Endangered Species Analysis: Process Book

Sydney Lundberg, Victoria Mache

Table of Contents

Overview and Motivation

We're building an educational website that traces how endangered and threatened species in the U.S. have changed from 1976–2019 and links those shifts to key policies and events. Using a USDA Research Data Archive dataset, we'll clean the CSV (fix dates, rename columns, handle missing values, and derive year/season, group shares, and spikes) in Python, then develop that data into interactive visuals. The core functionality pairs a streamgraph of species composition over time with a coordinated heatmap of year × taxonomic group, supported by stacked bars for quick snapshots and an annotated timeline that grounds patterns in real-world context. If we can find reliable state-level data, we'll add an interactive U.S. map to surface regional differences. Our goal is to spot when and where listings surge, highlighting everyday species people encounter, and sparking more informed conservation decisions.

Endangered-species data is public but fragmented and easy to misinterpret. That confusion hides meaningful trends, like years when listings jump after a policy change, and it can be difficult to realize that many species people recognize everyday are still at risk. Take monarch butterflies, they're common in yards and school gardens, yet multiple assessments in the last decade flagged major declines driven by habitat loss and climate pressures. By cleaning the data, translating the jargon, and aligning trends with concrete events and laws, we want to make those patterns obvious so teachers can teach them, journalists can explain them, and decision-makers can act on them.

Users

We're building this for people who want easier access to endangered species information, such as teachers pulling a quick lesson, students needing reliable facts to cite, local reporters and advocates who need a clean chart for a story, and policymakers who want to make informed decisions based on the most current trends. The site should be simple to use: choose a year or a species group, see what changed, see a note for context, and grab the chart or CSV with a citation.

Related Work

There are various conservation sites that already exist and inspire our own project. One of these organizations is the WWF, or World Wildlife Federation. They are similar to our project in that they try to make wildlife conservation more fun and digestible for the general public and those without a background in science. They do this by allowing people to donate towards specific species they care about by paying to "adopt" specific animals in their care. In addition, every two years, they publish the Living Planet Report, which distills big datasets like the Living Planet Index into clear visuals and short takeaways. That webpage has been a large inspiration for this assignment.

In order to add the historical annotations to our website, we will be using this United States Fish & Wildlife site to implement the effects of any Endangered Species Act implementations; https://www.fws.gov/page/endangered-species-act-amendments.

Questions

Primary questions:

- How has the total number of endangered and threatened species in the U.S. changed since 1976?
- Which taxonomic groups exhibit the largest absolute growth and the largest share changes over time?
- Which historical events or policy milestones coincide with major surges or slowdowns in listings?
- How might public awareness shift if the data behind everyday species at risk was more digestible?

Potential future questions/changes:

- If early charts show a few groups driving most of the totals, we'll shift the view to focus on relative shares so the smaller groups aren't completely drowned out.
- If a spike lines up with multiple nearby events, we'll group those events on the timeline and invite side-by-side comparisons of the surrounding years.
- If the state data can be found, we'll add more locative questions, i.e. "Which regions see the largest increase in endangered species listings in X decade?".

