

# ***Burmese Pythons in Southern Florida***

Gianluca Mangiapane

MUSA/LARP 745-401 Advanced Topics in GIS

## **Introduction**

When I was living in Australia and New Zealand, one of the major concerns of those countries were invasive species being introduced and impacting native flora and fauna populations. Rabbits and Cane Toads create extra food and habitation competition for Australian mammals and amphibians, while weasels and stoats in New Zealand ricked wiping out the bird population that had no defenses against those predators. Considering they are island nations with delicate ecosystems, Australia and New Zealand dedicate wildlife resources to curb and control invasive species populations. One can witness these programs in place even on popular nature walks through out the two countries, with traps and cages set up to find these pests.

I never took much thought of invasive species until seeing the effort Australia and New Zealand made throughout my time there, when I moved after my undergraduate degree. Growing up in Northern New Jersey, the only real invasive species I witnessed were the stink bugs that always seemed to come out in the summer, and even then, I only really thought of them as a nuisance more than harm to the local environment. When I came back to the United States this past year, I decided to find out more on invasive species in America that are currently becoming a major issue, and unfortunately found out that its not only retirees making Florida home, but also Burmese Pythons. First spotted in 1979, the Burmese Python population had boomed in the past 20 years. Now, the fragile everglades ecosystem is at risk, due to the python's voracious appetite and breeding skills, in a region that never had snakes of this magnitude in its evolution.

In 2017, Florida introduced a bounty system, where registered and certified hunters can catch Burmese pythons, and get paid depending on the size of the python. I find this an interesting method to combat the spread and population of the pythons, only because Great Britain introduced a very similar program during their rule in India to fight against an increasing population of cobras. This led to an increase in cobra populations, for hunters were releasing cobras into the wild to be caught and paid for, increasing the amount of money they got. It incentivized hunters to not completely wipe out their source of income, for why would they want to get rid of the snakes if they were getting paid for catching/killing them? This has led me to think are there different or better ways to identify and combat the spread of Burmese pythons?

## **Literature Review**

According to Michael Sarill at Berkeley's Rausser College of Natural Resources, as of 2013, scientists were estimating the invasive Burmese python species to be between 30,000 and 150,000 in South Florida (Sarill, 2016). That broad range

shows how hard it is to accurately assess just how many snakes there are in Florida, due to their camouflaged and nature.

However, there has been some ability to monitor the snake population due to the changes in prey animal population. I plan on using a sample of these bird species to see if they exist outside the everglades and what they're distribution is like to assess any similarities with Python sightings. Dove et Al. (2011) identified over 25 species of birds in the digestive tracts of 85 Burmese pythons collected in the everglades, showing that Florida birds take up a main staple of the Burmese Python diet. In this same framework, I will include looking at mammalian habitat areas in Florida as well, so see if there is a relationship between mammal dispersion and where the pythons have been sighted. In terms of mammals, over the last decade, mammal observations in the everglades have gone down, with 87.5% of the bobcat population declined, 94.1% of the white-tail deer population declined, 98.9% of opossum population declined, 99.3% of raccoon population declined, 100% of rabbit population declined and 100% of fox population declined in the Everglades (Sarill, 2016). The animals above have no instinctive fear of a snake the size of the Burmese Python, and as a result can be easy prey. The last snake the size of the Python to exist in the Everglades became extinct 16 million years ago (Sarill, 2016).

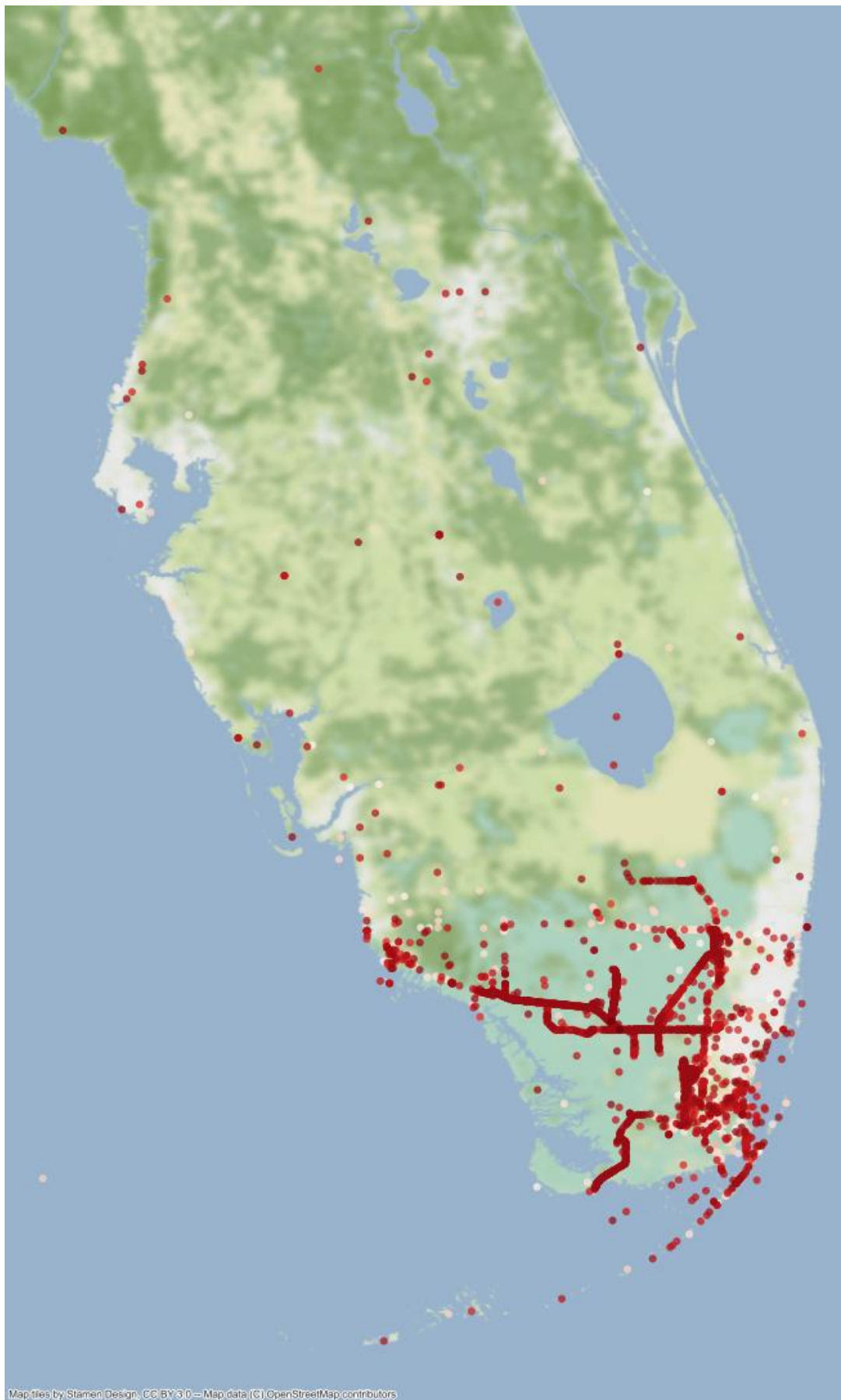
In 2016, the Wetland and Aquatic Research Center published a report that current industry methods for detection and control provide low detection rates, with the methods being detector dogs, remote sensing, attractant traps, judas snakes (methods currently used by hunters, in addition to old fashioned walking and searching) However, Environmental DNA is a promising method of detection, which works by detecting discarded DNA material (skin, droppings) in water, collecting those water samples, and then testing for that python DNA to determine if snakes are present in those areas, which helps to pinpoint specific areas to search for pythons in. Hunter et Al. (2015) did a study looking at the effect of Environmental DNA PCR testing in isolating and detecting Burmese Python DNA amongst various reptilian samples and found their occupancy models to have a 91% effectiveness rate on detecting Burmese pythons out of all other animal samples. Thus, environmental DNA is capable in being able to identify trails of Burmese Pythons, and help with search locations of the actual animal, since it can verify if they have passed in the area to a degree of accuracy. Bartoszek et Al. (2021) has taken the step further and investigated the spatial ecology of Burmese Pythons in Florida and found the mean annual home range to be  $7.5 \text{ km}^2 \pm 2.9 \text{ km}^2$  (95% kernel density estimate), with pythons moving at a maximum mean rate of 0.52 kilometers a day. They also found various land cover types that Burmese Pythons tended to move in, such as freshwater and saline wetlands, but avoided open water and urban land cover (Bartoszek et Al, 2021). Therefore, if Environmental DNA sampling can identify Burmese Python trails, and there is known patterns in spatial behavior of Pythons, then it can be easier to find and detect the python once Environmental DNA has confirmed they have been in the area. Yet sampling is limited and costly, and it would be over ambitious to sample the extremity of Florida. So, the question remains, where would be the areas best suited to sample, or at least is there a pattern in the sightings to identify features and locations that will have better probability of finding a Burmese python? Can I find correlations amongst certain animal, geographical, and human variables that help identify where pythons are more likely

to be, allow sampling and detecting efforts to focus on certain regions or areas over others, and thus not lose time and resources in areas that will yield inadequate results?

