Process Book

Initial Project Proposal

Basic Information:

Project name: Bird Tracker
Group members: Sakshi Jain (u1464049@umail.utah.edu), Sujit Kumar Kamaraj (u1466410@umail.utah.edu), Tanmay Sharma (tanmay.sharma@utah.edu , u1472860)

GitHub Link: https://github.com/orgs/dataviscourse2024/teams/bird-tracker/repositories

Background and Motivation: Recently, we found ourselves engrossed in bird watching, which inspired us to track different bird species across the United States. The project aims to create an interactive tool where users can search for a specific bird and see all the states where it is present. Additionally, by hovering over a state, users

will get information about the birds in that state, including their count and critical features. This project aims to enhance understanding of bird populations, migration patterns, and biodiversity across different regions. By making this data easily accessible and visually engaging, we hope to promote conservation efforts, educate users about the richness of bird species in the U.S., and contribute to the growing need for user-friendly tools in environmental sciences.

Project Objectives:

Build an interactive visualization tool that allows users to search for specific bird species and view the states where those species are present.

Enable state-level exploration by displaying the number of bird species in a selected state and critical features about each species (such as migratory or endangered status).

Build the first steps required to build a real-time bird-tracking visualization tool.

Promote understanding and awareness of bird biodiversity by creating a user-friendly interface for educational and research purposes.

Integrate hover-over functionality that offers detailed insights on each state, including the count of birds and their unique characteristics, to enhance user interaction and data accessibility.

Data:

We are initially going to use data that has been collected via the Monitoring Avian Productivity and Survivorship (MAPS) program. https://ibp-maps-data-exploration-tool.org/app/maps

We are also searching for other data that has more general information about the birds and their images, but we might have to compile this on our own as of now.

Data Processing:

- We will be doing some data cleanup as a lot of data is provided to us, and we will remove the data we don't require.
- We will also derive which state the birds are in from the region data they have given.
- We will also have to get bird image data and more data about the bird from a different source.
- The data will be displayed using an interactive map and a few plots to show different trends among the birds.
- We are taking the following details from the predefined dataset.

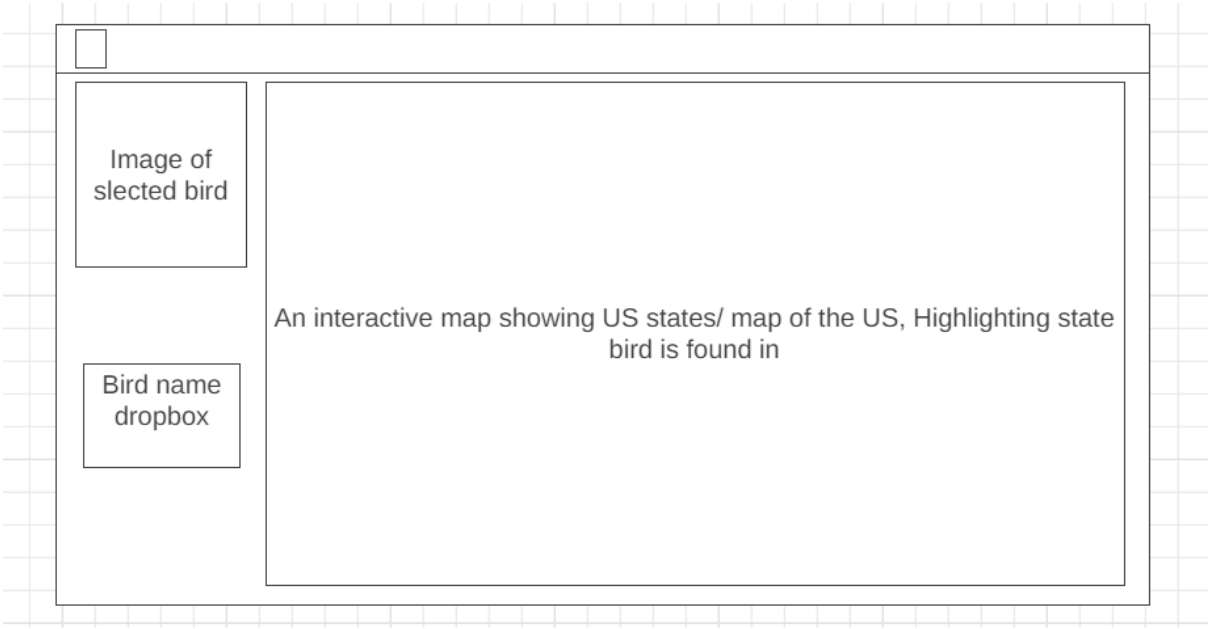
1. Scientific Name – The Latin name of the bird.
2. Common Name – The widely recognized name of the bird.
3. Bird Count – Total number of birds recorded in the selected state.