Data

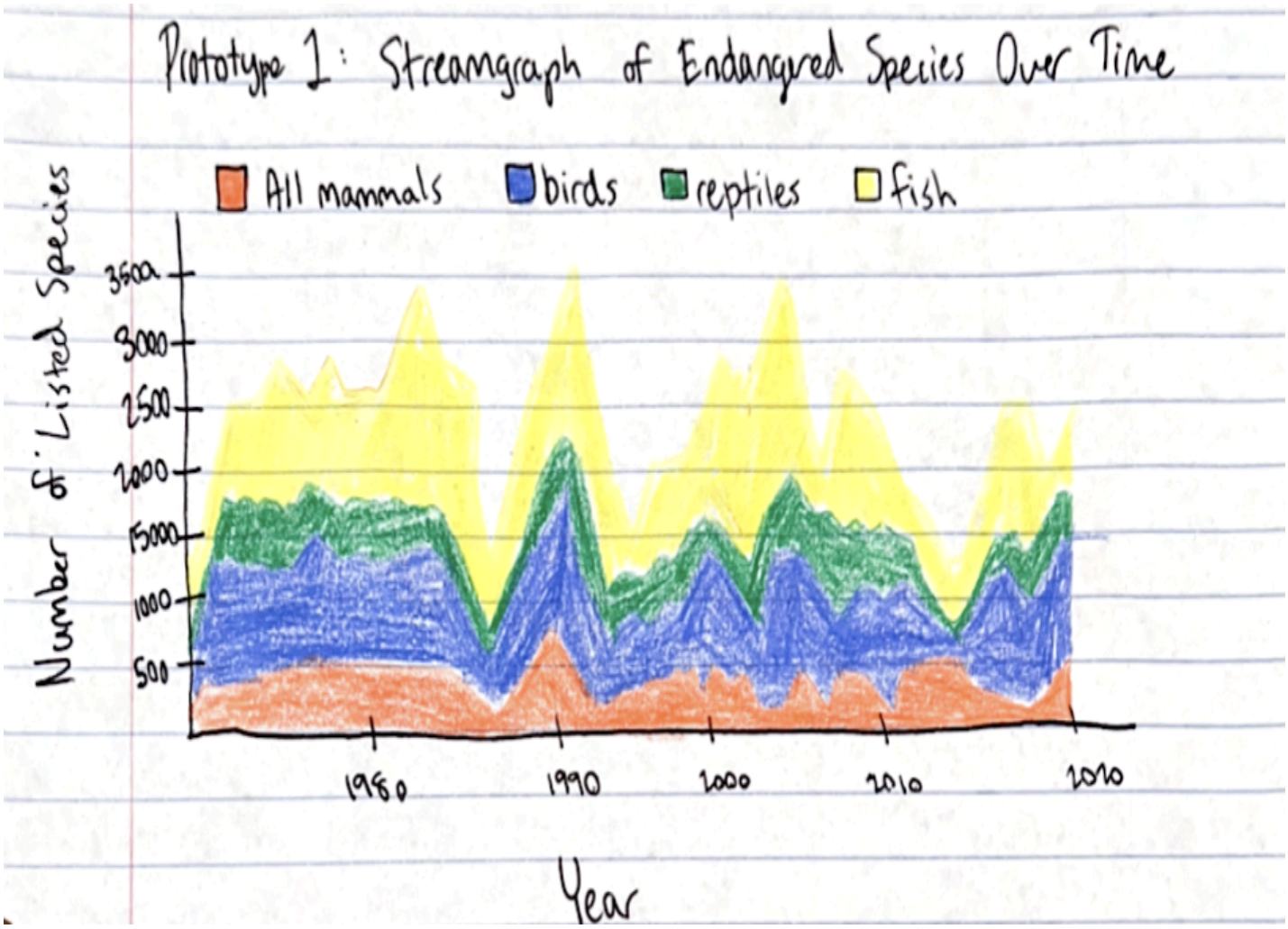
Our data is from the United States Department of Agriculture, specifically from their research data archive linked here. This was an ideal resource for us because we know this to be a reputable source with unbiased data. We will be able to download this data as a raw .csv file, which will make opening it in a readable format relatively uncomplicated. For processing, we don't expect extremely heavy cleanup, but a few important steps will be necessary:

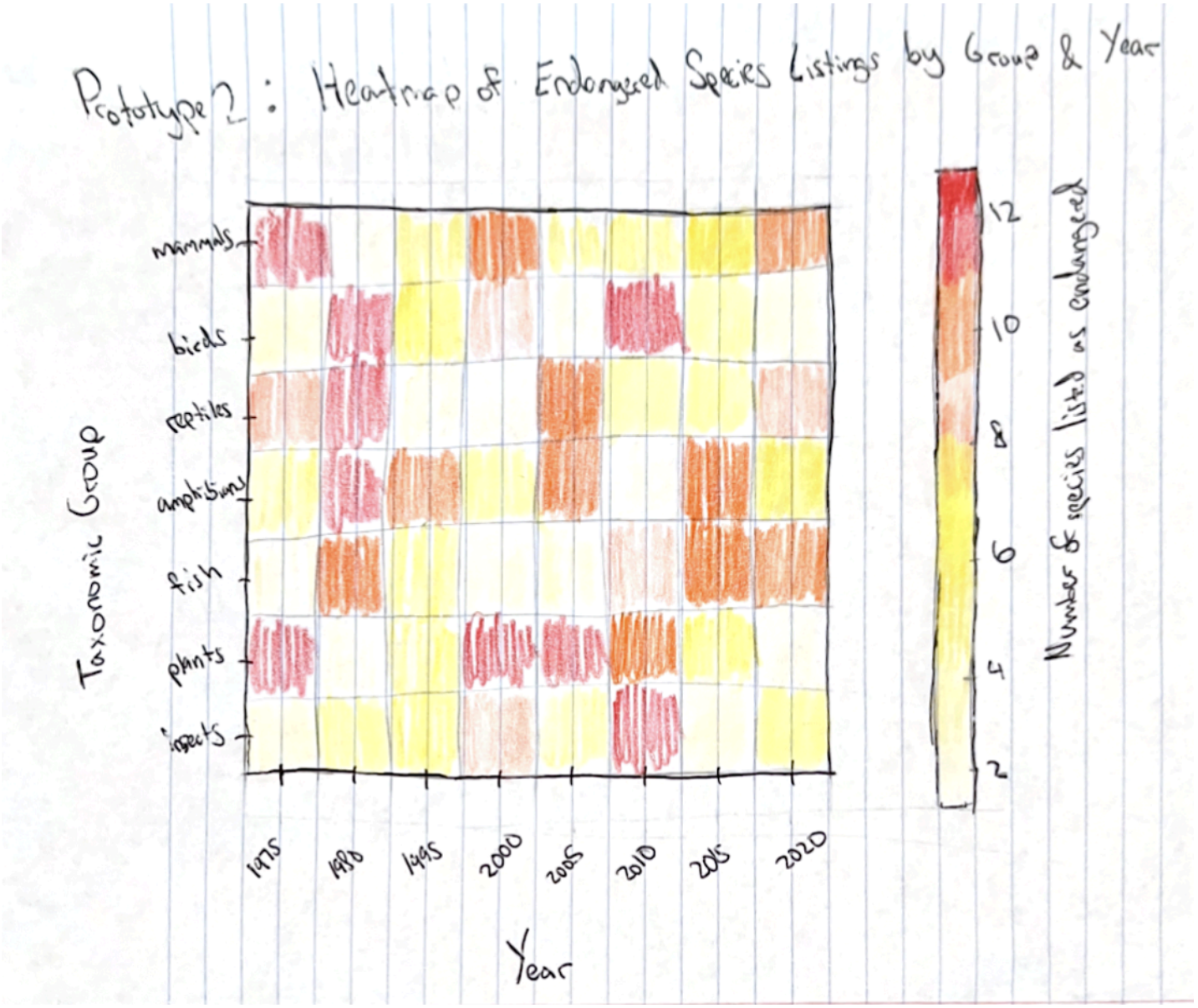
- The date column is stored as a text string: ex., 1 Jul 76. We want to convert this to a datetime object and also create new variables such as year and a categorical season (spring, summer, fall, winter) for simpler grouping and filtering.
- Most of the columns use shorthand titles, like all_m, which are not very intuitive. We will rename these to improve readability and clarity for users and ourselves. For example, all_m will become all_mammals.
- Some columns have missing values, especially in early years for groups like crustaceans or corals. If the missing values mean that no species were listed, we can just replace them with zeros. Otherwise, we can leave them out entirely. We will also remove extra "total" columns, such as all_ani and all_all, so we don't accidentally count the same species twice.
- We will derive some extra quantities to make the data more useful. For example, we'll calculate how many new species were added between time periods to show spikes and show growth over long stretches like decades.

- We plan to use Python Pandas to clean and prepare the data, then save it as JSON or CSV for the website. The interactive charts will most likely be built in D3.js or Plotly, since they work best with this data format.

