Sydney Lundberg u1405285

Victoria Mache u1390767

GitHub Repository

Endangered Species Analysis

The goal of this project is to create an educational website where users can explore the rise and fall of endangered species in the United States and how it is shaped by historical events. There are several motivations behind this project, both conservational and informational. Firstly, both of us feel very strongly about protecting the environment and conserving natural resources, including animal species. By having the ability to identify key catalysts for endangered populations, both positive and negative, policymakers can pass more effective legislation to protect these vulnerable groups. Moreover, the number of endangered species in the U.S. is extensive and always increasing, so the usefulness of having a central site that shows all of the various types of animals: mammals, amphibians, insects, etc., cannot be understated. This website would also raise public awareness of what animals should be treated with extra caution. For example, many people see Monarch butterflies around their gardens/backyards all the time, never knowing that in recent years their population has significantly declined and become at risk. Often, kids will try to catch and keep them, which can unfortunately result in their death. However, if they were better educated about endangered species, this could be prevented.

One of the main questions we will be trying to answer is how the total number of endangered and threatened species in the U.S. has changed since 1976. Which taxonomic groups are the most affected by historical events/legislation? And which of these events has been the most significant for the populations? Finally, we want to explore how public awareness would be different if people were aware of the everyday species that are at risk.

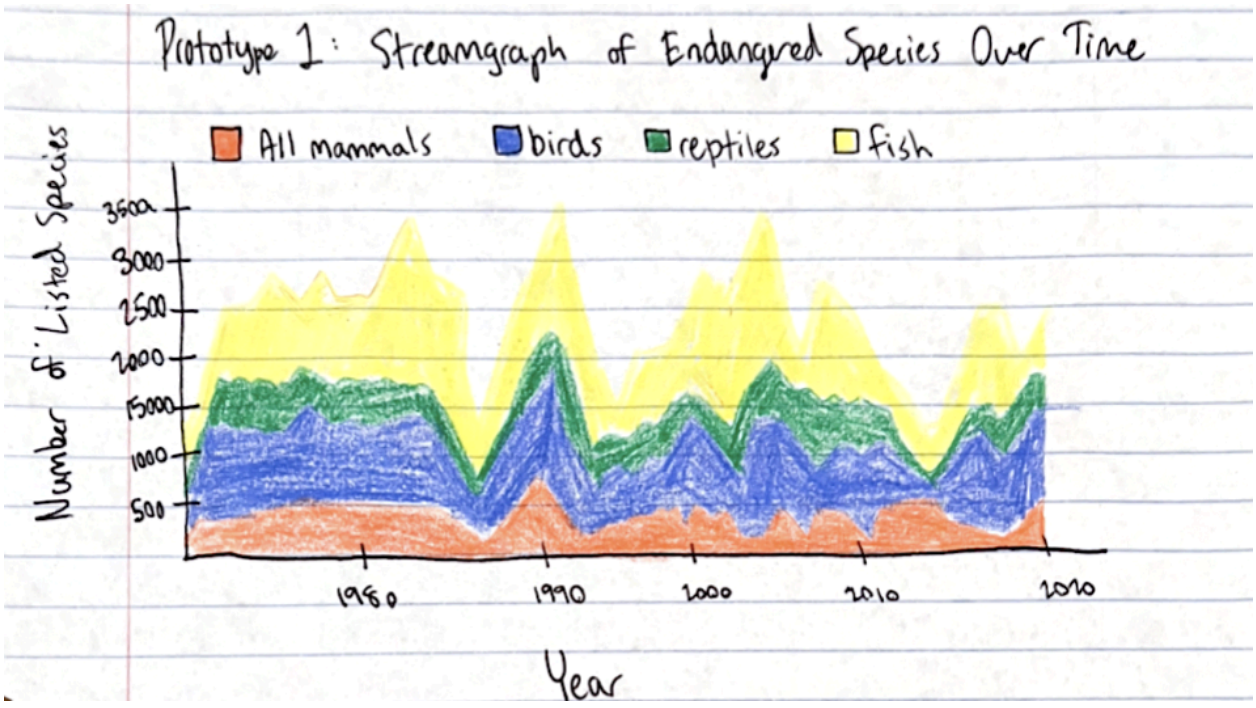
Our data is from the United States Department of Agriculture, specifically from their research data archive linked here. This was a desirable resource for us because we know this to be a reputable source with unbiased data. We will be able to download this data as a raw .csv file, which will make opening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

1. 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.
2. 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.
3. 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, the share of each group compared to the total to show balance between plants and animals, and 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.

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 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.