- We will compile our dataset for the following:

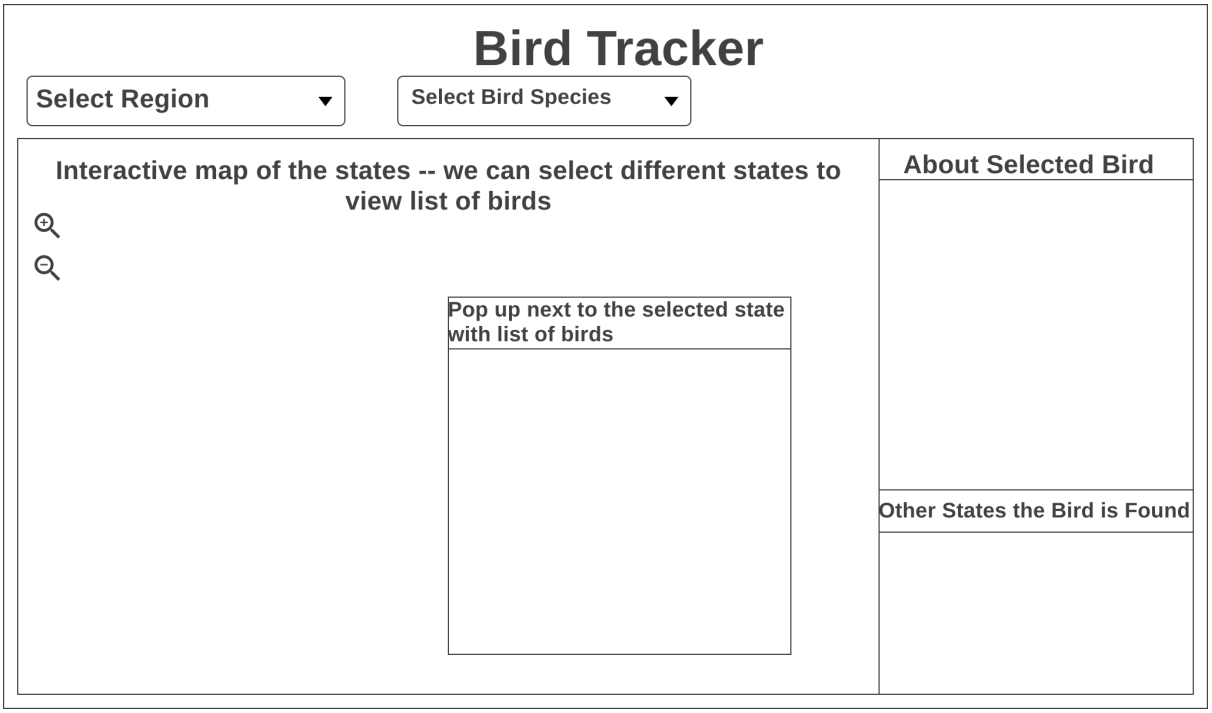
1. Bird Image – Image of the selected bird.
2. Interesting Facts – Interesting facts about the chosen bird.