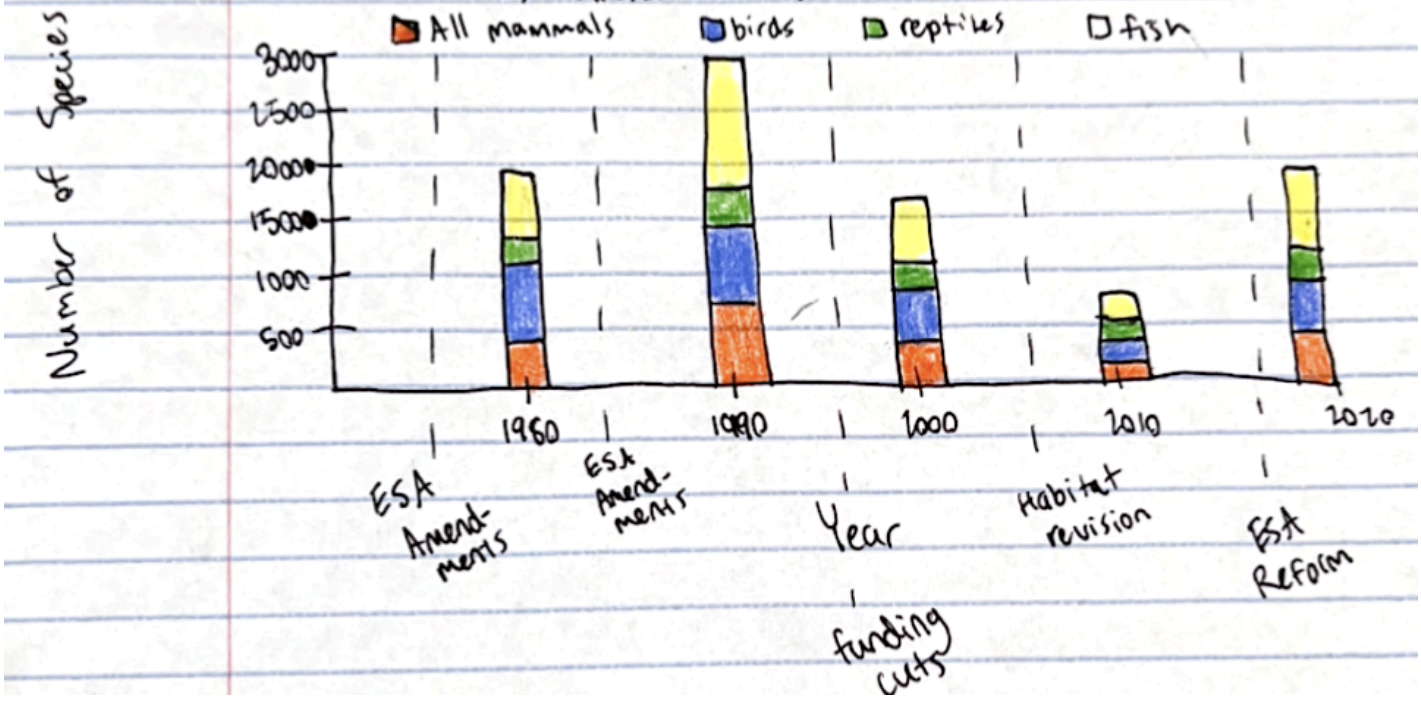
Initial Visualizations

In this first prototype of the website, we plan to display our data with a streamgraph showing the relative share of taxonomic groups of endangered species and the overall trend. This will be useful to us because the data set has a temporal focus, and it will highlight large surges in endangered populations. We will also make an animated timeline that will include the historical annotations, like legislation changes and policy milestones. Most of those will come from the Endangered Species Act and related amendments. Additionally, a heatmap could display the year vs. species group, with color intensity demonstrating frequency. This could be helpful in quickly spotting spikes in the endangered species listings. A stacked area chart would also be great to display what proportion of the total number of endangered species each different taxonomic group represents at a given time. The initial design sketches are shown below.