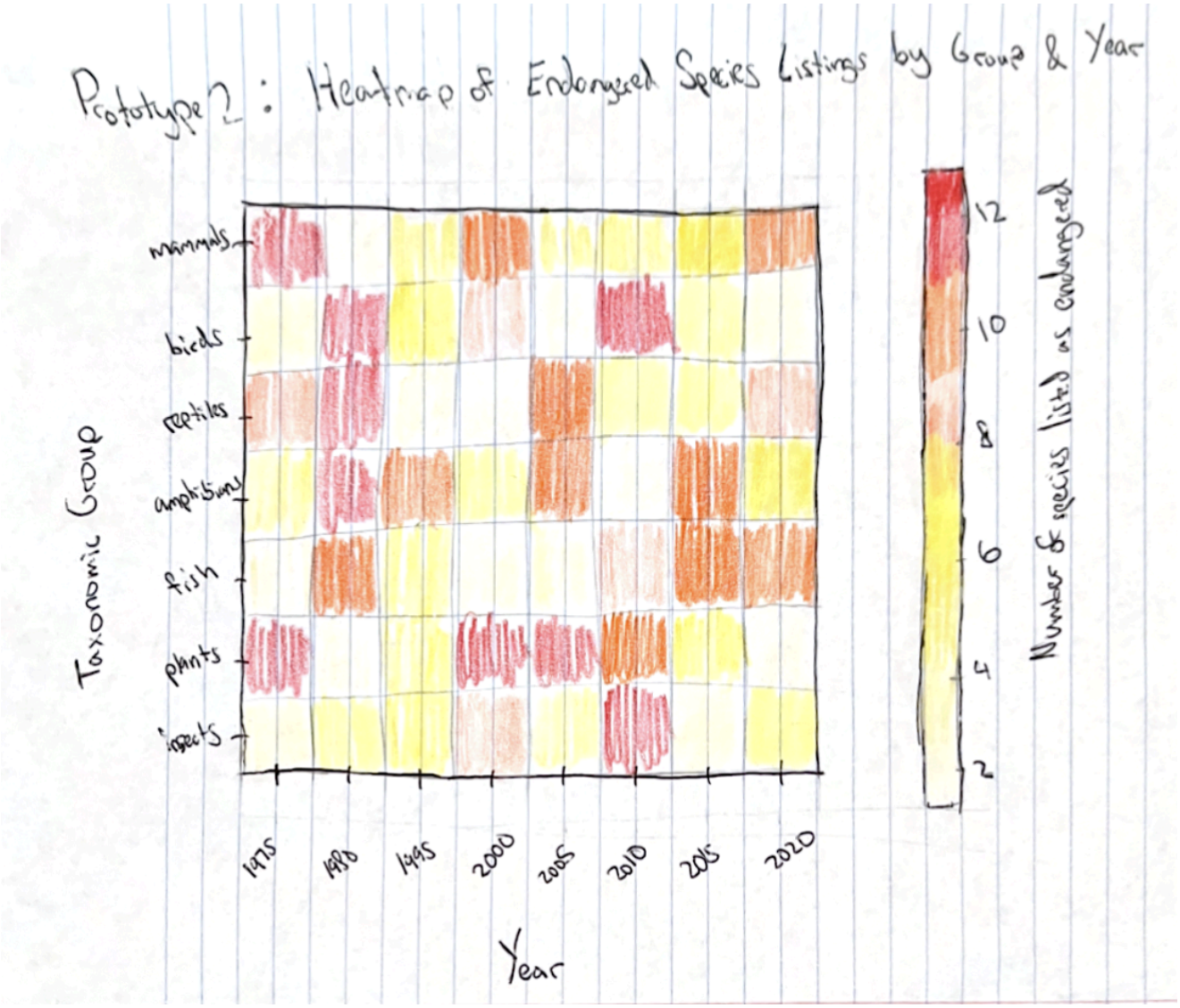
Prototype 1:



The streamgraph emphasizes fluctuation in the number of endangered species listings over time and makes any large surges or declines in listings more visually apparent. The