## **Feature Engineering and Data handling Methods**

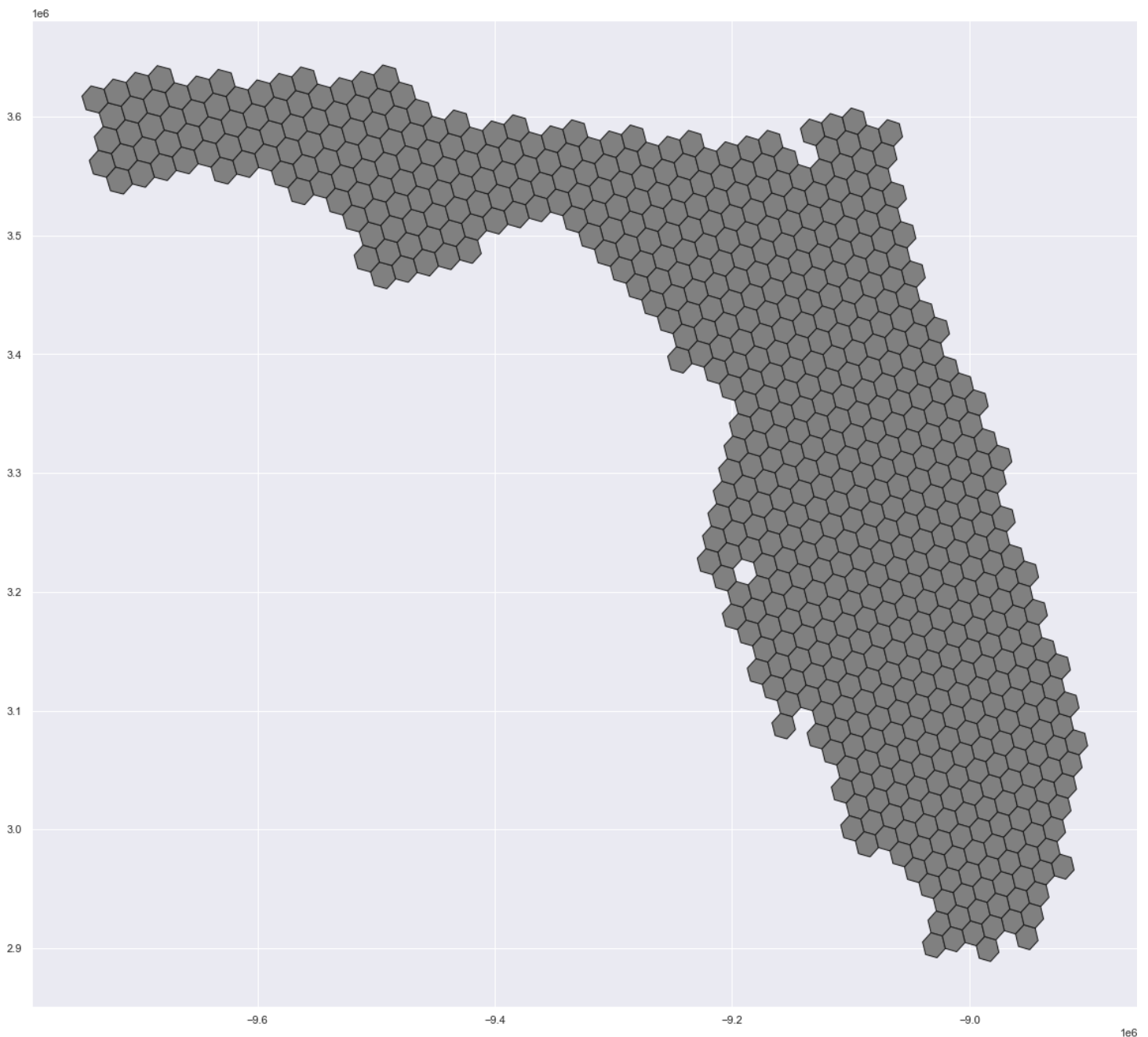
The University of Georgia's Center for Invasive Species and Ecosystem Health runs EddMapS, which has collected all known Burmese Python sightings from 1979 to current day, with their known coordinates, day collected, and some rudimentary information such as how it was sighted, any notes and comments, organization, etc, but for most of the sightings this was left blank. The only consistent data filled out was the date, coordinates, comments, and that the sighting has been verified by the Florida Fish and Wildlife Conservation Commission. For my research, I am looking at the years 2016 to 2020, which is one year before the start of the Florida Burmese Python bounty collecting program, and just one year into the COVID-19 pandemic that started. In that four-year time, 4,587 pythons were sighted and collected in the State of Florida, with their dispersion shown in Figure 1. Looking at that initial assessment, one can see sightings clustered in straight lines that extend in different directions, which indicate sightings that have occurred along to right next to roads. Sightings Cluster in Southern Florida, which makes sense due to the everglades being situated in Florida, and where the population of invasive Pythons first started. The area of Miami-Dade County has numerous sightings, which initial thoughts could be due to human population being most dense there, and thus more likely for someone to spot or find a python. However, there are sightings that do extend up into Central Florida, and out west into the Everglades, and down South into the Keys.

Figure 1: 4,587 python sightings in the state of Florida between 2016 and 2020



To assess spatial distribution patterns of the Pythons, I used Uber's H3 Hexagonal Hierarchical Spatial Indexing library that splits a set geographical boundary into hexagons of equal area, dependent on the resolution size set by the user. I set my grid to Resolution 5, which gives an average hexagonal area of 252 km<sup>2</sup>, and an average hexagonal edge length of 8.5 km. Resolution size was done for sake of computational power and time, and I want to explore if there is a way to find the most effective or optimized resolution size to use in analysis moving forward. Figure 2 shows the hexagonal split of the state of Florida into 645 equal hexagons. The empty hexagon on the middle left represents the body of water that is Tampa Bay. Already, one can see that python sightings for 2016-2020 did not stray into the Northern Florida or the Panhandle currently, but that doesn't mean to say sightings before or after this time period didn't occur on those areas.

**Figure 2: Hexagonal Grid of Florida**



After creating my hexagonal grid, I spatially joined my python Data to my grid, so that there was a count of python sightings per hexagon for the period. 116 hexagons out of the 645 hexagons had python sightings, with the highest

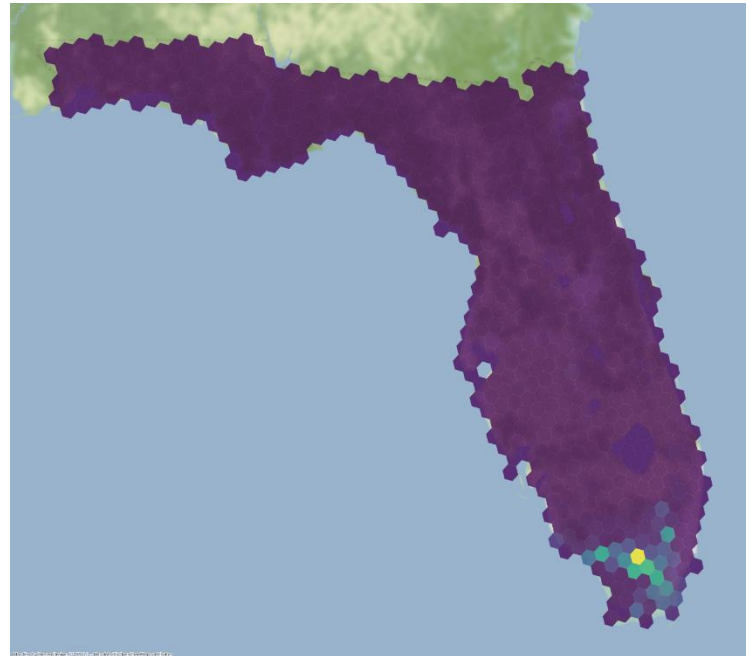
count being 512, show in Figure 3. Figure 4 is the distribution plotted in a choropleth map, and as stated earlier, the cluster of most sightings is distributed around Southern Florida.

**Figure 3: Count of Pythons per Hex**

	hex_id	pythonsPerHex
22	8544127bffffff	512
12	8544124ffffff	338
10	85441247ffffff	306
8	8544123bffffff	287
11	8544124bffffff	287
...	...	...
80	8544a8c3ffffff	1
81	8544a8dbffffff	1
82	8544a903ffffff	1
84	8544a90ffffff	1
115	8544f4cbffffff	1