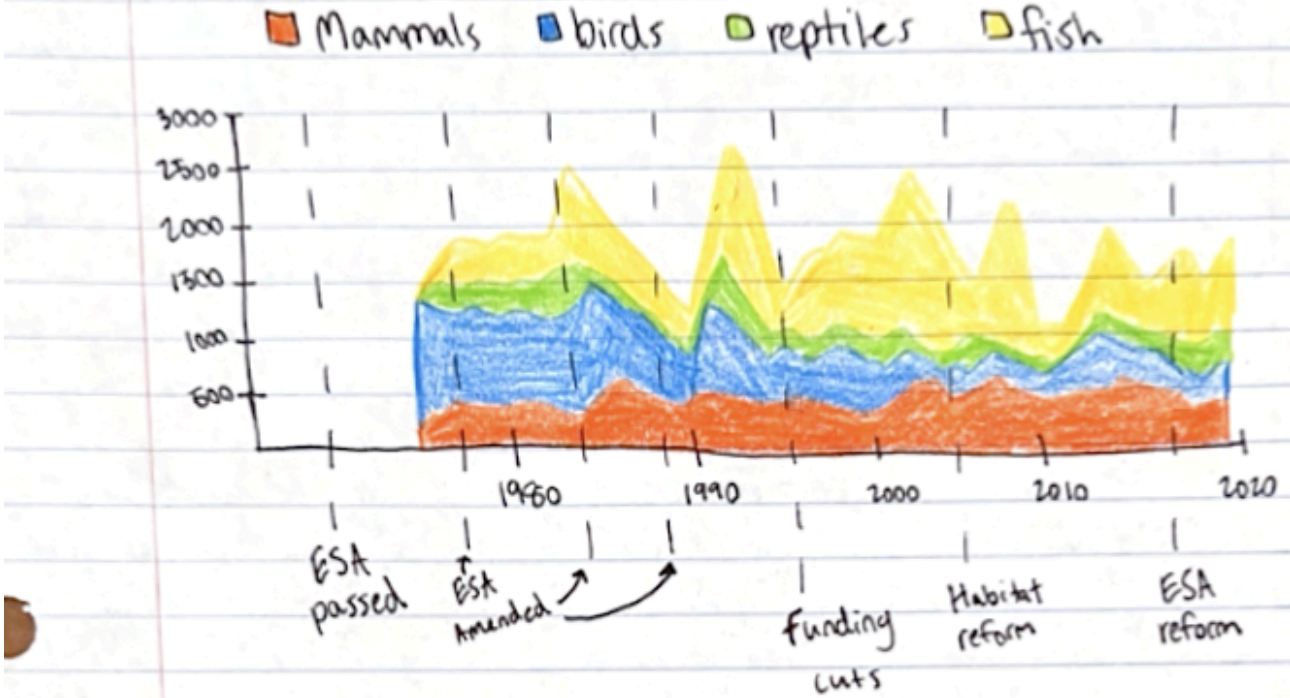

Prototype 1: Streamgraph of Endangered Species Over Time
All mammals
birds
reptiles
fish
Number of Listed Species
3500
3000
2500
2000
1500
1000
500
1980
1990
2000
2010
2020
Year

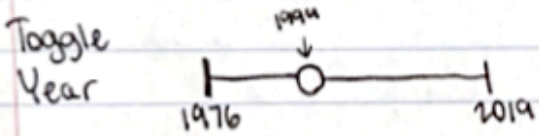

Prototype 2 : Heatmap of Endangered Species Listings by Group & Year
Taxonomic Group
mammals
birds
reptiles
amphibians
fish
plants
insects
1975
1980
1985
2000
2005
2010
2015
2020
Year
12
10
8
6
4
2
Number of species listed as endangered

Prototype 3: Stacked Bar Chart of Endangered Species w/ Historical Annotation

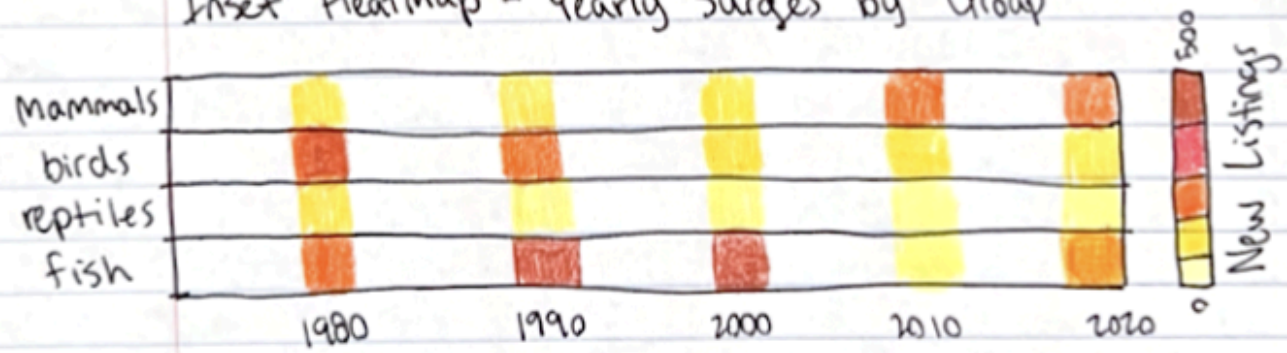


Prototype 4: Final Hybrid Animated Timeline Streamgraph + Annotations + Heatmap





Inset Heatmap - Yearly Surges by Group



Prototype 4 is our most important visualization, and what will be reflected on our website. This is the final combination of these previous prototypes. It will be animated once it's been made digital so that it moves in a stream as the years increase. During years with significant legislation and policy changes, or environmental events, these events will be represented in the graph, giving it a timeline-like feel. The user will also have a slider with which to pause on or select a particular year. Underneath the animated streamgraph, the heatmap will also be animated to highlight significant surges in the number of endangered species. I only represented one bar per decade in the heatmap for clarity, but in reality, there would be one generated for each year between 1976 and 2019. These two graphs will be moving in unison as years increase.