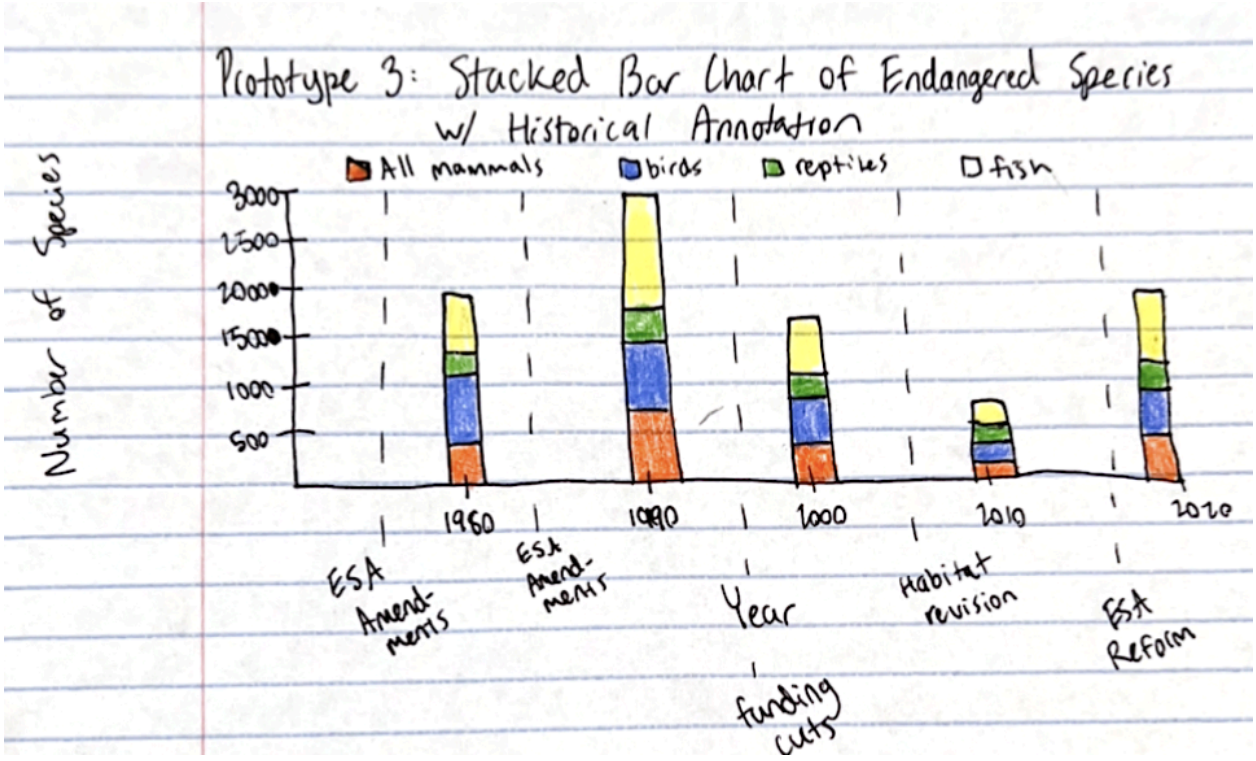
thickness of a stream at a given point in time encodes the relative size of that group within the total number of endangered species. It shows a high-level proportional view. The distinct colors of each taxonomic group ensure clear differentiation between groups. This prototype just shows a few selected representative taxonomic groups for clarity, but the real visualization will include them all.

Prototype 2:



This heatmap of endangered species listing by group and year visually demonstrates in what year which type of species had the most endangered listings. The darker colors show a higher frequency of species listings, and the lighter colors represent fewer listings. The colors make it easy to represent the years when many new species across groups were listed, which could possibly be linked to new amendments, policies, or environmental events. The heatmap is beneficial because it is an organized way to visualize how the three variables, taxonomic group, year, and number of species listed, are linked. This prototype just shows a few selected representative taxonomic groups for clarity, but the real visualization will include them all.