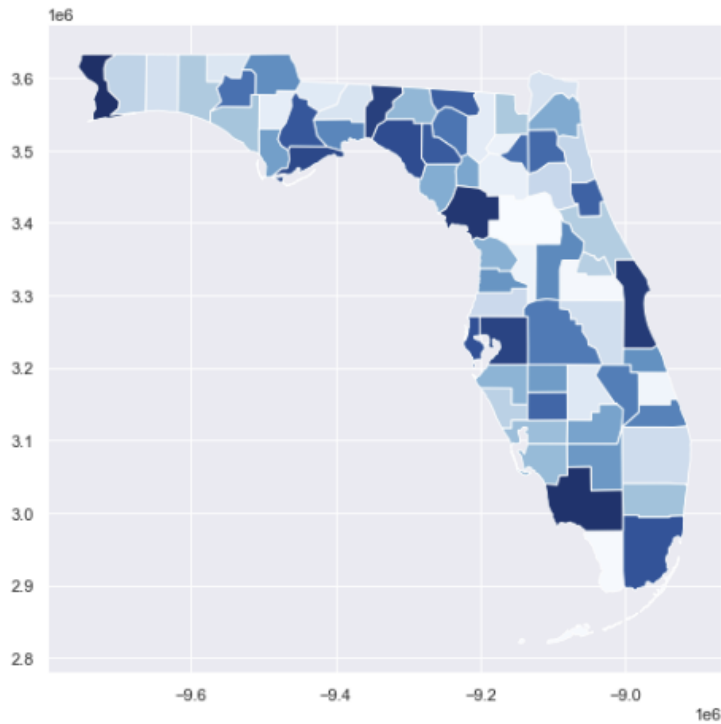
116 rows × 2 columns

**Figure 4: Choropleth map of Sightings per Hexagon**



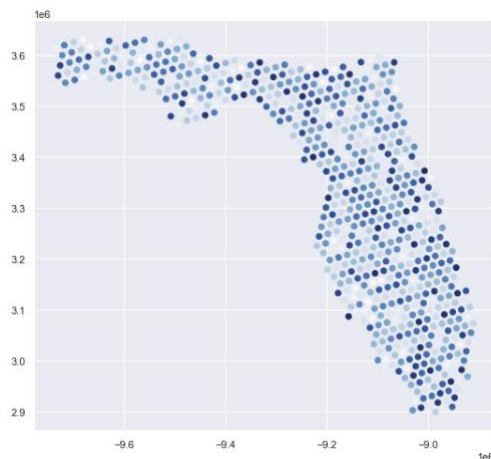
Once the pythons per hexagon was created, I then added in the count of months for when pythons were sighted, along with the corresponding county each hexagon is in. This will control for the county and will also assess if the time of the year (i.e., month) is correlated with sightings. Most sightings were found in the summer months but is this due to python activity or due to more humans out looking for them or travelling. The count of months was added in by essentially counting how many sightings per month were in each respective hexagon, and having that count added into the data frame. Thus, adding in the month variable will control for time of the year for sightings. I binarized the county data by One Hot encoding, in that for each county, a hexagon was assigned a 1 for being within that county, and a 0 if it was not in that county. This makes the categorical variable workable in my modelling methods. In addition to county data, I also found the median income per county and the population per county from the 2016 ACS API, and brought that down to the per hex level, by finding the population density and income density per hexagonal area. My thought process here is that the population again controls for human activity, and the income level controls for the Florida Bounty program. Since hunters are paid per snake and size of snake, my inference is that those who would potentially have lower income levels may be incentive or inclines to get certified and partake in the bounty program so earn more money. I used median income from the 2016 ACS variables, but perhaps a better move in the next round of research could be including poverty rate. Figure 5 shows the counties of Florida for reference.

**Figure 5: Counties of Florida**



Much of my feature engineering was finding the distance of each hexagon relative to other spatial variables. I took the centroid of each hexagon and applied a distance function to various other features to assess the nearest neighbor to each hexagonal centroid of that specific feature. Figure 6 below shows the centroids of each hexagon.

**Figure 6: Centroids of Florida hexagons**



For example, I read in Bird Sanctuary data in Florida, and found the nearest bird sanctuary to each respective hexagonal centroid, in units of meters. This distance was then created as a variable and merged with each respective hexagon. Thus, each hexagon had a count of python, along with the distance to the closest bird sanctuary. I applied this method for all animal and geographic features that I obtained from the state government of Florida's open geospatial data portal. In

addition to Bird Sanctuaries, I used eagle nest sites, bird nest sites mapped from 2000, land mammal habits, bird habitat areas from 2003, shorebird habitats, snowy plover nests, wading bird colonies, urban boundaries, lakes, springs, rivers, flowing water resources, wading bird rookeries, stormwater activity sites, waterbodies, and submerged lands. The respective hexagonal distances to the closest geometric feature of those datasets were then merged into my data frame.

Finally, since I noticed in the mapping of sightings that the points followed a pattern along roads, I pulled in traffic data from 2016 to 2020, with the average annual daily traffic density (AADT) per all roads and highways in Florida. I again took the nearest neighbor distance of each hexagonal centroid to points along road geometries, and for hexagons that had roads within their boundaries, I also merged the average four-year AADT to those hexagons, to control for humans driving along roads and making sightings just by happenstance of driving past a python.

Thus, my explanatory variable is `pythonsPerHex`, which represents the count of pythons in each hexagon, and my independent variables are as follows, below, in Table 1.

**Table 1: Independent Variables to be regressed against `pythonsPerHex` (Count of Pythons)**

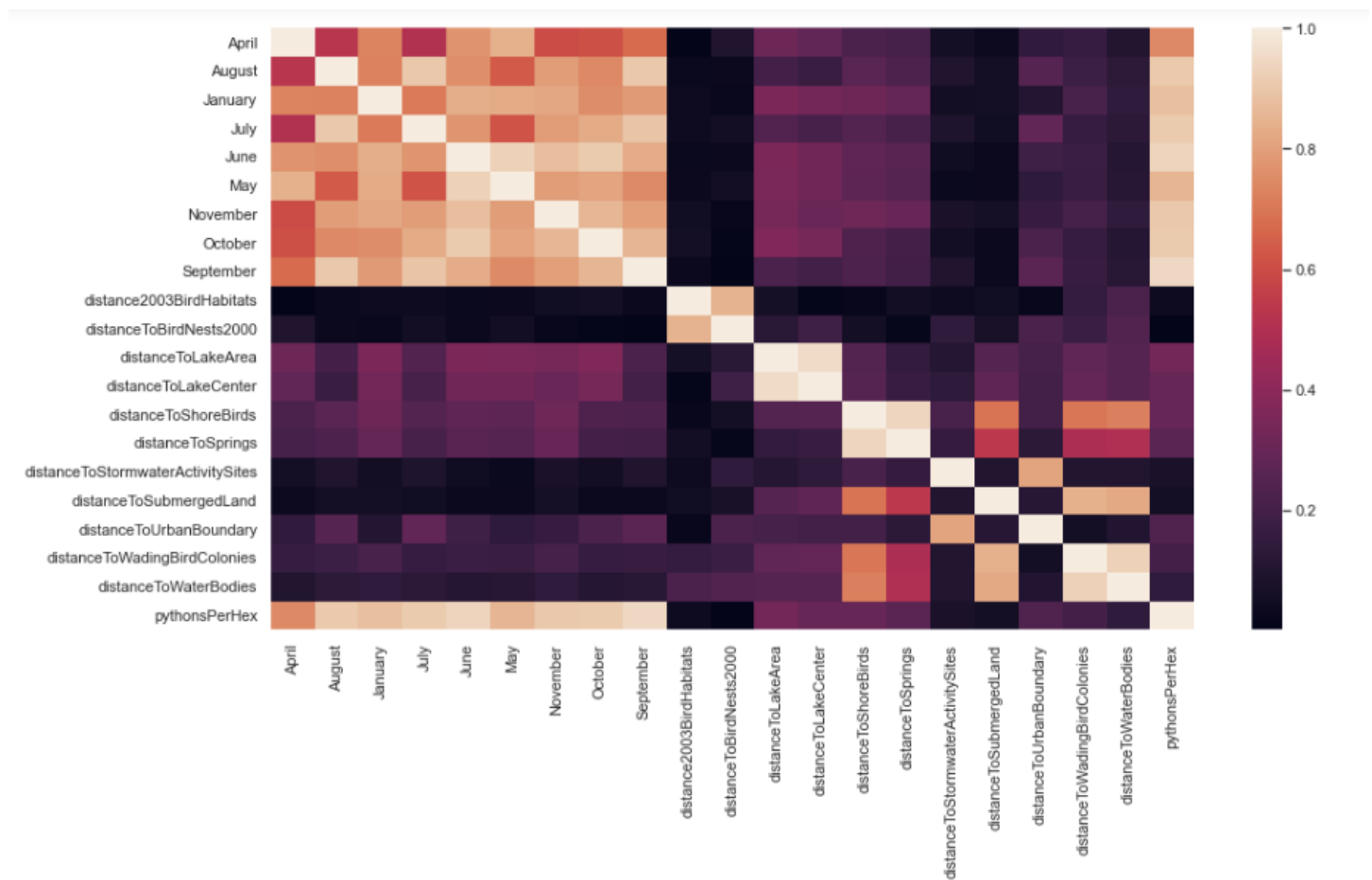
April	NAME_Citrus	NAME_Indian River	NAME_Pasco	distanceToMammalHabitats
August	NAME_Clay	NAME_Jackson	NAME_Pinellas	distance2003BirdHabitats
December	NAME_Collier	NAME_Jefferson	NAME_Polk	distanceToShoreBirds
February	NAME_Columbia	NAME_Lafayette	NAME_Putnam	distanceToRoad
January	NAME_DeSoto	NAME_Lake	NAME_Santa Rosa	AADT_x (AADT per Hex)
July	NAME_Dixie	NAME_Lee	NAME_Sarasota	distanceToPloverNests
June	NAME_Duval	NAME_Leon	NAME_Seminole	ploverNestsPerHex
March	NAME_Escambia	NAME_Levy	NAME_St. Johns	distanceToWadingBirdColonies
May	NAME_Flagler	NAME_Liberty	NAME_St. Lucie	distanceToUrbanBoundary
November	NAME_Franklin	NAME_Madison	NAME_Sumter	distanceToUrbanCenter
October	NAME_Gadsden	NAME_Manatee	NAME_Suwannee	distanceToLakeArea
September	NAME_Gilchrist	NAME_Marion	NAME_Taylor	distanceToLakeCenter
NAME (County)	NAME_Glades	NAME_Martin	NAME_Union	distanceToSprings
NAME_Alachua	NAME_Gulf	NAME_Miami-Dade	NAME_Volusia	distanceToRivers
NAME_Baker	NAME_Hamilton	NAME_Monroe	NAME_Wakulla	distanceToFlowingWater
NAME_Bay	NAME_Hardee	NAME_Nassau	NAME_Walton	distanceTo1999WadingBirdRookeries
NAME_Bradford	NAME_Hendry	NAME_Okaloosa	NAME_Washington	distanceToStormwaterActivitySites
NAME_Brevard	NAME_Hernando	NAME_Okeechobee	distanceToBirdSanctuaries	distanceToWaterBodies
NAME_Broward	NAME_Highlands	NAME_Orange	eagleNestsPerHex	distanceToSubmergedLand
NAME_Calhoun	NAME_Hillsborough	NAME_Osceola	distanceToEagleNests	INCOME_Hex
NAME_Charlotte	NAME_Holmes	NAME_Palm Beach	distanceToBirdNests2000	POP_Hex



However, there remains two more things I do before the modelling stage. First, I drop all hexagons with counts of 0 in from my data since I can't know for sure if those hexagons are counts of zero or not. Python sightings don't represent occurrences of an event happening without human observations, in that a hexagon with a count of zero could potentially have pythons in it, they just haven't been found yet. Including those hexagons will create bias in my dataset as being an actual count of zero and skewing the correlation of the independent variables have on python count. I can only include the hexagons with actual known counts of 1 or more, since that will provide a known instance of a sighting, and the relationship amongst the variables will have more meaning.

