Erin Witt
December 2023

Final Project: Reptile Radar

For my final assignment I developed Reptile Radar, a database of sighting data for different kinds of lizards, snakes and turtles. The database allows users to query information based on their coordinates, generating a paragraph detailing the types of animals sighted, along with information about them.

To construct this database, I used data from the Global Biodiversity Information Facility (gbif.org) and employed the Geospatial Processing tools available in Pgadmin 4. By analyzing sighting data for various reptile species, I aimed to create a comprehensive database offering information about animals in the user's vicinity. All the original tables that I used had the same columns and data types which meant that it was easy to join several tables together at a time. Its purpose is to use SQL tables to create user-friendly outputs. While a table may serve the needs of some, a paragraph describing potential encounters during a hike could be more useful for a broader audience. Despite the relatively small scale of this database (580,000 rows), compared to the extensive reptile sighting data available on GBIF, I focused on species commonly found in North America.

A kingsnake, (also known as a milk snake), one of the species featured in the Reptile Radar database. San Diego Zoo Wildlife Alliance, 2023

The construction process involved making a table and importing all CSV files into it. Additionally, I made a separate table containing facts about each animal in the larger database. This table, with only 11 rows, was designed to be concise, avoiding unnecessary length while still providing essential information. This approach allowed me to store pertinent details without duplicating information for each sighting entry.

	snakeid numeric	animalname text	color text	size text	fact text	active text	dangerous boolean
1	1	Garter Snake	black and yellow	51-76.2	From birth they are...	during the day	false
2	2	King Snake	red and black	96-154	Kingsnakes are co...	during the day	false
3	3	Rat Snake	black and white	91-152.4	They are exception...	during the day	false
4	4	Skink	brown, gray, and orange	7.5-81	Their babies are ca...	during the day	false
5	5	Canebrake	gray, pink or even yellow	76-154	When encountered ...	during the day and night	true
6	6	Massauga	gray or light brown	60.96-91.44	People are rarely bi...	during the day	true
7	7	Copperhead	brown or yellow	60.96-91.44	They give no warni...	during the day and night	true
8	8	Brown Anole	brown	12.7-22.86	These lizards have ...	during the day	false
9	9	Green Anole	green	20.32	In the wild they are ...	during the day	false
10	10	Snapping Turtle	brown	40-70	To hunt fish, they s...	during the day	true
11	11	Red-Eared Slider	green with red ears	12.7-20.32	They are a very po...	during the day	false

Total rows: 11 of 11 Query complete 00:00:00.116

In my "allreptiles" table, each entry included a unique "snakeid," geographic coordinates specified in SRID 4326, and other details about each sighting. To avoid redundancy, I did not store comprehensive data about each individual species within this table. Instead, I made a system where a specific ID number associated with each entry could be cross-referenced with the "snakeinfo" table when constructing sentences or retrieving additional details. This approach allowed for a more streamlined database structure, as the core information was consolidated in "allreptiles," and species-specific details were efficiently linked through the shared ID numbers in the "snakeinfo" table. This design not only optimized storage but also facilitated the retrieval of data about each species as needed during sentence formation.

	gbifid numeric	occurrenceid character varying (5000)	countrycode text	scientificname text
1	931030813	urn:catalog:CAS:HERP:255391	US	Agkistrodon contortrix (Linnaeus, 1766)
2	931030802	urn:catalog:CAS:HERP:255389	US	Agkistrodon contortrix (Linnaeus, 1766)
3	931030799	urn:catalog:CAS:HERP:255388	US	Agkistrodon contortrix (Linnaeus, 1766)

Total rows: 582524 of 582524 Query complete 00:00:02.680 Ln 147, Col 1

In summary, the described approach optimizes the database by minimizing redundant queries to the original dataset, utilizing temporary tables for attribute assignment, and implementing indexing for faster spatial searches. These optimizations collectively contribute to a more efficient and responsive system for providing information about reptile sightings based on user input coordinates.

The following report is a more detailed explanation of how I created this database and some insight into how it was optimized and how it generates its outputs.