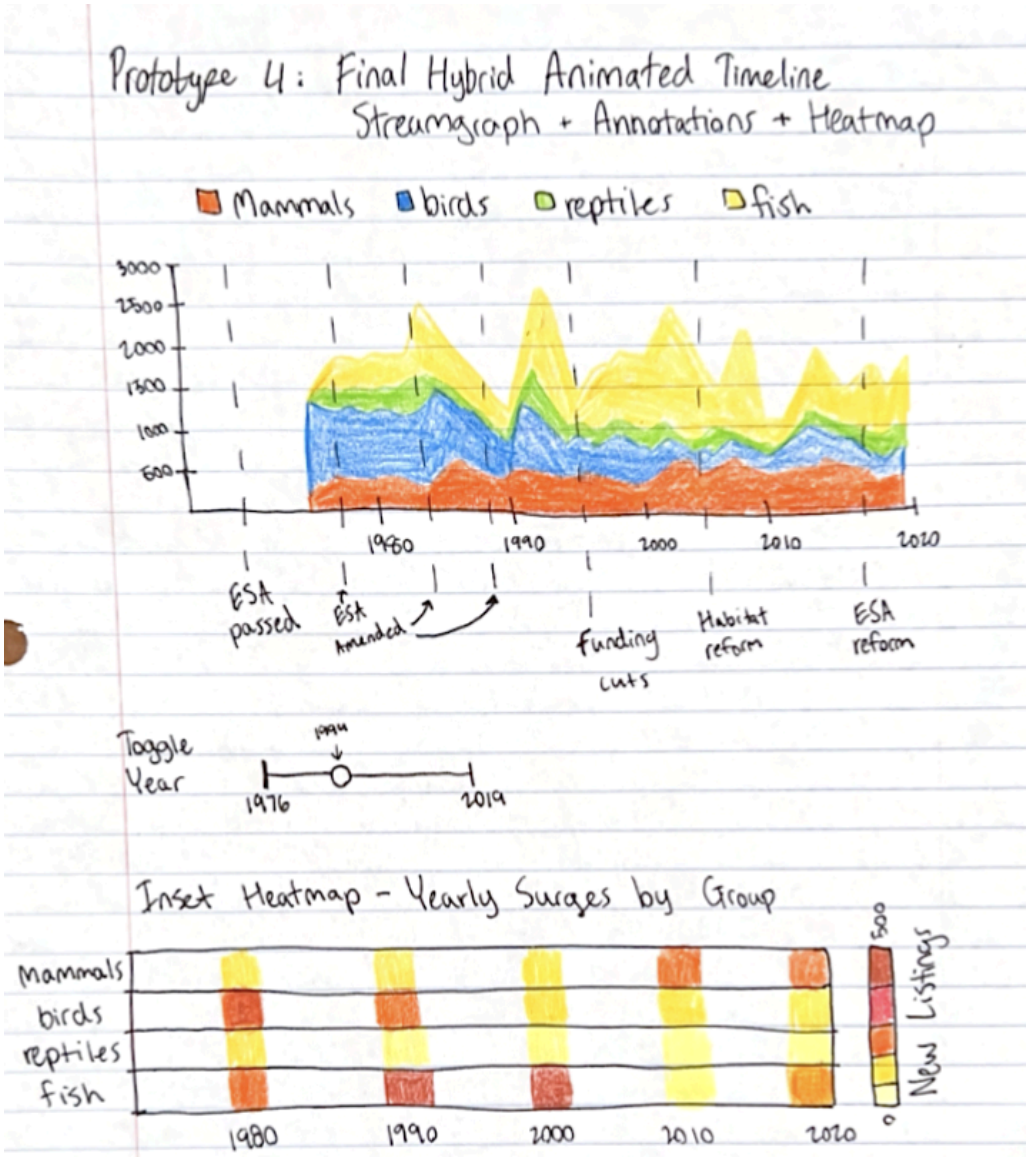
Prototype 3:



This bar chart shows the number of endangered species in a way that makes it easier to decipher which group makes up what proportion of the entire endangered animal population. I

only represented one bar per decade for clarity, but in reality, there would be one generated for each year between 1976 and 2019. These taxonomic groups are sorted by different distinct colorations. This prototype just shows a few selected representative taxonomic groups for clarity, but the real visualization will include them all. This graph also includes markers where relevant historical events occurred to give more context behind the rise and fall of endangered species.

Prototype 4:



This is the final combination of these previous visualizations. 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. This is very important to us, as we want to blend data visualization with storytelling, not only representing and analyzing the true data, but 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.

An optional feature would be an interactive map of the United States. If we are able to find data that categorizes the endangered species by state, we could use this to make a map. A slider at the top could go from 1976 to 2019. Clicking on each state would show the category of species and the total number of endangered species within the specified year range. Each state will be shaded darker or lighter depending on the number of endangered species. Alternatively, if we were for some reason not able to implement this map for a user-entered time range, we could also just zone in on one key year. If we are able to find the right corroborating geographical data, creating this additional visualization would provide further unique insights.

We plan to primarily meet and pair program the vast majority of this project, rather than divvying up specific tasks. Here is our tentative project schedule from here on out:

Week 5

- Complete project review with course staff.
- Begin initial coding setup and project structure.

Week 6

- Start early data cleaning and processing work.
- Set up a draft structure for the process book.

Week 7

- Complete dataset processing. Midterm exam this week, so focus just on finishing data prep.

Week 8/9

- Create simple exploratory visualizations.
- Expand on visualization drafts (begin streamgraph, heatmap, annotated timeline, etc.).

Week 10

- Have most visualizations drafted (may still have bugs/imperfections).
- Experiment with adding geographic/contextual data to supplement the existing dataset.

Week 11/12 (Milestone)

- Refine and polish initial visualizations.
- Integrate interactivity (filters, tooltips, toggles).
- Submit milestone with working prototype and current process book.

Week 13

- Focus on website implementation (embedding visualizations, layout design).
- Continue debugging and testing interactive features.

Week 14

- Complete website functionality.
- Upload visuals, process book, and data to GitHub Pages.

Week 15/16 (Final Submission)

- Record and edit a screencast demo.
- Incorporate peer feedback to make final tweaks to visualizations and site design.
- Submit the final project :)