary food choice than females. However, for large-sized alligators, both males and females do not prefer bird as their primary food choice (with probability close to 0).

After fitting our final model, we are interested in the outliers and potential influential cases. Based on diagnostic plots and corresponding identified outliers, we see that many cases have large studentized residuals, which means that they are outliers in terms of y (Bird). For example, some alligators are unusual in that they are large-sized Hancock male who have bird as their primary food choice. And some alligators are unusual in that they are large-sized George female who have bird as their primary food choice. We also observe that small-sized Trafford male alligators have large hat values, which means that they are outliers in terms of x (Lake, Gender, Size). However, none of the cases above have a large Cook's distance, suggesting that even though these cases are identified as outliers, they are not influential.

3.2.5 Other as Response

The deviance test between M1 and M2 rejects null hypothesis with p-value 3.185e-06, suggesting that the more complex model (M2) is more suitable for the data. The deviance test between M2 and M3 rejects null hypothesis with p-value 0.00928, suggesting that the more complex model (M3) is more suitable for the data. The large p-value (1) of GOF test for M3 indicates that this model fits the data well. Therefore, M3 will be our final model, which is: logit(π) = - 1.705 + 1.956 * *LO* - 17.861 * *LT* - 17.861 * *LG* - 0.492 * *GF* - 0.087 * *SG* - 1.012 * *LO*:*GF* + 19.548 * *LT*:*GF* + 39.624 * *LG*:*GF* - 19.730 * *LO*:*SG* + 16.475 * *LT*:*SG* + 19.535 * *LG*:*SG* - 0.488 * *GF*:*SG* + 1.992 * *LO*:*GF*:*SG* - 34.954 * *LT*:*GF*:*SG* - 41.522 * *LG*:*GF*:*SG* where π = P(Other = 1 | Lake, Gender, Size).

We can then interpret the estimated coefficients similar to section 3.2.1. For example, the estimated coefficient for the term *LO*, *GF*, *SG*, *LO:GF*, *LO:SG*, *GF:SG*, *LO:GF:SG* are 1.956, -0.492, -0.087, -1.012, -19.730, -0.488, 1.992 respectively, which means that Oklawaha female alligators with size greater than 2.3 meters are exp(1.956 - 0.492 - 0.087 - 1.012 - 19.730 - 0.488 + 1.992) = 1.750e-8 times as likely to have "Other" as their primary food choice compared to the reference group. Other estimated coefficients can be interpreted using the same way.

Based on the Lake*Gender*Size effect plot, we could find that for small-sized alligators, males would more preferred to choose "Other" as their primary food choice than females in Lake Hancock and Oklawaha, but less preferred in Lake Trafford and George. For large-sized alligators, males have a higher probability to choose "Other" as their primary food choice in all lakes except Lake Oklawaha where both males and females do not prefer "Other" as their primary food choice (with probability close to 0). Also, we could find that for female alligators, small-sized alligators always have a higher probability to choose “Other” as their primary food choice than large-sized alligators.

After fitting our final model, we are interested in the outliers and potential influential cases. Based on diagnostic plots and corresponding identified outliers, we see that many cases have large studentized residuals, which means that they are outliers in terms of y (Other). For example, some alligators are unusual in that they are large-sized Trafford male who have “Other” as their primary food choice. And some alligators are unusual in that they are large-sized George female who have “Other” as their primary food choice. We also observe that small-sized Trafford male alligators who have “Other” as primary food choice have large hat values, which means that they are outliers in terms of x (Lake, Gender, Size). However, none of the cases above have a large Cook’s distance, suggesting that even though these cases are identified as outliers, they are not influential.

3.3 Logistic Regression Models for Polytomous Response

In this section, instead of creating binary responses and fitting logistic regression models for each primary food type, we will directly fit generalized logit models for the polytomous responses Primary Food Type. For brevity, we will use the same abbreviations as section 3.2 when writing the models (e.g. *LakeOklawaha:GenderFemale* will be denoted by *LO:GF*). Also, we define π_j to be P(Primary Food Type = j | Lake, Gender, Size). Here, j can be 1, …, 5, representing Fish, Invertebrate, Reptile, Bird, Other, respectively. The reference level is Fish.