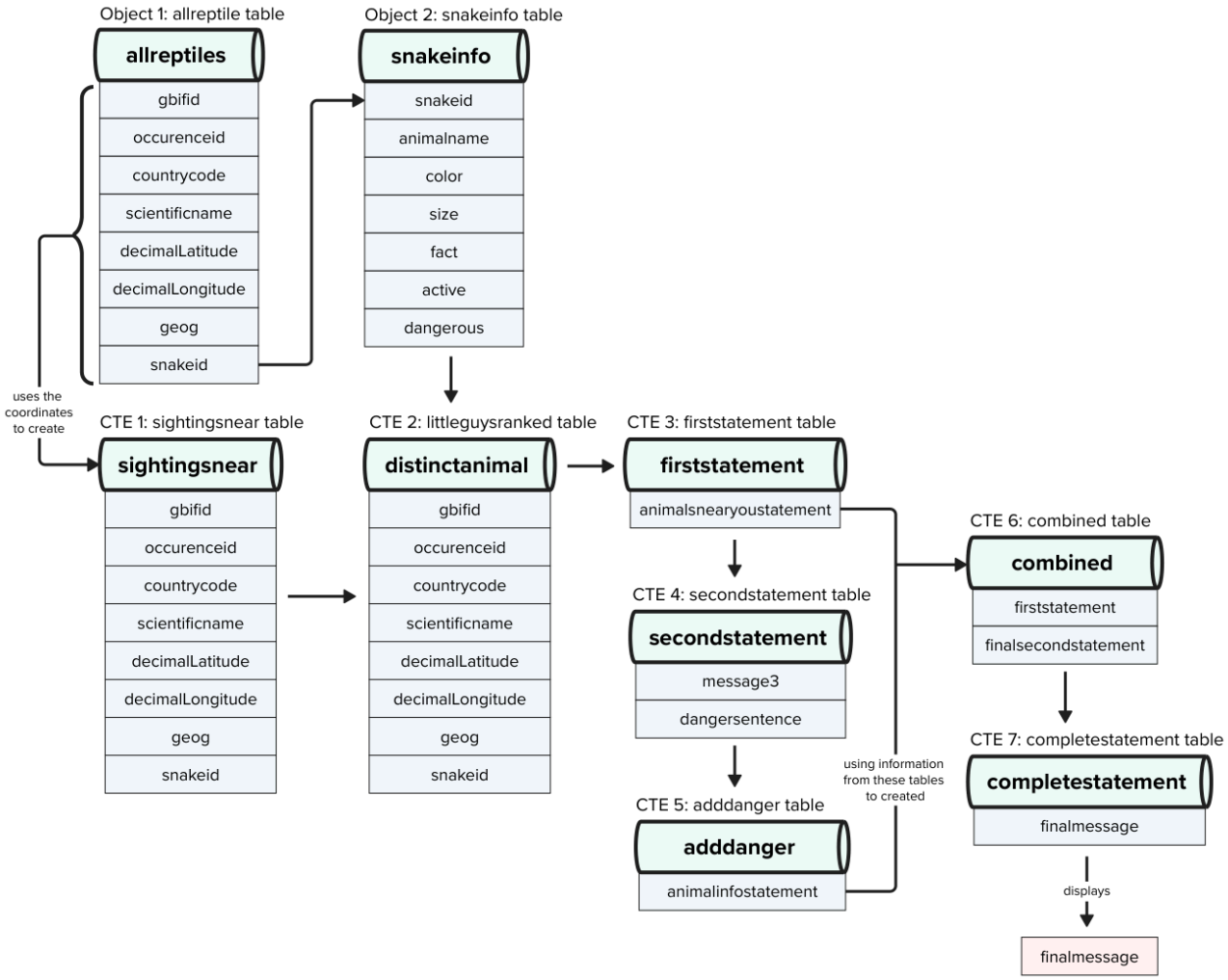
Table of Information about the Reptiles

All of this information is stored in a table called “snakeinfo”. It uses the “SnakeID” column to cross reference the list of all animal sightings that is spit out when you insert the coordinates. This information is used to concat sentences about the animals that are found nearby.