The second, and last action before I model is filtering out the independent variables that have an absolute Pearson correlation greater than 0.8, when compared to at least one other independent variable in my data sample. I found a function that filters out and shows the correlation matrix of all independent variables that have at least a correlation with another variable that's greater than 0.8. Those variables are then removed from my data sample, since a remaining variable will already represent that relationship in the model.

**Figure 7: Correlation Matrix of variables with Pearson correlation value above 0.8**



The independent variables dropped are, 'August', 'January', 'July', 'June', 'May', 'November', 'October', 'September', 'distance2003BirdHabitats', 'distanceToLakeCenter', 'distanceToSprings', 'distanceToStormwaterActivitySites', 'distanceToWaterBodies', and 'distanceToSubmergedLand'.

This leaves me with the following independent variables to use in modelling, as shown in table 2.

**Table 2: Post Correlation Matrix Filtered Out Independent Variable**

April	NAME_Citrus	NAME_Gulf	NAME_Lee
December	NAME_Clay	NAME_Hamilton	NAME_Leon
February	NAME_Collier	NAME_Hardee	NAME_Levy
March	NAME_Columbia	NAME_Hendry	NAME_Liberty
NAME	NAME_DeSoto	NAME_Hernando	NAME_Madison
NAME_Alachua	NAME_Dixie	NAME_Highlands	NAME_Manatee
NAME_Baker	NAME_Duval	NAME_Hillsborough	NAME_Marion
NAME_Bay	NAME_Escambia	NAME_Holmes	NAME_Martin
NAME_Bradford	NAME_Flagler	NAME_Indian_River	NAME_Miami_Dade
NAME_Brevard	NAME_Franklin	NAME_Jackson	NAME_Monroe
NAME_Broward	NAME_Gadsden	NAME_Jefferson	NAME_Nassau
NAME_Calhoun	NAME_Gilchrist	NAME_Lafayette	NAME_Okaloosa
NAME_Charlotte	NAME_Glades	NAME_Lake	NAME_Okeechobee
NAME_Orange	NAME_Suwannee	distanceToRoad	
NAME_Osceola	NAME_Taylor	AADT_x	
NAME_Palm_Beach	NAME_Union	distanceToPloverNests	
NAME_Pasco	NAME_Volusia	ploverNestsPerHex	
NAME_Pinellas	NAME_Wakulla	distanceToWadingBirdColonies	
NAME_Polk	NAME_Walton	distanceToUrbanBoundary	
NAME_Putnam	NAME_Washington	distanceToUrbanCenter	
NAME_Santa_Rosa	distanceToBirdSanctuaries	distanceToLakeArea	
NAME_Sarasota	eagleNestsPerHex	distanceToRivers	
NAME_Seminole	distanceToEagleNests	distanceToFlowingWater	
NAME_St_Johns	distanceToBirdNests2000	distanceTo1999WadingBirdRookeries	
NAME_St_Lucie	distanceToMammalHabitats	INCOME_Hex	
NAME_Sumter	distanceToShoreBirds	POP_Hex	

## Model Selection

Since I am looking to model the relationship between the independent variables and the count of pythons per hexagon, I know the base model to use will be a Poisson model. However, one of the assumptions of a generalized Poisson model is that the mean and variance of counts are equal, but in looking at our data of pythonsPerHex, the mean is 37.06 and the variance is 6,638, so the assumption of a generalize Poisson model where variance = mean is void. Therefore, the model I use is a negative binomial regression, which seeks to determine the value of  $\alpha$  (which is assumed to be 1 in a general Poisson model) before regressing the variables. Therefore, my steps are as follows;

1. Run a generalized Poisson model first, to get the vector of the fitted rates, or  $\lambda$
2. Find the auxiliary OLS regression on the data set, which gives a value of  $\alpha$ , and needs to be significant

3. Use that  $\alpha$  to fit the negative binomial regression model and make predictions.

I found the alpha value to be 0.059208, with a T-statistic score of 2.524969. When compared to the T-statistics distribution table, with DOF of 115 (Number of observations – 1) and 95% significance level, the T-value for a two-tailed test is 1.980808, which is less than the T-value of my found alpha, thus my alpha value is significant, and I can use the negative binomial distribution.

The negative binomial distribution follows the same principle as the Poisson distribution, where the coefficients are the log-likelihood (probability) of the count increasing due to the relationship with the independent variable.

## **Results**

In first looking at the goodness of fit for the model (Figure 8: Negative Binomial Regression Summary), the pseudo R<sup>2</sup> is 1, which is not necessarily meaningful, since the Pseudo R<sup>2</sup> is a measure trying to replicate the R<sup>2</sup> in OLS regression, and thus needs to be taken with a grain of salt. The Pearson Chi<sup>2</sup> is a measure of deviance and independence of the variables, and it is 321, with a value being closer to 0 the better. This is useful when comparing to other models for goodness of fit measure.

Let's now look at the coefficients, and which are significant. To determine that, its necessary to look at the p-value, and see which ones are less than 0.05. When looking at the p-values, if they are less than 0.05, then it is said that we are 95% confident that the coefficient falls along a range of values that do not include 0, which means that a relationship exists between the independent variable and the explanatory variable. If the co-efficient is 0, then it's an acceptance that no relationship exists between the independent variable, and the explanatory variable. Looking at the summary result, one can see significant variables exist. The months of March and February are statistically significant with p-values less than 0.05, and positive coefficients, which means python sightings have a higher probability of occurring in hexes during those months. Other variables with positive coefficients and p-values less than 0.05 are distance to distance to 1999 bird rookeries, income per hex, Mammal Habitats, and distance to urban boundaries, which means python sightings have a higher probability of occurring in hexagons that are further away from urban boundaries, 1999 bird rookery sites, and mammal habitats (i.e., distance increases). For income per hex, this means that python sightings have a higher probability of occurring in hexagons with a higher per hexagonal income density. Variables with negative coefficients that are statistically significant of p-values less than 0.05 are distance to bird sanctuaries, distance to shore birds, distance to roads, AADT per hex, distance to plover nests, distance to wading bird colonies, and distance to flowing water. For the distance variables, this means python sightings have a higher probability of occurring in hexagons that are closer in distance to those locations, such as bird sanctuaries, bird colonies, rivers, roads, etc. Interestingly, for AADT per hex, this means that python sightings have a higher probability of occurring in hexagons with a lower AADT value. Majority of the counties (except those with NaN coefficients) are statistically significant with p-values less than 0.05, with a mix of negative and positive coefficients.

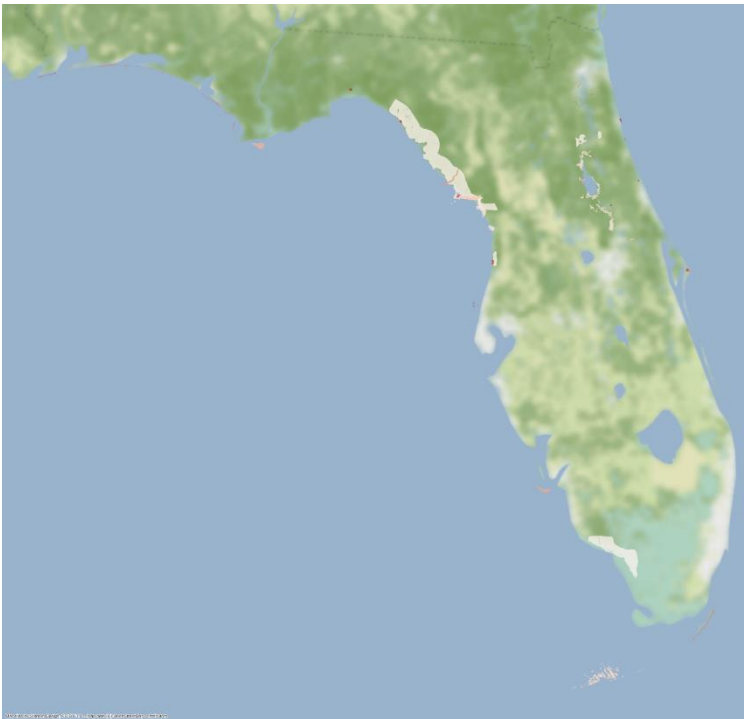
Figure 8: Negative Binomial Regression Summary