sThree alternative prototype designs for our visualization

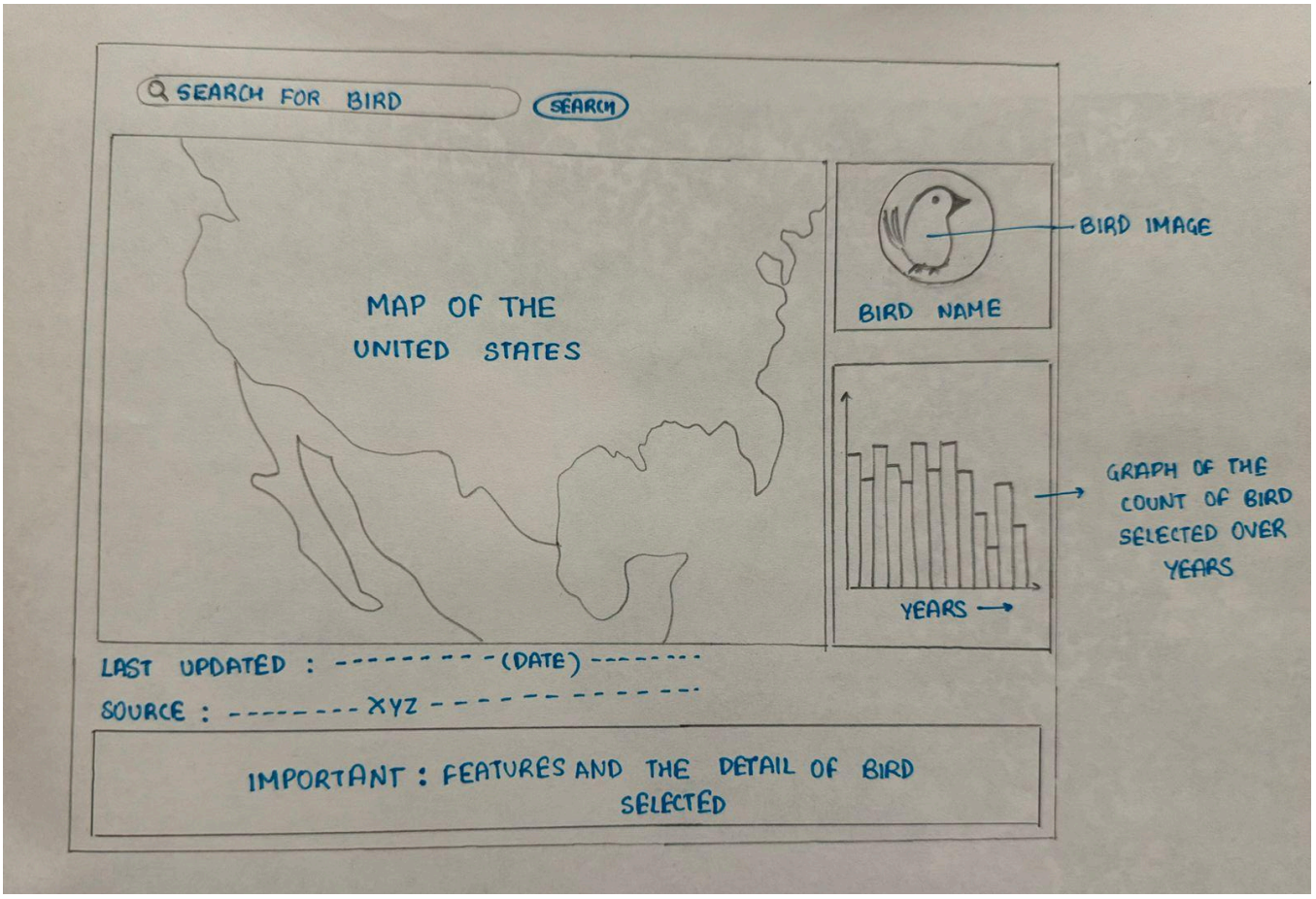
Tanmay:



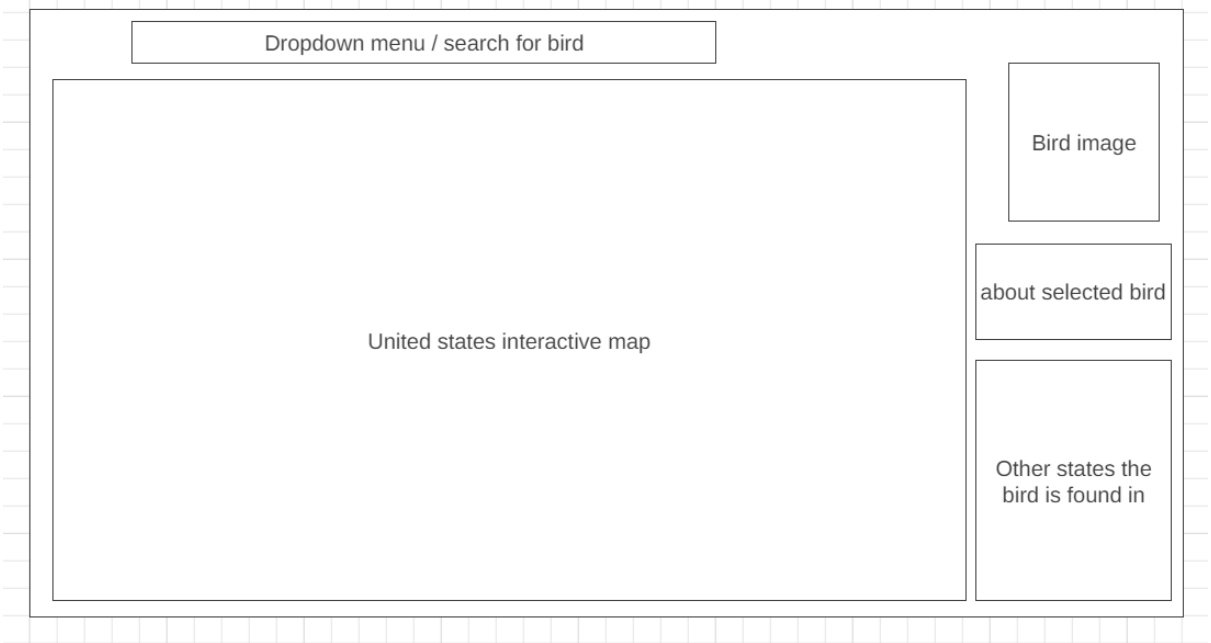
Sujit:



Sakshi:



Final design Sketch



Design description and justification

- **Interactive Map:-** this is the core aspect of our design. We will show a map of the United States with the state borders, and each state will be selectable. Also, when a bird is selected, the state will be highlighted
- **Dropdown menu:-** will have birds available to be selected or searched for.
- **Bird Image:-** will have an image of the bird selected from the dropdown menu.
- **About Bird:-** will have the number of birds captured for observation and other facts about the bird.
- **Other states:-** the bird is found in will be a list of other states where we can find the bird

Data cleanup

Our initial dataset has a lot of data for the bird species that are not required, almost 30 fields, so we reduced these to 4

- BCR
- BirdName
- Scientific Name
- Bird count

Datasets

We have five datasets as of now

- Bird data cleaned
- Bird info
- US states(for D3)
- BCR to State
- Images

Bird data cleaned

This dataset has been “cleaned” from the main dataset which we are using for our visualization. It has been reduced from 30 columns of data to 4 that are relevant to us

Bird info

This dataset has migration patterns, life spans, fun facts

BCR to State

This dataset has the data of which state is a part of which Bird Conservation Region

Images

This holds all images of the birds

Initial Code creation

Library used

- D3.js

Stack

- HTML
- CSS
- JS

Code

- Created an interactive map of the USA, using D3.js which is interactive.
- We created a dropdown for selecting birds.
- We added a function to show the selected bird’s image and a few facts about it
- At this point you can hover over a state and it will register as such in the console