Snake ID	Name	D	Color	Length (cm)	Fact	When active
1	Garter Snake	N	Black and Yellow	51-76.2 cm	From birth they are able to hunt by themselves.	Day
2	King Snake	N	Red and Black	96-154 cm	Kingsnakes are constrictors, meaning they hunt by restricting bloodflow to the brain of their prey.	Day
3	Ratsnake	N	Black and white	91 - 152.4 cm	They are exceptionally good at climbing trees.	At day or night, depending on the season.
4	Skinks	N	Brown Grey Orange	7.5 - 81 cm	Their babies are called skinklets!	During the day
5	Canebrakes	Y	Grey or even pink and yellow	76-152 cm	When encountered in the wild, they often remain docile.	during both day and night.
6	Massasaugas	Y	Gray or light brown	60.96 - 91.44 cm	People are rarely bitten by them because they are shy.	during the daytime.
7	Copperheads	Y	Brown and yellow	60.96 - 91.44 cm	They give no warning and can strike instantly.	They are active during the day.
8	Brown Anoles	N	Brown	12.7 - 22.86 cm	These lizards have a dewlap, an expandable flap of skin under their chin. They use it to scare away predators.	They are active during the day.
9	Green Anoles	N	Green	20.32 cm	In the wild they are skittish and nervous but if handled enough by humans, they become more tame.	They are active during the day.
10	Snapping Turtle	Y	Brown	40–70 cm	To hunt fish, they stay underwater and keep their mouths open.	They are active during the day.
11	Red-eared Sliders	N	Green with red ears	12.7 - 20.32 cm	They are a very popular pet but are also invasive for that reason.	They are active during the day.

Object Relational Diagram / Workflow Diagram

Here is a diagram of how my main query searches the database and provides information to the user. The process starts at the allreptiles table and ends at the finalstatement box.



Final Code

Here is a detailed breakdown of how I made this work.

```
-Created a table to input the csv files from GBIF.
CREATE TABLE gartersnake (
gbifID numeric,
datasetKey text,
occurrenceID text,
kingdom text,
phylum text,
…
establishmentMeans text,
lastInterpreted text,
mediaType text,
issue text )

-I copied the csvs into the tables I made.
COPY gartersnake FROM
'C:\Users\erins\Desktop\0024257-231120084113126\0024257-23112008
4113126.csv' WITH CSV HEADER DELIMITER '     ';

-From what I imported, I then cut down the tables to save
storage. I did this for each animal, I just won't write it all
out here.
CREATE TABLE gartersnake2
SELECT
     gbifid,
     occurrenceid,
     scientificName,
     decimalLatitude,
     decimalLongitude,
     year
FROM gartersnake
WHERE decimallatitude IS NOT NULL
```

```sql
--I created the snakeinfo table and put in all the information
about the reptiles. I also added the snakeID column.
CREATE TABLE snakeinfo
ADD COLUMN snakeid numeric,
ADD COLUMN animalname text,
ADD COLUMN color text,
ADD COLUMN size text,
ADD COLUMN fact text,
ADD COLUMN active text,
ADD COLUMN dangerous boolean,
ADD COLUMN geog geography(POINT, 4326);

--I added all the information about each individual animal.
INSERT INTO snakeinfo
     (snakeid, animalname, color, size, fact, active, dangerous)
VALUES
     (1, 'Garter Snake', 'black and yellow', '51-76.2', 'fact',
     FALSE),
     (2, 'King Snake','red and black', '96-154', 'fact', 'during the
     day', FALSE),
     (3, 'Rat Snake', 'black and white', '91-152.4', 'fact', 'during
     the day', FALSE),
     (4, 'Skink', 'brown, gray, and orange', '7.5-81', 'fact', FALSE),
     (5, 'Canebrake', 'gray, pink or even yellow', '76-154', 'fact',
     'during the day and night', TRUE),
     (6, 'Massauga', 'gray or light brown', '60.96-91.44', 'fact',
     'during the day', TRUE),
     (7, 'Copperhead', 'brown or yellow', '60.96-91.44', 'fact',
     'during the day and night', TRUE),
     (8, 'Brown Anole', 'brown', '12.7-22.86', 'fact', 'during the
     day', FALSE),
     (9, 'Green Anole', 'green', '20.32', 'fact', 'during the day',
     FALSE),
     (10, 'Snapping Turtle', 'brown', '40-70', 'fact', 'during the
     day', TRUE),
     (11, 'Red-Eared Slider', 'green with red ears', '12.7-20.32',
     'fact', 'during the day', FALSE)
```

```
--Then I combined all the individual tables into one big one.
CREATE TABLE allreptiles
AS SELECT * FROM gartersnake2
     UNION ALL
SELECT * FROM kingsnake2
     UNION ALL
SELECT * FROM ratsnake2
     UNION ALL
SELECT * FROM skinks2
     UNION ALL
SELECT * FROM canebrakes2
     UNION ALL
SELECT * FROM massasaugas2
     UNION ALL
SELECT * FROM copperheads2
     UNION ALL
SELECT * FROM brownanoles2
     UNION ALL
SELECT * FROM greenanoles2
     UNION ALL
SELECT * FROM snappingturtles2
     UNION ALL
SELECT * FROM redearedsliders2

--I created the points from text for each sighting and created an
index for the entire table.
UPDATE allreptiles
     SET geog = ST_GeogFromText('POINT(' || decimalLongitude ||
     ' ' || decimalLatitude ||')');

CREATE INDEX allreptilesindex
     ON allreptiles
     USING GIST(geog);
```