Generalized Linear Model Regression Results						
Dep. Variable:	pythonsPerHex	No. Observations:	116			
Model:	GLM	Df Residuals:	64			
Model Family:	NegativeBinomial	Df Model:	51			
Link Function:	Log	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-396.55			
Date:	Fri, 29 Apr 2022	Deviance:	297.26			
Time:	12:30:06	Pearson chi2:	329.			
No. Iterations:	19	Pseudo R-squ. (CS):	1.000			
Covariance Type:	nonrobust					
	coef	std err	z	P> z	[0.025	0.975]
April	0.0043	0.008	0.516	0.606	-0.012	0.021
December	0.0148	0.013	1.164	0.245	-0.010	0.040
February	0.0944	0.015	6.250	0.000	0.065	0.124
March	0.0599	0.022	2.707	0.007	0.017	0.103
NAME_Alachua	-0.0003	3.6e-05	-7.888	0.000	-0.000	-0.000
NAME_Baker	0.0002	2.54e-05	7.888	0.000	0.000	0.000
NAME_Bay	-0.0001	1.88e-05	-7.888	0.000	-0.000	-0.000
NAME_Bradford	-6.896e-05	8.74e-06	-7.888	0.000	-8.61e-05	-5.18e-05
NAME_Brevard	722.0045	91.381	7.901	0.000	542.900	901.109
NAME_Broward	706.1855	89.488	7.891	0.000	530.793	881.578
NAME_Calhoun	-3.144e-05	3.99e-06	-7.888	0.000	-3.93e-05	-2.36e-05
NAME_Charlotte	707.7680	90.002	7.864	0.000	531.368	884.168
NAME_Citrus	0.0003	3.27e-05	7.888	0.000	0.000	0.000
NAME_Clay	-0.0004	5.42e-05	-7.888	0.000	-0.001	-0.000
NAME_Collier	706.4850	89.485	7.895	0.000	531.098	881.872
NAME_Columbia	0.0018	0.000	7.888	0.000	0.001	0.002
NAME_DeSoto	0.0030	0.000	7.888	0.000	0.002	0.004
NAME_Dixie	-7.114e-07	9.02e-08	-7.889	0.000	-8.88e-07	-5.35e-07
NAME_Duval	-3.19e-07	4.04e-08	-7.888	0.000	-3.98e-07	-2.4e-07
NAME_Escambia	1.349e-07	1.71e-08	7.889	0.000	1.01e-07	1.68e-07
NAME_Flagler	-9.735e-08	1.23e-08	-7.888	0.000	-1.22e-07	-7.32e-08
NAME_Franklin	-3.868e-08	4.9e-09	-7.888	0.000	-4.83e-08	-2.91e-08
NAME_Gadsden	-1.53e-07	1.94e-08	-7.888	0.000	-1.91e-07	-1.15e-07
NAME_Gilchrist	-7.536e-08	9.55e-09	-7.888	0.000	-9.41e-08	-5.66e-08
NAME_Glades	709.7953	89.807	7.904	0.000	533.777	885.814
NAME_Gulf	-8.903e-08	1.13e-08	-7.888	0.000	-1.11e-07	-6.69e-08
NAME_Hamilton	1.856e-07	2.35e-08	7.888	0.000	1.39e-07	2.32e-07
NAME_Hardee	714.2807	90.563	7.887	0.000	536.780	891.782
NAME_Hendry	707.5156	89.712	7.887	0.000	531.684	883.347
NAME_Hernando	720.7121	91.381	7.887	0.000	541.609	899.816
NAME_Highlands	713.7619	90.345	7.900	0.000	536.689	890.835
NAME_Hillsborough	718.5606	91.069	7.890	0.000	540.069	897.052
NAME_Holmes	6.405e-09	8.12e-10	7.888	0.000	4.81e-09	8e-09
NAME_Indian_River	716.9050	90.804	7.895	0.000	538.933	894.877
NAME_Jackson	3.023e-11	3.83e-12	7.888	0.000	2.27e-11	3.77e-11
NAME_Jefferson	-3.551e-10	4.5e-11	-7.888	0.000	-4.43e-10	-2.67e-10
NAME_Lafayette	2.644e-10	3.35e-11	7.888	0.000	1.99e-10	3.3e-10
NAME_Lake	725.4454	91.624	7.918	0.000	545.866	905.025
NAME_Lee	705.5795	89.632	7.872	0.000	529.905	881.254
NAME_Leon	1.638e-11	2.08e-12	7.888	0.000	1.23e-11	2.04e-11
NAME_Levy	734.3963	92.825	7.912	0.000	552.462	916.330
NAME_Liberty	-4.017e-10	5.09e-11	-7.888	0.000	-5.02e-10	-3.02e-10
NAME_Madison	-1.104e-10	1.4e-11	-7.888	0.000	-1.38e-10	-8.3e-11
NAME_Manatee	712.0051	90.624	7.857	0.000	534.385	889.625
NAME_Marion	732.9907	92.474	7.926	0.000	551.745	914.237
NAME_Martin	709.7529	89.995	7.887	0.000	533.366	886.140
NAME_Miami_Dade	706.7782	89.409	7.905	0.000	531.539	882.018
NAME_Monroe	704.0096	89.447	7.871	0.000	528.696	879.323
NAME_Nassau	-1.203e-11	1.53e-12	-7.888	0.000	-1.5e-11	-9.04e-12
NAME_Okaloosa	3.633e-13	4.6e-14	7.891	0.000	2.73e-13	4.54e-13
NAME_Okeechobee	712.8410	90.086	7.913	0.000	536.275	889.407
NAME_Orange	719.5533	91.290	7.882	0.000	540.628	898.479
NAME_Osceola	718.7135	90.990	7.899	0.000	540.376	897.051
NAME_Palm_Beach	707.5910	89.670	7.891	0.000	531.841	883.341
NAME_Pasco	719.5940	91.146	7.895	0.000	540.951	898.237
NAME_Pinellas	711.6359	91.185	7.804	0.000	532.917	890.354
NAME_Polk	717.4389	90.581	7.920	0.000	539.903	894.975
NAME_Putnam	0	0	nan	nan	0	0
NAME_Santa_Rosa	0	0	nan	nan	0	0
NAME_Sarasota	708.5289	90.386	7.839	0.000	531.376	885.682
NAME_Seminole	718.4464	91.334	7.866	0.000	539.436	897.457
NAME_St_Johns	0	0	nan	nan	0	0
NAME_St_Lucie	711.3301	90.170	7.889	0.000	534.600	888.060
NAME_Sumter	0	0	nan	nan	0	0
NAME_Suwannee	0	0	nan	nan	0	0
NAME_Taylor	0	0	nan	nan	0	0
NAME_Union	0	0	nan	nan	0	0
NAME_Volusia	0	0	nan	nan	0	0
NAME_Wakulla	0	0	nan	nan	0	0
NAME_Walton	0	0	nan	nan	0	0
NAME_Washington	0	0	nan	nan	0	0
distanceToBirdSanctuaries	-4.784e-05	6.74e-06	-7.101	0.000	-6.1e-05	-3.46e-05
eagleNestsPerHex	-0.0437	0.031	-1.417	0.157	-0.104	0.017
distanceToEagleNests	6.353e-06	7.29e-06	0.871	0.384	-7.94e-06	2.06e-05
distanceToBirdNests2000	-1.185e-05	6.18e-06	-1.916	0.055	-2.4e-05	2.71e-07
distanceToMammalHabitats	2.247e-05	5.97e-06	3.764	0.000	1.08e-05	3.42e-05
distanceToShoreBirds	-7.329e-05	9.3e-06	-7.877	0.000	-9.15e-05	-5.51e-05
distanceToRoad	-5.524e-05	1.25e-05	-4.426	0.000	-7.97e-05	-3.08e-05
AADT_x	-1.938e-07	3.5e-08	-5.538	0.000	-2.62e-07	-1.25e-07
distanceToPloverNests	-8.795e-05	9.84e-06	-8.940	0.000	-0.000	-6.87e-05
ploverNestsPerHex	0.1107	0.038	2.876	0.004	0.035	0.186
distanceToWadingBirdColonies	-1.763e-05	7.23e-06	-2.438	0.015	-3.18e-05	-3.46e-06
distanceToUrbanBoundary	3.62e-05	7.24e-06	4.999	0.000	2.2e-05	5.04e-05
distanceToUrbanCenter	-4.74e-06	5.34e-06	-0.888	0.374	-1.52e-05	7.2e-06
distanceToLakeArea	-1.193e-05	6.39e-06	-1.869	0.062	-2.45e-05	5.83e-07
distanceToRivers	7.008e-08	6.07e-06	0.012	0.991	-1.18e-05	1.2e-05
distanceToFlowingWater	-3.739e-05	1.17e-05	-3.199	0.001	-6.03e-05	-1.45e-05
distanceTo1999WadingBirdRookeries	3.764e-05	1.5e-05	2.501	0.012	8.15e-06	6.71e-05
INCOME_Hex	0.0007	0.000	4.596	0.000	0.000	0.001
POP_Hex	8.005e-07	2.84e-06	0.282	0.778	-4.77e-06	6.37e-06