Visualization so far 10/23/2024

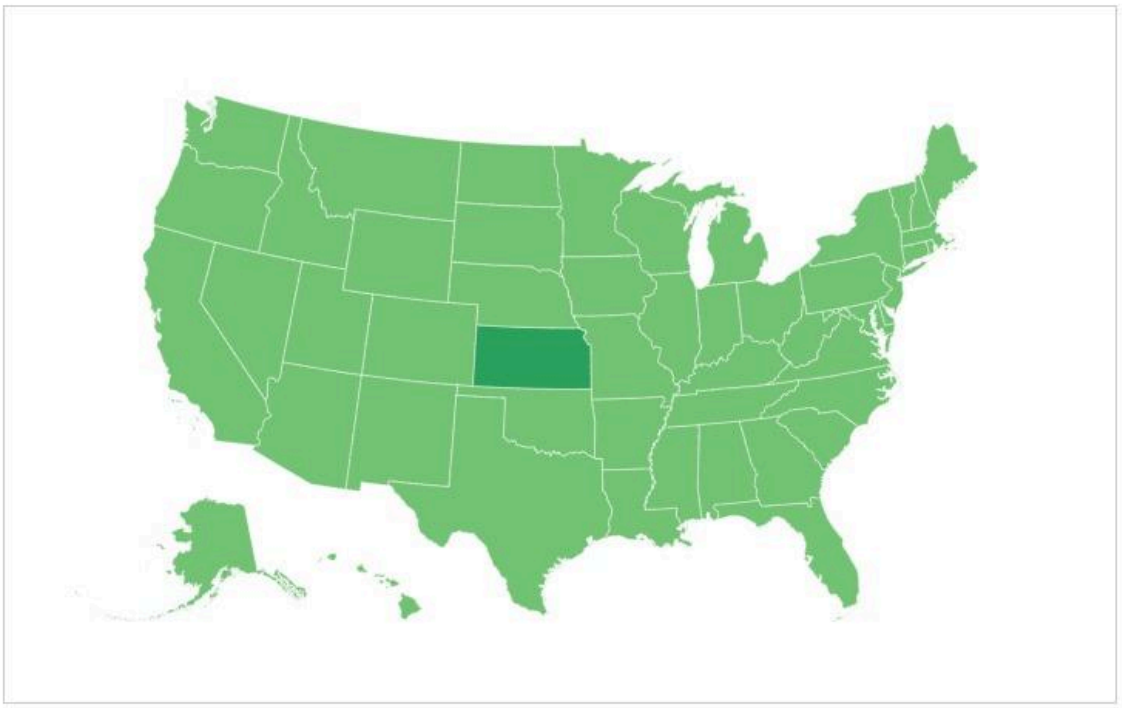


Figure: Highlight due to hover over state

Figure: Initial screen

Figure: Screen after bird selection

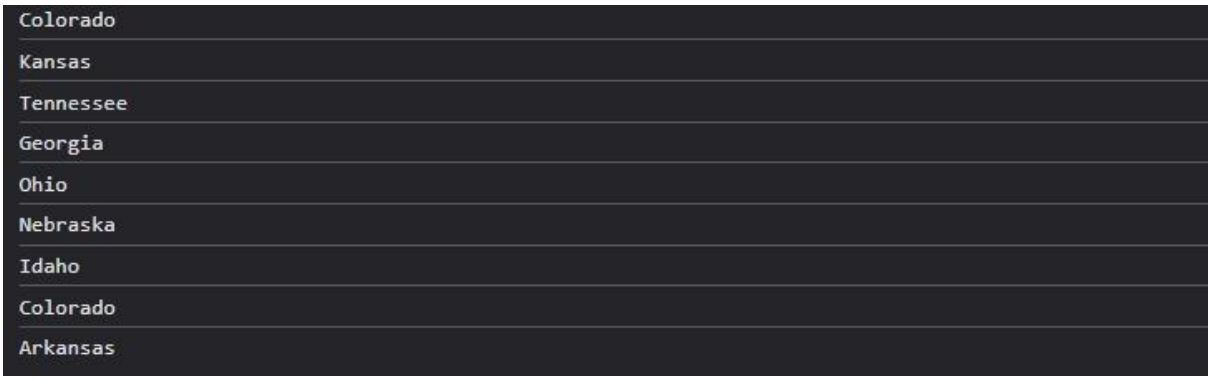


Figure: When State is clicked, it shows up on the console

Mapping BCR(region) to state

We have created a table which maps The Bird Conservation Region to specific states