-Then I verified that they all had the correct SRID.

```
SELECT DISTINCT ST_SRID(geog)
FROM allreptiles
```

	st_srid integer
1	4326

Total rows: 1 of 1

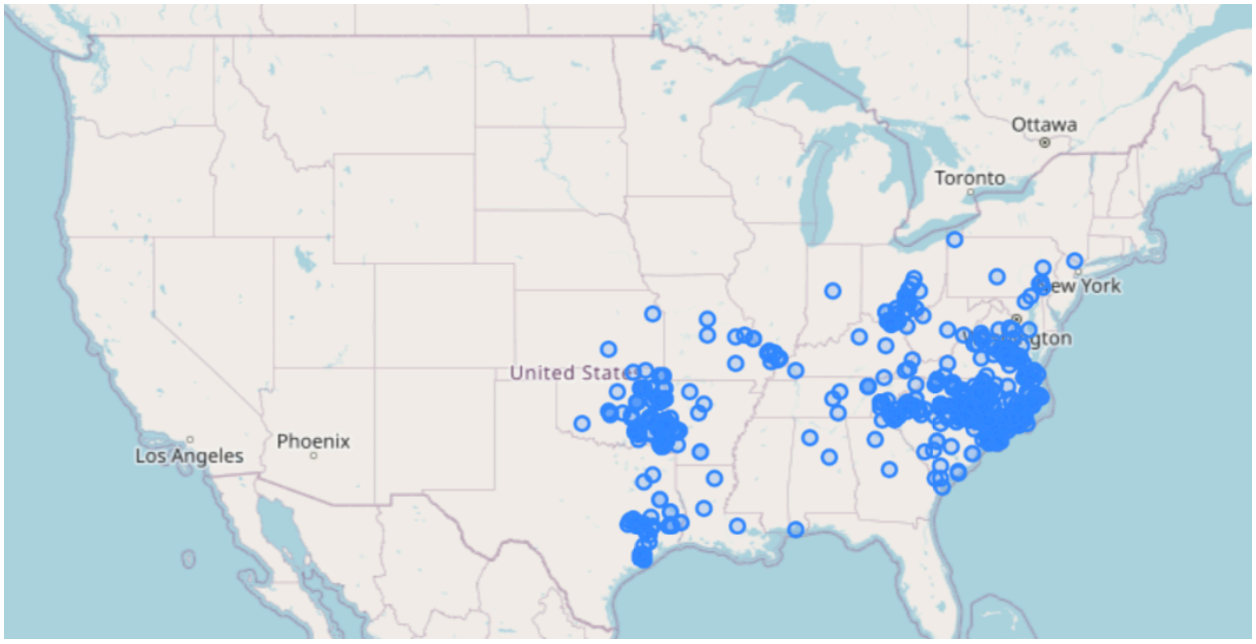
-Then I verified that all my points were valid. This showed that there were no geography points that were invalid.

```
SELECT ST_IsValid(geog::geometry) AS valid
     FROM allreptiles WHERE ST_IsValid(geog::geometry) = FALSE
```