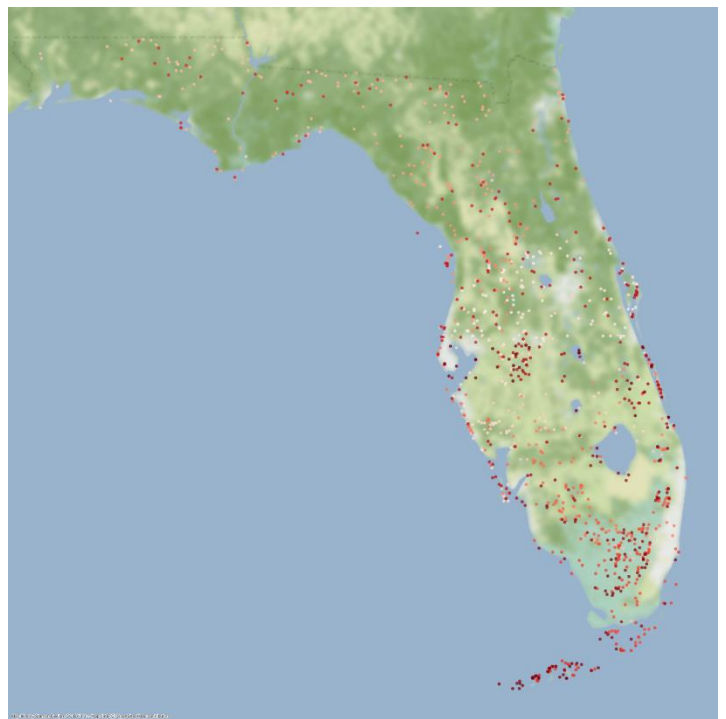
## Discussion

Let's start with variables that were found to have statistically significant positive coefficients. The months of March and February tie into the start of python breeding season, which could mean they are more active in finding mates and breeding grounds, which could lead to more sightings due to more activity. I would recommend Environmental DNA sampling to occur the most in those two months, as there exists that relationship of increased python counts. Additional research can be done to assess any changes in temperature or weather that correlates with these months that can help explain why February and March is significant. The positive coefficient of distance to 1999 bird rookeries is interesting, but perhaps this can be explained by how old the data is, being more than 20 years old of these sites. It would be telling to find out if these sites are still active, or perhaps have already been wiped out by pythons in the past twenty years up to this point? Its speculation until further research yields more. The positive coefficient with distance to urban boundaries make sense as Bartoszek et Al found in their results Burmese pythons tend to avoid urban areas, but the positive coefficient with mammal habitats is interesting, contradicting my inference that mammals would attract pythons due to being prey. Perhaps again, with those earlier mentions of declining mammal observations, Pythons are moving away from mammalian diets since there are less opportunities? Figure 9 shows the mapping of Mammal Habitats and Figure 10 shows the mapping of 1999 rookeries. Looking at the dispersion of the two maps, it doesn't seem like its due to cluster away from where the pythons are sighted (one could make the argument for mammal habitats up north, but they are distributed in the everglades in the western area as well), so I can't make inferences if its due to just being in areas pythons are not sighted. I would recommend for Environmental DNA sampling to not use mammal habitats, or these 1999 Bird rookery sites are places to sample.

**Figure 8: Mammal Habitats**



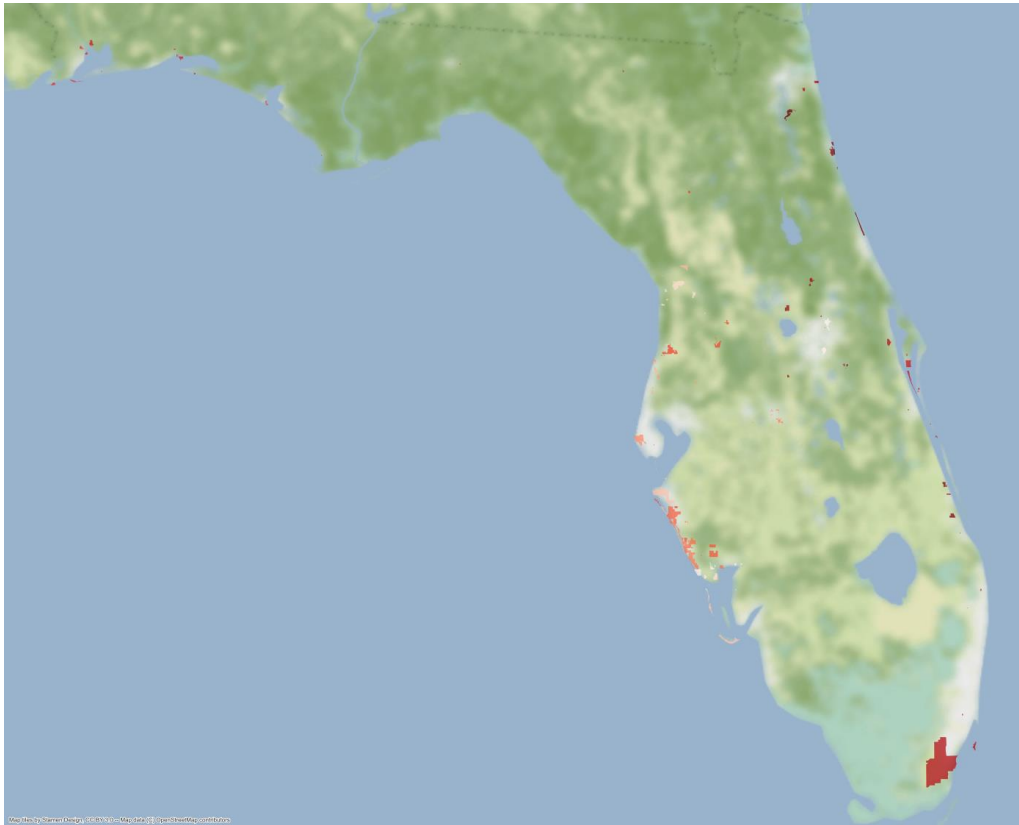
**Figure 9: 1999 Bird Rookeries**



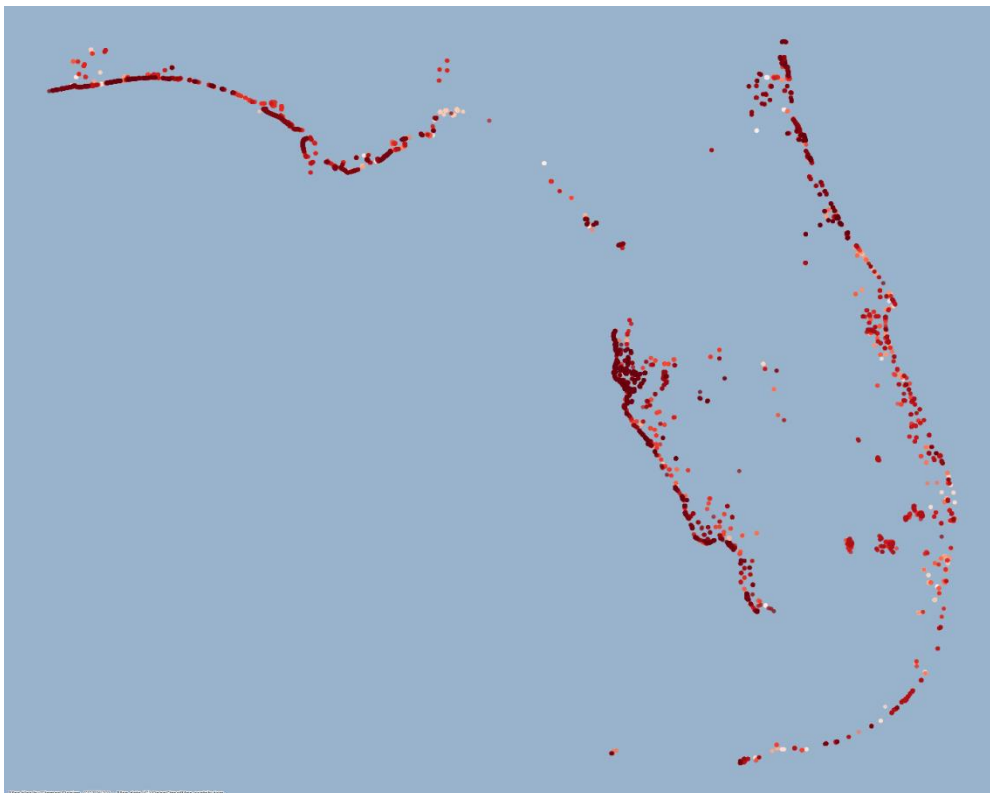
For income hex, I thought hexagons with lower income levels would tend to attract more python hunters for the bounties, but hexagons with higher income level had higher probability of having higher python counts. This could be potentially due to higher income levels means more human population or activity, and so more chances of being spotted. Finally, numerous counties had positive coefficients, and the probability of python sightings increase in those counties, but that could be due to if those counties have higher human population levels and therefore would naturally have more sightings. Until more data and modeling are done to account for human presence in sightings, I would not make county inferences from the data yet.

For the variables with negative coefficients that were statistically significant, such as distance to bird sanctuaries, distance to shore birds, distance to plover nests, distance to wading bird colonies, and distance to flowing water, I would recommend environmental DNA sampling to occur in areas that are closer to those features, which sounds obvious. But if you look at Bird Sanctuaries for instance, figure 10, it does cluster in patterns close to where Pythons are spotted. It's a large part of southern Florida and knowing there is at least a relationship there can help with pinpointing Environmental DNA sampling areas. Bird Sanctuaries also extend upwards in parts of Northern Florida, so Environmental DNA sampling should occur close to those areas to assess spread of the pythons up north as well. Figure 11 is the distribution of shorebirds, and Figure 12 is the distribution of wading birds. I would recommend environmental sampling to occur at the wading bird points, to help locate and identify pythons in the areas, and the same for the Shorebirds. However, the value in the shorebirds is that they extend up north and sampling up there can help monitor again if the pythons are spreading up north. Finally, the distribution of Snow Plovers, shown in Figure 13, is concentrated in Northern Florida, so its interesting that it is statistically significant. However, perhaps this variable controls for the spread of the pythons up north, and since it is significant, shows that pythons are expanding past the main range of southern Florida, and slowly making their way up north.