State	Bird Conservation Region(s)
Alabama	27 - Southeastern Coastal Plain, 28 - Appalachian Mountains
Alaska	1 - Aleutian/Bering Sea Islands, 2 - Western Alaska, 3 - Arctic Plains/Mountains, 4 - Northwestern Interior Forest
Arizona	33 - Sonoran/Mojave Deserts, 34 - Sierra Madre Occidental, 35 - Chihuahuan Desert
Arkansas	24 - Central Hardwoods, 25 - West Gulf Coastal Plain, 26 - Mississippi Alluvial Valley
California	5 - Northern Pacific Rainforest, 32 - Coastal California, 33 - Sonoran/Mojave Deserts
Colorado	10 - Northern Rockies, 16 - Southern Rockies/Colorado Plateau, 18 - Shortgrass Prairie
Connecticut	30 - New England/Mid-Atlantic Coast, 14 - Atlantic Northern Forest
Delaware	30 - New England/Mid-Atlantic Coast

Florida	31 - Peninsular Florida, 27 - Southeastern Coastal Plain
Georgia	27 - Southeastern Coastal Plain, 28 - Appalachian Mountains
Hawaii	N/A (Not part of the continental Bird Conservation Regions)
Idaho	9 - Great Basin, 10 - Northern Rockies, 11 - Prairie Potholes
Illinois	22 - Eastern Tallgrass Prairie, 23 - Prairie Hardwood Transition
Indiana	22 - Eastern Tallgrass Prairie, 23 - Prairie Hardwood Transition
Iowa	11 - Prairie Potholes, 22 - Eastern Tallgrass Prairie, 23 - Prairie Hardwood Transition
Kansas	18 - Shortgrass Prairie, 19 - Central Mixed Grass Prairie, 20 - Edwards Plateau
Kentucky	24 - Central Hardwoods, 28 - Appalachian Mountains
Louisiana	25 - West Gulf Coastal Plain, 26 - Mississippi Alluvial Valley
Maine	14 - Atlantic Northern Forest, 30 - New England/Mid-Atlantic Coast
Maryland	30 - New England/Mid-Atlantic Coast, 29 - Piedmont
Massachusetts	30 - New England/Mid-Atlantic Coast
Michigan	12 - Boreal Hardwood Transition, 23 - Prairie Hardwood Transition
Minnesota	11 - Prairie Potholes, 12 - Boreal Hardwood Transition
Mississippi	26 - Mississippi Alluvial Valley, 27 - Southeastern Coastal Plain
Missouri	24 - Central Hardwoods, 25 - West Gulf Coastal Plain
Montana	9 - Great Basin, 10 - Northern Rockies, 11 - Prairie Potholes
Nebraska	11 - Prairie Potholes, 18 - Shortgrass Prairie, 19 - Central Mixed Grass Prairie
Nevada	9 - Great Basin, 33 - Sonoran/Mojave Deserts
New Hampshire	14 - Atlantic Northern Forest, 30 - New England/Mid-Atlantic Coast
New Jersey	30 - New England/Mid-Atlantic Coast
New Mexico	16 - Southern Rockies/Colorado Plateau, 35 - Chihuahuan Desert
New York	13 - Lower Great Lakes/St. Lawrence Plain, 14 - Atlantic Northern Forest
North Carolina	27 - Southeastern Coastal Plain, 28 - Appalachian Mountains, 29 - Piedmont
North Dakota	11 - Prairie Potholes
Ohio	13 - Lower Great Lakes/St. Lawrence Plain, 23 - Prairie Hardwood Transition
Oklahoma	19 - Central Mixed Grass Prairie, 20 - Edwards Plateau

Oregon	5 - Northern Pacific Rainforest, 9 - Great Basin
Pennsylvania	13 - Lower Great Lakes/St. Lawrence Plain, 28 - Appalachian Mountains
Rhode Island	30 - New England/Mid-Atlantic Coast
South Carolina	27 - Southeastern Coastal Plain, 28 - Appalachian Mountains
South Dakota	11 - Prairie Potholes, 17 - Badlands and Prairies, 18 - Shortgrass Prairie
Tennessee	24 - Central Hardwoods, 28 - Appalachian Mountains
Texas	20 - Edwards Plateau, 21 - Oaks and Prairies, 36 - Tamaulipan Brushlands
Utah	9 - Great Basin, 16 - Southern Rockies/Colorado Plateau
Vermont	14 - Atlantic Northern Forest
Virginia	29 - Piedmont, 30 - New England/Mid-Atlantic Coast, 28 - Appalachian Mountains
Washington	5 - Northern Pacific Rainforest, 9 - Great Basin
West Virginia	28 - Appalachian Mountains
Wisconsin	12 - Boreal Hardwood Transition, 23 - Prairie Hardwood Transition
Wyoming	10 - Northern Rockies, 16 - Southern Rockies/Colorado Plateau, 17 - Badlands and Prairies