One of the must-have features that represents the main passion behind and the goal of this project is the interactive timeline. We want to blend data visualization with storytelling, offering potential hypotheses for why the data is behaving in such a way. Within this timeline, the user must be able to filter certain time periods, specific species, or certain historical events. These features allow the timeline to be truly interactive and tailored to the needs of the user.

Project Review With Staff

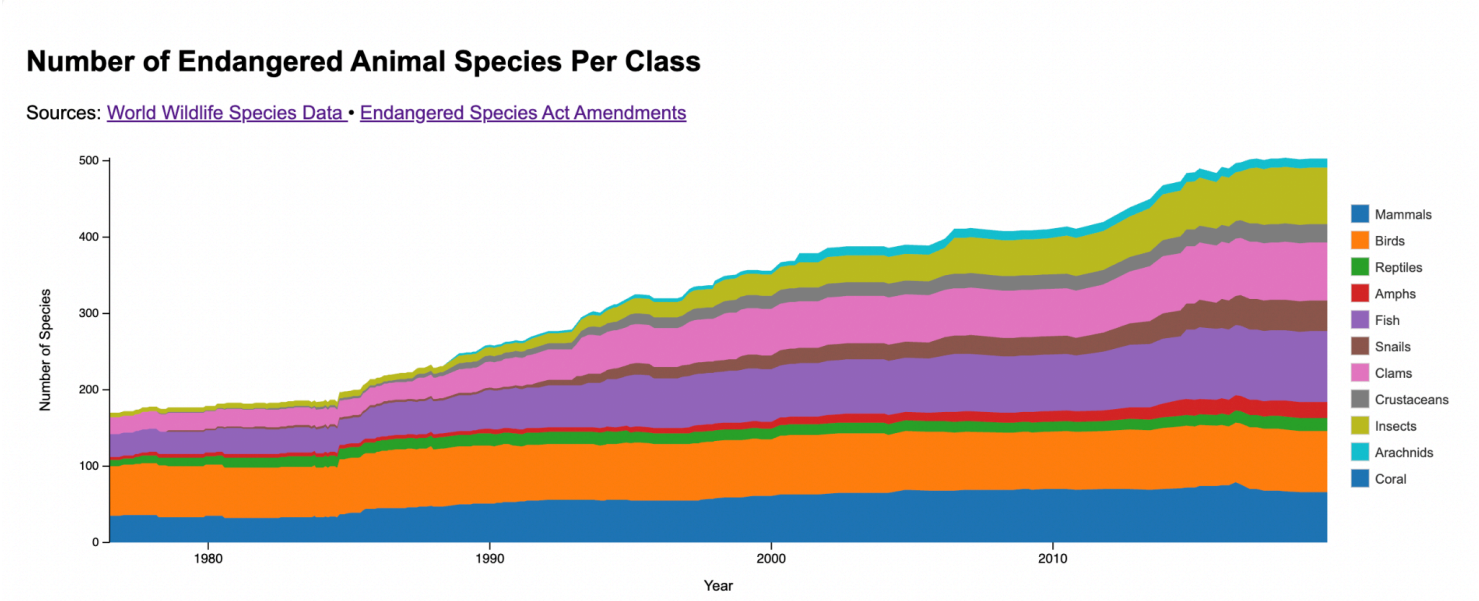
Upon reviewing with Rifat Ara Proma, we were given valuable constructive criticism. She suggested we add a little more depth to the project in order to create a more well-rounded picture of the endangered species status over time. As a result, we introduced a second data set from the Virginia Open Data Portal. It shows the counts of all U.S. species listed as threatened or endangered under the Endangered Species Act of 1973, where the count for each year is the number of species added to the list in that year.