Similar to section 3.2, we first fit 3 different generalized logit models: main effect model (M1), model that allow all 2-way interactions (M2), model that allow 3-way interaction (M3). However, a warning was issued when fitting M3, indicating that the model (M3) was too complex relative to the sample size. Therefore, we will not consider M3 in this section. The BIC of M1 is 8.9925 lower than that of M2, but its AIC is 85.9015 higher than M2’s AIC. Based on this, we believe that M2 is generally better than M1. Then, we fit all possible models that allow 2-way interactions and compare them based on AIC and BIC. The best models we find are M2 and the model that allow interaction between Lake and Size, and interaction between Gender and Size (M4). M2 has a lower AIC compare to M4 (564.7914 vs. 575.2975), and M4 has a lower BIC compare to M2 (710.8604 vs. 741.0231).

We first look at M2, which is:

- logit(π_2/π_1) = - 2.757 + 3.000 * *LO* + 18.831 * *LT* + 18.469 * *LG* + 1.570 * *GF* + 2.064 * *SG* - 2.720 * *LO*:*GF* + 1.514 * *LT*:*GF* + 0.972 * *LG*:*GF* - 0.402 * *LO*:*SG* - 18.942 * *LT*:*SG* - 16.968 * *LG*:*SG* - 2.333 * *GF*:*SG*
- logit(π_3/π_1) = - 1.847 - 14.874 * *LO* + 17.959 * *LT* - 13.190 * *LG* + 0.125 * *GF* + 2.796 * *SG* + 1.500 * *LO*:*GF* + 4.867 * *LT*:*GF* + 19.155 * *LG*:*GF* + 15.244 * *LO*:*SG* + 21.600 * *LT*:*SG* - 4.726 * *LG*:*SG* - 1.621 * *GF*:*SG*
- logit(π_4/π_1) = - 0.645 + 2.013 * *LO* + 16.800 * *LT* + 16.909 * *LG* - 1.918 * *GF* - 0.404 * *SG* - 0.353 * *LO*:*GF* + 8.173 * *LT*:*GF* + 0.938 * *LG*:*GF* - 2.156 * *LO*:*SG* - 45.688 * *LT*:*SG* - 20.255 * *LG*:*SG* + 4.952 * *GF*:*SG*
- logit(π_5/π_1) = - 1.923 + 3.999 * *LO* + 15.951 * *LT* + 16.531 * *LG* + 0.978 * *GF* + 2.492 * *SG* - 3.715 * *LO*:*GF* + 7.032 * *LT*:*GF* + 5.104 * *LG*:*GF* - 28.930 * *LO*:*SG* - 19.564 * *LT*:*SG* - 16.218 * *LG*:*SG* - 7.499 * *GF*:*SG*

We can then interpret the estimated coefficients in terms of odds ratio based on this model. For example, the estimated coefficient for the term *LO*, *GF*, *SG*, *LO:GF*, *LO:SG*, *GF:SG* in logit(π_2/π_1) are 3.0, 1.57, 2.064, 2.72, - 0.402, - 2.333, respectively, which means that Oklawaha female alligators with size greater than 2.3 meters are exp(3.0 + 1.57 + 2.064 + 2.72 - 0.402 - 2.333) = 749.196 times as likely to have invertebrate as their primary food choice compared to fish. Other estimated coefficients can be interpreted using the same way.