Current files in the project

```
bird-tracker-visualization/
│
├── data/
│   └── cleaned_data.csv
│
├── src/
│   ├── preprocessing/
│   │   └── data_cleaning.py
│   │
│   ├── js/
│   │   ├── main.js
│   │   ├── utils.js
│   │   └── data_processing.js
│   │
│   ├── css/
│   │   └── styles.css
│   │
│   └── html/
│       └── index.html
│
└── README.md
```

Meeting with TA

We discussed quite a few things we could improve in our project and it also lead us to realize that we needed to add more things these were finalized in our next meeting as written in the next paragraph

Meeting for the project- discussion on adding more things to our visualization and derived attributes 10/22/2-24

Ideas

- Timeline for
    - Breeding status
    - Number of birds in the state
- Life span

We found more data for creating a timeline-based approach for the graphs, including bird breeding and the number of birds seen across the years.
Based on the number of birds and breeding status, we can derive whether the conservation of the bird is working or not
After meeting with the TA, she suggested exploring more on ways to **promote understanding and awareness** and include more derived attributes.

Showing how bird numbers have changed based on current weather/temperature changes.

We tried getting datasets for urbanization to compare with bird numbers but couldn't get a good enough dataset/ one which don't require as much work

Switching direction

Our initial approach led to us having to do a lot more preprocessing while having very limited useful data; we have found a few more datasets from the same source that provide a lot more valuable information in a way that we can extract more of its potential so we are planning to use that now along with our existing dataset. This is also to address the TAs point. With the new data, we can add more details about the birds and create a timeline, allowing us to show more patterns like birds spotted based on year, breeding numbers based on year, etc. This also allows us to make derived attributes to show the trends of birds over a set of years.

12 November 2021
Following the TAs' suggestions, we made significant enhancements to our webpage to provide a more comprehensive and informative experience for users. We expanded the range of datasets incorporated into the project, enabling us to present a broader and more detailed analysis. To improve navigation and clarity, we created distinct tabs on the webpage, each dedicated to specific information. These tabs include:

- **Breeding Status:** Displaying the reproductive behaviors and patterns of various bird species.
- **State-wise Bird Distribution:** Highlighting the number of different bird species found across states in the US. Also we added **Life Span**: Providing insights into the typical life expectancy of these birds.

Additionally, we integrated a **Process Book** into the webpage, detailing the project's methodology, including the steps we followed for data collection, analysis, and visualization. This addition not only showcases our workflow but also offers transparency into how the project was developed. These improvements have made the webpage more engaging and useful, aligning with our goal to raise awareness about bird biodiversity and conservation.

29 November 2022

1) Thought to remove youtube link from index page and create a new tab in navbar.
2) Change the default bird image
3) Change the color palette and use a color blind friendly palette.

2 - 5 th December
Created a lot of code for making the breeding data available
Made a test set with songbirds in California from 1992 to 2018 and used it to test the data processing file we made.