	valid boolean

Total rows: 0 of 0

-Here is what the map viewer looked like while displaying 1000 out of 582524 rows of the allreptiles dataset.



```
--Finally, I made a query that uses a coordinate, creates a list
of all sightings within 10000 meters of the person and puts them
into a list called "sightingsnear". Then, distinct rows are
picked and given traits based on their snakeid and put into CTE
"distinctanimals". A statement is made based on the contents of
"distinctanimals" this makes a list of all the different kinds
of animals nearby. Using "distinctanimals", a second statement
is created using the information that was appended from the
"snakeinfo" table. The danger statement is appended and the
"finalsecondstatement" is created. They are combined in a CTE
called "combined" and the finalmessage is displayed.
WITH sightingsnear AS (
     SELECT *
     FROM allreptiles
     WHERE ST_DWithin(
        geog,
        ST_SetSRID(ST_MakePoint(-122.409463, 37.559172), 4326),
        10000
    )
),
distinctanimals AS (
     SELECT
        snakeinfo.animalname,
          snakeinfo.color,
          snakeinfo.size,
          snakeinfo.fact,
          snakeinfo.active,
          snakeinfo.dangerous,
        ROW_NUMBER() OVER () AS ranked
     FROM
        sightingsnear
     JOIN snakeinfo ON sightingsnear.snakeid = snakeinfo.snakeid
     GROUP BY DISTINCT snakeinfo.animalname, snakeinfo.color,
     snakeinfo.size, snakeinfo.fact, snakeinfo.active,
     snakeinfo.dangerous
```

```
),
firststatement AS (
    SELECT
          CASE
          WHEN COUNT(*) = 1 THEN
          CONCAT('There are ', MAX(animalname), 's near you.')
          WHEN COUNT(*) = 2 THEN
          CONCAT('There are ', MAX(animalname), 's and  ',
          MIN(animalname), 's near you.')
          WHEN COUNT(*) >= 3 THEN
          CONCAT('There are ',
          STRING_AGG(animalname, 's, ' ORDER BY ranked), 's,
          ','near you.')
               ELSE 'There are no recorded animal sightings near
               you.'
          END AS animalsnearyoustatement
     FROM distinctanimals
),
secondstatement AS (
     SELECT
          animalname,
          ROW_NUMBER() OVER () AS ranked,
          CONCAT('The ', animalname, ' is ', color, ' in color
          and can grow up to ', size, ' centimeters long. They
          are active ', active, '. ', fact) as message3,
               CASE
          WHEN dangerous = TRUE THEN 'These animals are
          dangerous. If you see one in the wild, keep a good
          distance.'
               END AS dangersentence
     FROM distinctanimals
),
adddanger AS (
     SELECT
```

```
          STRING_AGG((message3 || CASE WHEN dangersentence IS
          NOT NULL THEN ' ' || dangersentence ELSE '' END), ' ')
          AS animalinfostatement
     FROM secondstatement
),
combined AS (
     SELECT animalsnearyoustatement FROM firststatement
          UNION ALL
     SELECT animalinfostatement FROM adddanger
),
completestatement AS (
     SELECT
     STRING_AGG(animalsnearyoustatement, ' ') AS finalmessage
     FROM combined
)
SELECT finalmessage FROM completestatement
```

FINAL OUTPUT FROM THIS STATEMENT:
"There are Garter Snakes, Rat Snakes, Skinks, Snapping Turtles, near you.

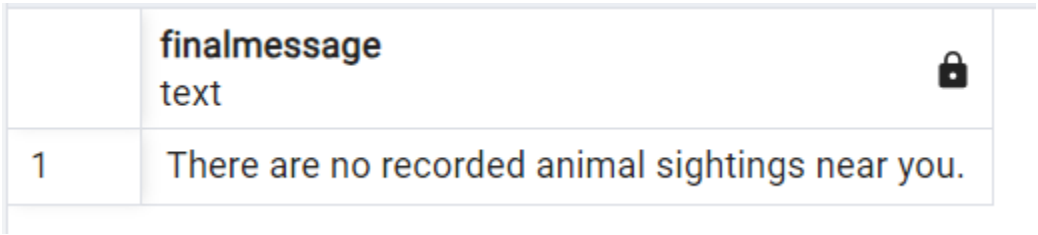
The Garter Snake is brown with a yellow stripe down their back in color and can grow up to 51-76.2 centimeters long. They are active during the day time. From birth they are able to hunt by themselves.

The Rat Snake is black and white in color and can grow up to 91-152.4 centimeters long. They are active during the day time. They are exceptionally good at climbing trees.

The Skink is brown, grey, and orange in color and can grow up to 7.5-81 centimeters long. They are active during the day time. Their babies are called skinklets!

The Snapping Turtle is brown in color and can grow up to 40-70 centimeters long. They are active during the day. To hunt fish, they stay underwater and keep their mouths open. These animals are dangerous. If you see one in the wild, keep a good distance."

