Biodiversity - Elephant in the room

Keerthana Manoharan (km734), Rocio Perez (rp529), Karan Rajpal (kr377)

Contents

Introduction

Our Earth is the only planet in the universe known to harbor life. Radiometric dating and other sources date the age of our Earth to be about 4.54 billion years. Over a period of four billion years, various species have come into existence on this planet – from simple life to complex organisms. Throughout history, millions of species have thrived and millions have gone extinct. Today, however, we are at the brink of upending the fragile biodiversity!

As part of this project, we wanted to gain a deeper understanding of the issues that surround the endangered species, the threats they face in their habitats and where we can focus our efforts in bring balance to the biodiversity of our planet.

Data

Source

Our data source is the IUCN Red List of Threatened Species(http://www.iucnredlist.org/). They provide an API to fetch different data about species, regions, habitats, threats and conservation measures(http://apiv3.iucnredlist.org/).

For our project we needed data like species count per category, total species count and region where the threatened species were concentrated which were straightforward to fetch from the API.

We also wanted to show the total number of species that are affected by potential threats in each habitat spread across all the world - this was a lot more challenging!

Fetching Data
We used both Python and JavaScript to fetch data to see which works best.

The threats API provided all the threats faced by a specific species. What we needed was a count of the threats faced by each of the species in each of the threatened categories. To do this, we first fetched and stored all the species in these categories as an array of IDs.

Next, we made 25000+ API calls to obtain the count of threats. Our first version failed because we put them in a for loop and JavaScript initiates network calls as soon as the request is made - and NOT sequentially. After a limit of around 1500 network calls, the IUCN API started rejecting our requests. So we limited the number of concurrent calls we made, waited for them to finish execution, and then continued with the rest of the calls. This operation was done in a loop till we had all the data we required for our visualization. Our final version also incorporated a small timeout before each call to maximize the number of concurrent calls we could make.

We applied the same learning to our habitats data and thus fetched all the data we needed on threats and habitats.

Cleaning and Transforming Data
After fetching all the data, we discarded data that wasn't relevant for our visualization purposes. This involved removing data like timing, scope and severity from threats which were not part of our study. We applied the same method on the habitat data as well. This reduced the size of our data a lot and we could easily use the data to try out our visualizations.

Steps for cleaning Threats Data
Following is the procedure we have used to clean the data for all the Threats.

1) Loop through the list of species to fetch the threats from the RedList API for each one.
   API:
   http://apiv3.iucnredlist.org/api/v3/threats/species/id/'+species+'?token=9bb4facb6d23f48efbf424bb05c0c1ef1cf6f468393bc745d42179ac4aca5fee

   Returned Data


```
{
	"id": "43",
	"result": [{
		"code": "1.1",
		"title": "Housing & urban areas",
		"timing": "Ongoing",
		"scope": null,
		"severity": null,
		"score": null
	}, {
		"code": "7.2",
```

```
            "title": "Dams & water management/use",
            "timing": "Ongoing",
            "scope": null,
            "severity": null,
            "score": null
      },….]
}
```

2) Create an object for each category of species - “VU”, “CR” and “EN”. Store the result of each call in an Object whose key is the id of the species, under it’s relevant category.

Transformed Data

```
{
      "VU": {
            "43": [{
                  "code": "1.1",
                  "title": "Housing & urban areas",
                  "timing": "Ongoing",
                  "scope": null,
                  "severity": null,
                  "score": null
            }, {
                  "code": "7.2",
                  "title": "Dams & water management/use",
                  "timing": "Ongoing",
                  "scope": null,
                  "severity": null,
                  "score": null
            }, ….],
            "140": [{
                  "code": "1.1",
                  "title": "Housing & urban areas",
                  "timing": "Ongoing",
                  "scope": null,
                  "severity": null,
                  "score": "Low Impact: 3"
            }, {... }
      },
      "CR": {
                  "3": [{
                        "code": "5.3",
                        …..
                         }
                  ]
            }
        }
}
```

3) Remove unnecessary data and store only the Threat IDs for each species as an array.