Step 1

Remove all the extra unneeded data

```
import pandas as pd



input_file = "datavis_processing/banding.csv"
output_file = "output.csv"

columns_to_keep = ["LOC", "SPEC", "year"]

try:

    data = pd.read_csv(input_file)



    filtered_data = data[columns_to_keep]



    filtered_data.to_csv(output_file, index=False)

    print(f"New CSV file saved as: {output_file}")
except FileNotFoundError:
    print(f"File {input_file} not found. Please check the file path.")
except KeyError as e:
    print(f"Column {e} not found in the CSV file. Please ensure the
specified columns exist.")
```

Step 2

Compile the individual bird data to sum up count for specific bird in specific location

```
input_file = "output.csv"  # Replace with the path to your input file
output_file = "grouped_data.csv"  # Specify the output file name
fda= filtered_data
try:

   data = pd.read_csv(input_file)
```

```
    grouped_data = fda.groupby(["LOC", "SPEC",
"year"]).size().reset_index(name="count")

    grouped_data.to_csv(output_file, index=False)

    print(f"Grouped data saved as: ")
except FileNotFoundError:
    print(f"File {input_file} not found. Please check the file path.")
except KeyError as e:
    print(f"Column {e} not found in the CSV file. Please ensure the
specified columns exist.")
```

Step 3

Create a legend for loc to state

```
import pandas as pd
input_file = "loc_state.csv"
output_file = "loc_state_legend.csv"
columns_to_keep = ["LOC","STATE"]

Try:

    data = pd.read_csv(input_file)


    filtered_data = data[columns_to_keep]

    filtered_data.to_csv(output_file, index=False)

    print(f"New CSV file saved as: {output_file}")
except FileNotFoundError:
    print(f"File {input_file} not found. Please check the file path.")
except KeyError as e:
    print(f"Column {e} not found in the CSV file. Please ensure the
specified columns exist.")
```

Step 4 mash it together

```
import pandas as pd


input_file = "MAPS_BANDING_capture_data.csv"
```

```
output_file = "output.csv"

columns_to_keep = ["LOC", "SPEC", "year"]
filtered_data=pd.read_csv(input_file,low_memory=False)
try:

    data = pd.read_csv(input_file,low_memory=False)

    filtered_data = data[columns_to_keep]


except FileNotFoundError:
    print(f"File {input_file} not found. Please check the file path.")
except KeyError as e:
    print(f"Column {e} not found in the CSV file. Please ensure the
specified columns exist.")


#input_file = "output.csv"
#output_file = "grouped_data.csv"

fda= filtered_data
Try:
   # data = pd.read_csv(input_file)

        grouped_data = fda.groupby(["LOC", "SPEC",
"year"]).size().reset_index(name="count")


    #grouped_data.to_csv(output_file, index=False)

    print(f"Grouped data saved as: ")
except FileNotFoundError:
    print(f"File {input_file} not found. Please check the file path.")
except KeyError as e:
    print(f"Column {e} not found in the CSV file. Please ensure the
specified columns exist.")

input_file = "MAPS_STATION_location_and_operations.csv"
output_file = "loc_state_legend.csv"


columns_to_keep = ["LOC","STATE"]
```

```
try:

    data = pd.read_csv(input_file,low_memory=False)

    filtered_data = data[columns_to_keep]

    filtered_data.to_csv(output_file, index=False)

    print(f"New CSV file saved as: {output_file}")
except FileNotFoundError:
    print(f"File {input_file} not found. Please check the file path.")
except KeyError as e:
    print(f"Column {e} not found in the CSV file. Please ensure the
specified columns exist.")

#grouped_data_file = "grouped_data.csv"
state_mapping_file = "loc_state_legend.csv"  output_file =
"updated_grouped_data.csv"

try:

    state_mapping = pd.read_csv(state_mapping_file)
    merged_data = grouped_data.merge(state_mapping, on="LOC",
how="left")


    merged_data["LOC"] =
merged_data["STATE"].fillna(merged_data["LOC"])


    merged_data.drop(columns=["STATE"], inplace=True)


    merged_data.to_csv(output_file, index=False)

    print(f"Updated grouped data saved as: {output_file}")
except FileNotFoundError as e:
    print(f"File not found: {e.filename}")
except KeyError as e:
    print(f"Column {e} not found in the file. Please ensure the column
exists.")
```

We finalized the page and then began the long process of making it functional. We tried several methods to create it and initially started from scratch. However, we encountered issues with the code, so we decided to use the base layout we had already created for the main page and built upon it. We removed the bar graphs and replaced them with a line graph. The primary challenge we faced was aligning the two data sets, as there were some JavaScript issues. Ultimately, we were able to resolve these problems and successfully match the data.