IF THERE ARE NO RECORDED SIGHTINGS IT SAYS:



Data Dictionary

Data Dictionary for Reptile Radar

Tables

'allreptiles'

- Description: This table has 582524 rows worth of sighting information for various types of reptiles and animals all over the world from GBIF.com.
- 'Gbifid' (numeric) Unique identifier.
- 'decimalLatitude' (numeric) Latitude value.
- 'decimalLongitude' (numeric) Longitude value.
- 'geog' (geography) A geography column created using its long and lat from the other columns. In 4326 SRID.
- 'snakeid' (numeric) Each species has a unique id number to cross reference it to the snakeinfo table.

*** ‘snakeinfo’
  - Description: This table has 11 rows and has a unique number for each type of snake and their corresponding information which is used to form the final sentences.
  - ‘snakeid’ (text) Each species has a unique id number to cross reference it to the allreptiles table.
  - ‘animalname’ (text) The name of each species of animal.
  - ‘color’ (text) The common coloring of each animal.
  - ‘size’ (text) An average length.
  - ‘fact’ (text) A short fun fact about each animal.
  - ‘active’ (text) What time of day they are most active at.
  - ‘dangerous’ (text) Whether or not the animal is dangerous to humans.

Indexes

‘allreptileindex’
  - Description: A spatial index created on the ‘geog’ column which was created using the decimalLat and decimalLong column for each animal in the table.
  - **Type:** GiST.

Functions

ST_IsValid
  - Description: Was used to test the validity of the points before they were used in the database.
  - Parameters:
    - ‘geog’ (geography) Column I checked the validity of.

ST_DWithin
  - Description: Used to compare the user input coordinates and the sighting data in the allreptiles table.
  - Parameters:
    - ‘

ST_MakePoint

- Description: Used to turn the user's input coordinates into a point that can be used in the ST_DWithin function and used to compare the two.
- Parameters:
    - 'Lat' (numeric): user input first point
    - 'Long' (numeric): user input second point

Datasources

Common Garter Snake
Thamnophis Fitzinger, 1843 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

King Snakes or Milk Snakes
Lampropeltis Fitzinger, 1843 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Ratsnakes Dataset
Pantherophis Fitzinger, 1843 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Skink Dataset
Plestiodon Duméril & Bibron, 1839 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Canebrake Rattlesnake
Crotalus horridus Linnaeus, 1758 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Massasauga Rattlesnake

Sistrurus catenatus (Rafinesque, 1818) in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Copperhead Rattlesnake
Agkistrodon contortrix (Linnaeus, 1766) in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Brown Anole
Anolis sagrei Duméril & Bibron, 1837 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Green Anole
Anolis carolinensis Voigt, 1832 in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

Snapping Turtles
Chelydra serpentina (Linnaeus, 1758) in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2023-12-12.

National Park Service. (2022 March 25). *Garter Snake Fast Facts.* Retrieved December 10th from https://www.nps.gov/articles/000/garter-snake-fast-facts.htm

Szalay, J., (2016 Feburary 29) Kingsnake Facts. Retrieved December 10th from https://www.livescience.com/53890-kingsnake.html

Greer, H., (2022). *Pantherophis obsoletus Eastern Rat Snake.* Animal Diversity Web. Retrieved December 10th from https://animaldiversity.org/accounts/Pantherophis_obsoletus/#economic_importance_positive

Hall, H., (2023 March 29). *Skink Lizard.* A-Z Animals. Retrieved December 10th from https://a-z-animals.com/animals/skink-lizard/

Wilson, J.D., (2005). *Canebrake / Timber Rattlesnake (Crotalus horridus) - Venomous.* Retrieved December 10th from https://srelherp.uga.edu/snakes/crohor.htm

Biggs. B., (2022 31 July). *Copperhead snakes: Facts, bites & babies.* Live Science. Retrieved December 10th from https://www.livescience.com/43641-copperhead-snake.html

Vallie. S., (2022 24 November). *What to Know About Brown Anoles.* WebMD. Retrieved December 10th from https://www.webmd.com/pets/what-to-know-about-brown-anoles

McLeod. L., (2021 3 March). *Green Anole: Species Profile Characteristics, Housing, Diet, and Other Information.* The Spruce Pets. Retrieved December 13th from https://www.thesprucepets.com/green-anoles-pets-1236900

The Editors of Encyclopedia Britanica. (2023 3 May). *snapping turtle*. Encyclopedia Britanica. Retrieved December 13th from https://www.britannica.com/animal/snapping-turtle

Pace. L., (2022 21 July). Red-Eared Slider Trachemys scripta elegans. A-Z Animals. Retrieved December 13th from https://a-z-animals.com/animals/red-eared-slider/