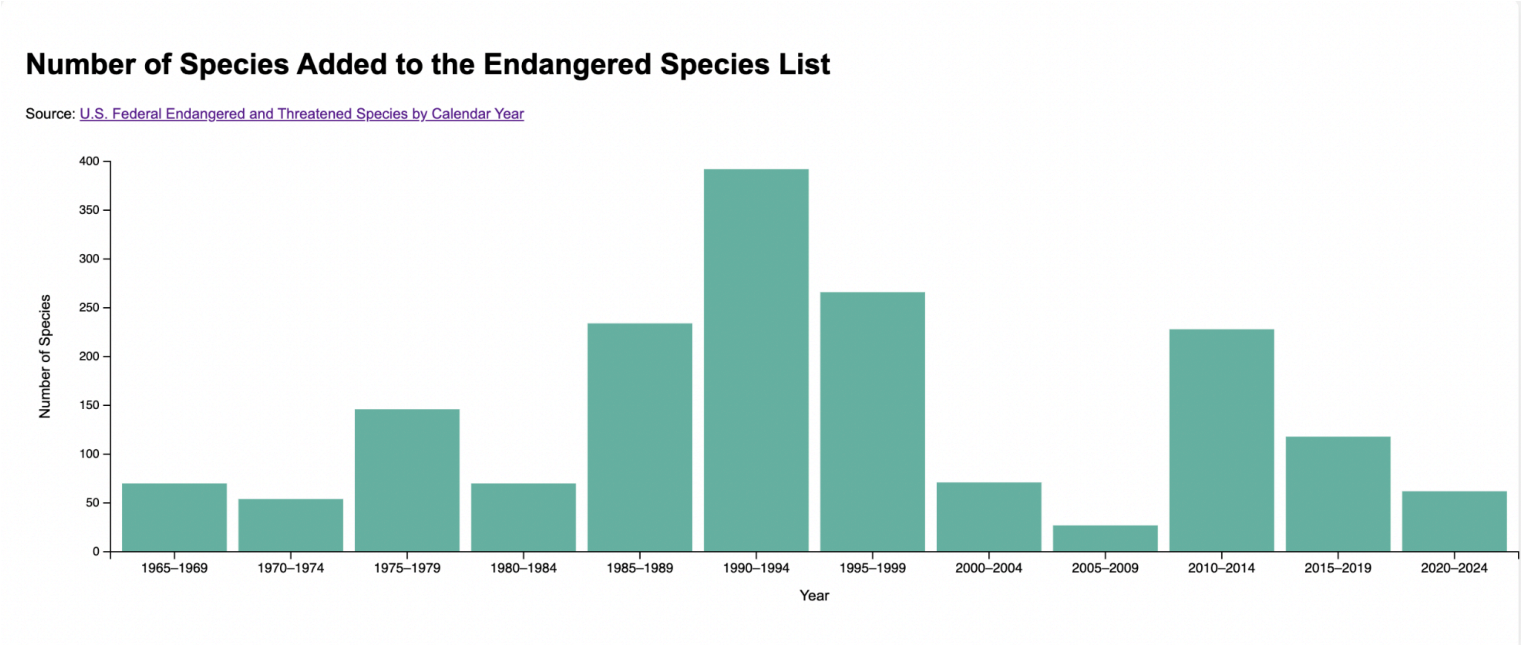
See the dataset here: https://data.virginia.gov/dataset/u-s-federal-endangered-and-threatened-species-by-calendar-year

This data set is different from the other one we are using, because instead of showing the number of endangered species per type of animal, it shows just species that are newly added (or readded) to that list in that specific year. We will use this data set to create some auxiliary visualizations on the website to further compare with our original dataset.

Exploratory Data Analysis - Milestone

We began by implementing the streamgraph from our proposal. This gave us insight into which classes have the most endangered species: at this point its animals and birds. The next step was to use our second data set, which shows the number of new species added to the endangered species list each year. We represented this as a histogram to explore the variation in frequency. We have not yet implemented the historical information about the Endangered Species Act key dates to identify causation points in these trends. We intend to embed this in both the streamgraph and the histogram for further contextualization.





Once we created the first two graphs, we reflected on the story the data told, and found that it gave a lot of useful insight into general species trends in the U.S. However, the United States is enormous and contains so much biodiversity and variation in terrain and climates. After a discussion, we came to the conclusion that we would actually like to have a geographical aspect to this analysis as well. So, we decided to also create a visualization that is a map of the 50 states, which when the user hovers over each state, it will show the current number of endangered species that are native to that state. We generated a dataset to represent this using an LLM, and incorporated outside knowledge of general species trends in each state for the distribution. I.e. Hawaii is known for having numerous endangered species, so, in this data set, it has the largest quantity. This data is cleaned and ready to go. Creating the map visualization is still a work in progress, it is partially interactive, as the hovering feature works, however the

gradient is not yet appearing.