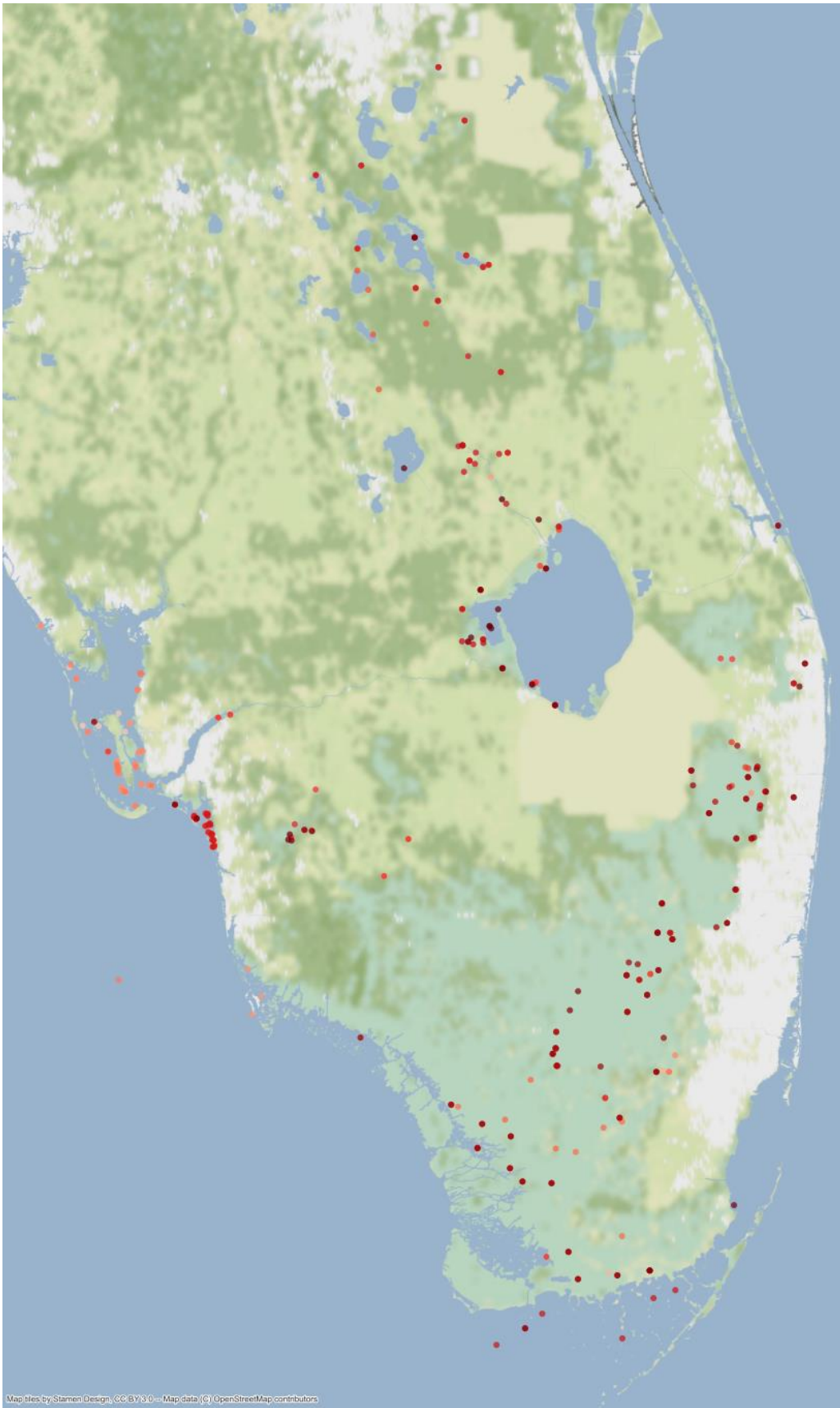
**Figure 10: Bird Sanctuaries**



**Figure 11: Shorebirds Distribution**



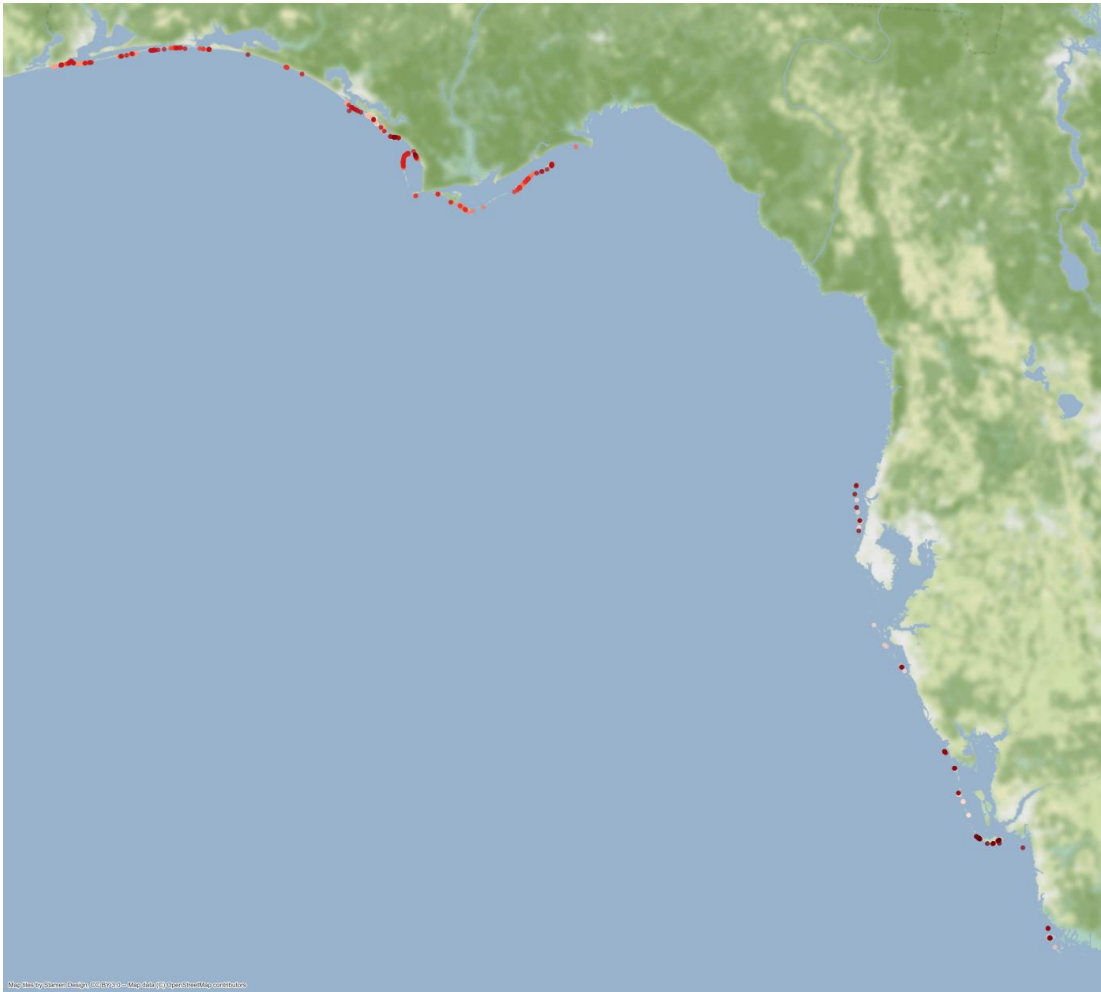
**Figure 11: Wading Bird Colonies**



Map tiles by Stamen Design, CC-BY, 3.0 - Map data (C) OpenStreetMap contributors



**Figure 12: Snowy Plover Nests**

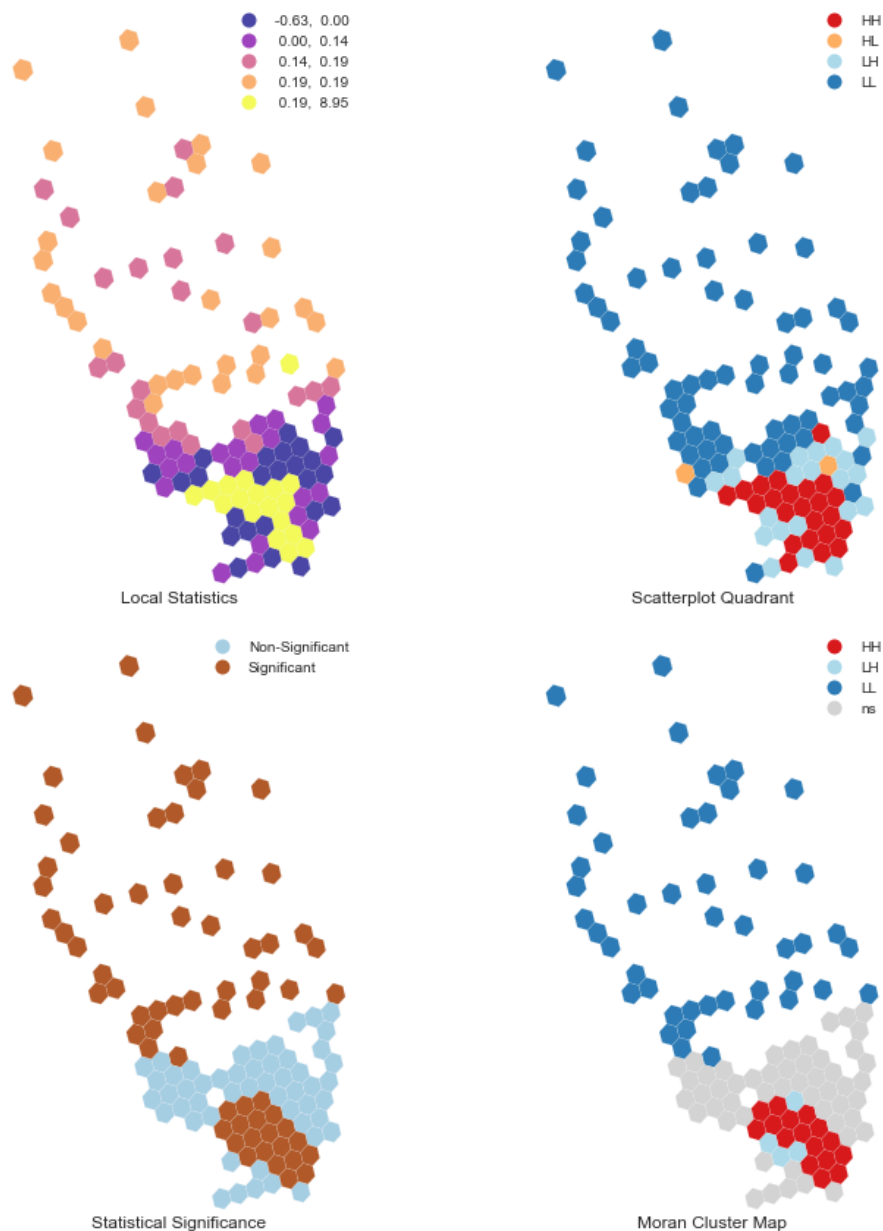


Finally, distance to roads had a negative coefficient that was statistically significant, meaning as hexagons had a closer distance to roads, the probability of the count of pythons in the hexagon went up. Conversely, if a hexagon had a high AADT rate, then the probability of having more pythons in the hexagon goes down. This makes sense, in the way that there are pythons being sighted along roads, as a circumstance of a driver seeing them on the or next to the road. So, there are more chances of finding the pythons along the roads, but on roads that do not have higher traffic volume. Pythons tend to avoid urban areas and that is where traffic is the densest, so it logical that areas with more traffic density are places pythons avoid. Python sightings on roads probably happen on highways or county roads through wildlife or non-urban areas, and those roads should be priority for environmental DNA sampling sites.

## Spatial Autocorrelation Assessment

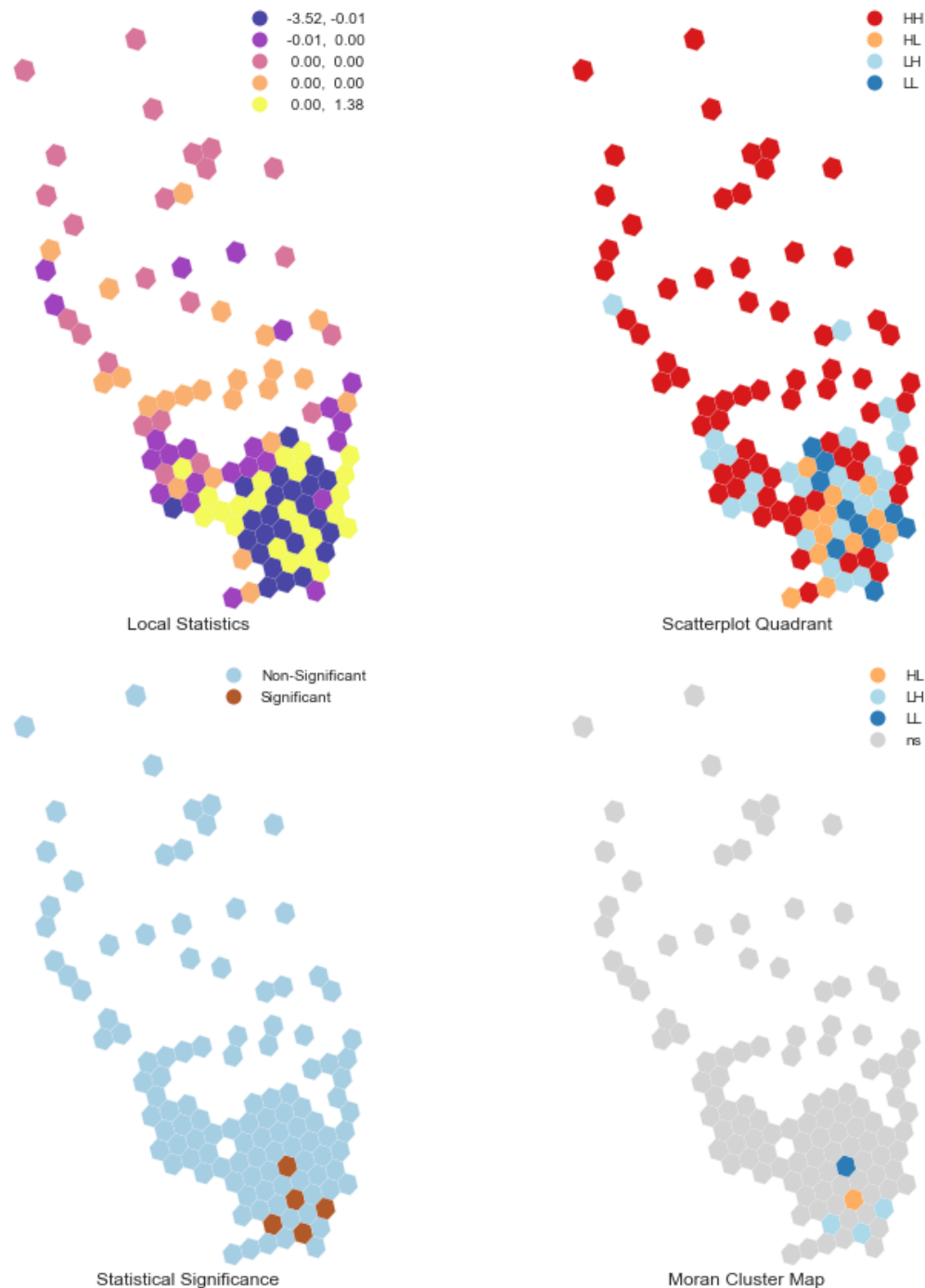
Since this is a model looking at the spatial relationship amongst Burmese Pythons, I need to factor and assess spatial autocorrelation in my model. Spatial Autocorrelation indicates that there is clustering in a map, in essentially following Tobler's law that near things are more related than distant things. For a regression model, this can present an issue when looking at residuals and predictions. If spatial distribution is not captured correctly in the model, then spatial autocorrelation will also bleed into the residuals and errors of predictions. In other words, you want your data to have spatial pattern and clustering, but you don't want that same pattern to appear in the residuals, otherwise it means the spatial relationship hasn't been fully accounted for. To assess spatial autocorrelation, we find the Local Moran's I, i.e., calculate the LISA statistics. I first performed this for the regular pythonsPerHex count, along with calculating the weight of its dispersion (i.e., relationship to its neighbors), shown in Figure 13.

**Figure 13: Moran's I of Python Count pr Hex**



Looking at the top left plot, there is clustering of similar values; the top Right plot illustrates High-High cluster, high-low outlier, low-low cluster, and non-significant areas; the bottom left is the statistical significance of spatial location, and the bottom left is the Moran Cluster map. Essentially the takeaway from here is that there is clustering and patterns in the spatial distribution of the python count per hex. Now let's look at the spatial distribution of the residuals, shown in Figure 14. As evident, there is less clustering and more random pattern. And when looking at the significance of the Moran's I, there is less than the distribution of the pythons per Hex. Some areas are still significant, and I think the model and resolution needs to go deeper and finer to assess spatial autocorrelation in those areas. I only used H3 resolution of 5, and to focus on a more detailed area of Florida, I would need to go to a higher resolution.