Based on the Lake*Gender*Size effect plot of M2 (shown in Figure 2), we could find that: For Size > 2.3 Male Alligators, in Lake Hancock, Probability of Primary Food Type dominant by Reptile and Other, some will also choose Fish but only few will choose Invertebrate and Bird; in Lake Oklawaha, Probability of Primary Food Type dominant by Invertebrate and some will choose Reptile, with few or no probability on others types; in Lake Trafford, Probability of Primary Food Type dominant by Fish and some will choose Invertebrate, with few or no probability on others types; in Lake George, Probability of Primary Food Type dominant by both Invertebrate and Other, some will also choose Fish but no one will choose other two types. For Size <= 2.3 Male Alligators, in Lake Hancock, Probability of Primary Food Type is dominant by Fish, some will also choose Bird, only few will choose other three types; in Lake Oklawaha, Probability of Primary Food Type is dominant by Other, some will also choose Bird, few or no will choose others types; in Lake Trafford, Invertebrate, Reptile, and Bird are three dominant Primary Food Type with same probability and few will choose Other type; in Lake George, Probability of Primary Food Type is dominant by Bird and some will choose Invertebrate, few will choose Other, no one will choose other two types. For Size > 2.3 Female Alligators, in Lake Hancock, Probability of Primary Food Type is dominant by Bird, few or no will choose others types; in Lake Oklawaha, Reptile and Bird are two dominant Probability of Primary Food Type with same probability, few or no will choose others types; in Lake Trafford, Probability of Primary Food Type is dominant by Reptile, some will also choose Fish and Invertebrate, few or no will choose others two types; in Lake George, there is no dominant Primary Food Type, but Invertebrate will be more popular, the second popular one will be Reptile, only few will choose others three types. For Size <= 2.3 Female Alligators, in Lake Hancock, Probability of Primary Food Type is dominant by Fish, some will choose Invertebrate and Other, few will choose others two types; in Lake Oklawaha, Fish is the most popular Primary Food Type, others three types except Reptile will have similar probability for each; in Lake Trafford, Bird is the most popular Primary Food Type, the second will be Other, few or no will choose others three types; In Lake George, Probability of Primary Food Type is dominant by Other, few will choose Invertebrate and others three types has 0 probability.

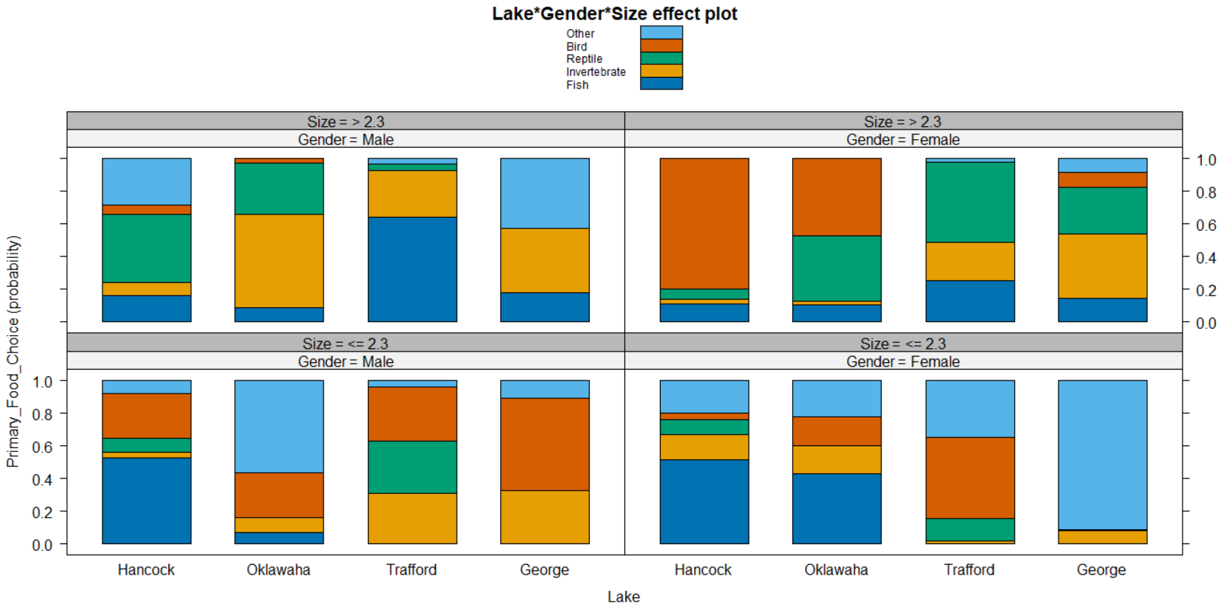

Figure 2: Effect plot of generalized logit model M2

We then look at M4, which is:

- logit(π_2/π_1) = - 1.355 + 1.215 * *LO* + 17.496 * *LT* + 17.715 * *LG* - 1.033 * *GF* + 0.315 * *SG* + 1.047 * *LO:SG* - 17.028 * *LT:SG* - 15.589 * *LG:SG* + 0.737 * *GF:SG*
- logit(π_3/π_1) = - 1.252 - 14.081 * *LO* + 18.018 * *LT* - 4.252 * *LG* - 1.424 * *GF* + 0.775 * *SG* + 15.268 * *LO:SG* - 19.399 * *LT:SG* + 3.308 * *LG:SG* + 3.241 * *GF:SG*
- logit(π_4/π_1) = - 0.525 + 1.189 * *LO* + 18.063 * *LT* + 17.397 * *LG* - 1.580 * *GF* - 1.042 * *SG* - 1.521 * *LO:SG* - 38.826 * *LT:SG* - 21.015 * *LG:SG* + 6.154 * *GF:SG*
- logit(π_5/π_1) = - 1.309 + 2.165 * *LO* + 17.228 * *LT* + 18.249 * *LG* - 0.146 * *GF* + 1.109 * *SG* - 21.410 * *LO:SG* - 19.724 * *LT:SG* - 16.933 * *LG:SG* - 1.651 * *GF:SG*

We can then interpret the estimated coefficients using the similar way as M2. For example, the estimated coefficient for the term *LO*, *GF*, *SG*, *LO:SG*, *GF:SG* in logit(π_2/π_1) are 1.215, - 1.033, 0.315, 1.047, 0.737 respectively, which means that Oklawaha female alligators with size greater than 2.3 meters are exp(1.215 - 1.033 + 0.315 + 1.047 + 0.737) = 9.786 times as likely to have invertebrate as their primary food choice compared to fish. Other estimated coefficients can be interpreted using the same way.

Based on Lake*Size effect plot of M4 (shown in Figure 3), we could find that probability varies on different Lake and Size. For Size <= 2.3, in Lake Hancock, Probability of Primary Food Type is dominant by Fish, the others four types have the similar probability; in Lake Oklawaha, Probability of Primary Food Type is dominant by Other, some will choose the other three types except Reptile; in Lake Traffrod, the most popular Primary Food Type will be Bird, some will choose the other three types except Fish; in Lake George, Probability of Primary Food Type is dominant by Other, some will choose Invertebrate and Bird but no one will choose the other two types. For Size > 2.3, in Lake Hancock, the most popular type will be Bird, the second will be Reptile, some will choose Fish but only few will choose Invertebrate and Other; in Lake Oklawaha, Probability of Primary Food Type is dominant by Reptile, some will choose Invertebrate, few will choose Bird and Fish and no one will choose Other type; in Lake Trafford, Probability of Primary Food Type is dominant by Fish, some will choose Invertebrate and Reptile, few or no will choose others two types; in Lake George, the most popular Primary Food Type will be Invertebrate, some will choose Other, Fish, and Reptile, only few will choose Bird.

Based on Gender*Size effect plot of M4 (shown in Figure 3), we could find that probability varies on different Gender and Size. For Size <= 2.3, Male's most popular Primary Food Type is Bird, the second will be Other, the third will be Invertebrate, and other two types with 0 probability; however, Female differ from Male will be the most popular will be Other and second will be Other, remain the same for everything else. For Size > 2.3, Male's most popular Primary Food Type will be both Fish and Invertebrate, some will choose Reptile, the other two types will have 0 probability; for Female, the most popular Primary Food Type will be Reptile, some will choose Fish and Invertebrate, only few will choose Bird and no one will choose Other.