Transformed Data


```
{
	"VU": {
		"43": [1, 7],
		"81": [11, 5, 7, 8, 9],
		"140": [1, 2, 3, 5],
		...
	}
}
```


4) Finally, we run a count to get the number of species under each category for each threat

Transformed Data


```
[
	{
		"Threat": "Residential/ Commercial",
		"CR": 1010,
		"EN": 1564,
		"VU": 2156
	},
	{
		"Threat": "Agriculture/ Aquaculture",
		"CR": 1947,
		"EN": 2990,
		"VU": 3575
	},
	….
]
```


Integration

We followed the same procedure for habitats. After we had our individual datasets for both threats and habitats, we had to map the data for threats and habitats into a single datasource for use with our heatmap. We wrote a script to match species, habitats and threats with species as a common key.

We now had 3 variables in a single datasource.

Datasets

We have gathered all our data into the following sources for each graph :

- **Sizes** - Size of each category of species along relevant information for plotting.
- **Threats** - List of all the threats. Each threat includes the number of species it affects in the 3 categories - CR, EN and VU.
- **Habitats** - List of all the habitats. Each habitat includes the number of species it affects in the 3 categories - CR, EN and VU.
- **Threat-Habitat** - The number of species affected in each habitat by each threat(We use 3 variables to build this dataset)
- **Countries** - List of all the countries with data for plotting the map.
- **Codes-iso** - List of all the country names along with their ISO specified country codes.
- **World-50** - Data to plot each country with d3.

Telling our story - Mapping data to visual elements

To show the scale of the problem we're talking about, we decided to use circles to compare the relative number of species in different categories. The 3 categories we are focussing on - Vulnerable, Endangered and Critically Endangered, are all highlighted in increasing shades of red which indicates the level of danger of the situation.

Visualization 1 - Circles
We chose to use sequentially laid circles to indicate the total number of species in each category. Shown next to each other, this brings out the extent to which various species are affected with respect to the total number of species. We have used a sqrt scale which best represents the sizes of circles based on its radius. The last circle that has all the species also includes the species which do not come under the previous 7 classifications of the Red List - these are the species which IUCN does not have sufficient data or they have not been evaluated yet.

Visualization 2 - World Map
We chose a scale of 9 multi-hues in order to show the country-wise distribution of the total number of endangered species across the globe. The darker colours correspond to a higher concentration of threatened species in that area.

Visualization 3&4 - Stacked Bar Chart
We chose stacked bar chart so that we could make our visual representation comparable in term of the total number of species per category that are affected in each habitat .

Visualization 5 - Heatmap
In this graph, we wanted to present data on three different variables - threats, habitats, and the total number of species that is affected by these threats in each habitat. Clustering of data, which is inherent of heat maps, helped us analyze the most critical threat that affected a huge population of the species in the most popular habitat.

Story

We initially thought that the biggest threats to Biodiversity would be more along the lines of Climate change and Pollution. We were surprised to find that these threats were not the largest, and instead the largest threats were Agriculture/Aquaculture and Biological Resource Use.

We expected the largest number of threatened species to be focussed in the Forests and this was true. Going deeper, we found that the threats were also mostly affecting species in the Forest habitat.

This means that going forward we need to focus our efforts into tackling issues like :

- Illegal trade of Wildlife
- Hunting and Poaching
- Overfishing
- Agricultural Methods

References

We have taken inspiration from https://bl.ocks.org/
- https://bl.ocks.org/mbostock/3886208
- http://bl.ocks.org/tjdecke/5558084
-More specifically, we have used code snippets to achieve word wrapping from
https://bl.ocks.org/mbostock/7555321

Data Source: http://www.iucnredlist.org/
http://stackoverflow.com