**Figure 14: Moran's I of Residuals per Hex**



However, my model, for the most part, has considered spatial autocorrelation, and it is not as significant in the errors as it is in the actual python distribution.

## **Conclusion**

While this model provided statistical significance in on numerous explanatory variables, I recommend for Environmental DNA sample sites to prioritize areas and regions close to wading bird colonies as the initial first step. This feature seems to address both the statistical significance of flowing water sites, for wading birds would be in flowing water, along with being a prey animal for Burmese Pythons. Furthermore, based on this model, I recommend focusing sampling actions specifically during the months of February and March. If Environmental DNA sampling higher than normal results from those areas during those months for Burmese Python DNA, then I would recommend expanding sampling to sites closer to shorebird distribution, bird sanctuaries, and more bodies of flowing water.

For future next steps on modelling, I want to include specific temperature and elevation data, tied to a certain time to year, to see if ground temperature, solar radiation, and vegetation levels have a relationship to where pythons are, in addition to elevation. Another dataset I would like to possibly include are identified breeding ground locations and nests, if recorded. I also want to try and increase the resolution dimensions of the study, to try and have a finer tuned hexagonal area of study, to try and make more detailed relationships.

## **References**

- Dove, CJ, Snow, RW, Rochford MR, Mazzotti, FJ. (2011) Birds Consumed by the Invasive Burmese Python (*Python molurus bivittatus*) in Everglades National Park, Florida, USA. *The Wilson Journal of Ornithology*. 123 (1): 126–131.
- Hunter ME, Oyler-McCance SJ, Dorazio RM, Fike JA, Smith BJ, Hunter CT, et al. (2015) Environmental DNA (eDNA) Sampling Improves Occurrence and Detection Estimates of Invasive Burmese Pythons. *PLoS ONE* 10(4): e0121655. <https://doi.org/10.1371/journal.pone.0121655>
- Sarill, M. (2016). Burmese pythons in the everglades: International & Executive Programs: UC Berkeley. Burmese Pythons in the Everglades | International & Executive Programs | UC Berkeley. Retrieved April 29, 2022, from <https://iep.berkeley.edu/content/burmese-pythons-everglades>

## **References for modelling**

- Rey, S., Arribas-Bel, D. and Wolf, L., 2020. *Geographic Data Science with Python*. 1st ed.
- Date, S. *Time Series Analysis, Regression and Forecasting*. Time Series Reasoning. <https://timeseriesreasoning.com/contents/negative-binomial-regression-model/>