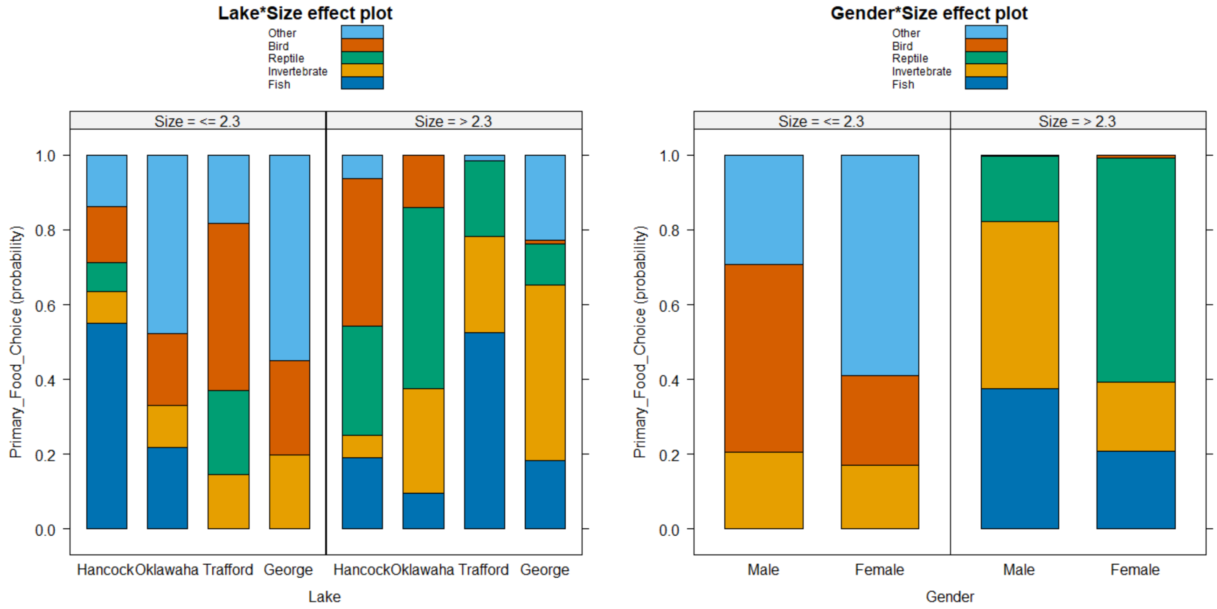

Figure 3: Effect plots of generalized logit model M4

4 Discussion

Our study aims to explore the associations between the response (primary food type of alligators) and each independent variables (lake of capture, gender, and size) with conditioning on the other remaining independent variables, and the associations between independent variables. In section 2, the stratified analysis tells us that there is an association between the response and each independent variable given one other independent variable. Then the doubledecker plot give us some rough idea about the dependence of primary food type of alligators on independent variables. However, this dependence is still unclear. In order to further investigate the associations mentioned above, we make the mosaic matrices and fit log-linear models under different association assumptions. The final well fitted log-linear model suggests that given Primary Food Type and Size, Lake and Gender will be independent of each other, and that there is an association between Lake and Gender which does not depend on other variables. We are also interested in how the combination of independent variables, along with the potential interactions between them, affects the primary food choice of alligators. Therefore, we fit several logistic regression models, some of which are for binary response variables and some of which are for polytomous response variables. Our fitted logistic regression models for binary response fit the data well and are really helpful for us to see how independent variables affect each primary food type by looking at the estimated coefficients.

Our study is strongly limited by the number of samples in the dataset. When we try to investigate the marginal distribution of the response by constructing one-way frequency tables, we found that only L9 can be used due to the lack of data. Also, we cannot fit generalized logit model which allows 3-way interactions since it is too complex, relative to our small sample size. If we can find additional datasets about the primary food type of alligators in Florida lakes, then we might be able to have a more detailed and meaningful analysis.

5 Peer assessment

Zihan Wang is responsible for rewriting code & analysis & summary of section 1 – 3.1. Yiquan Xiao is responsible for rewriting code & analysis & summary of section 3.2 & 3.3. Yusen Wang is responsible for code & analysis of section 3.3 and writing of section 4.

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.