National Park Biodiversity

Process Methodology

Data Preparation

- Tidy Dataset
- Identify the data related to selected problems
- Clean and standardize data

Analyze

- Calculate overall proportion
- Group by different attribute (park type, size,...) and analyze related attribute (location, Parkcode, regional information)
- Draft some visualizations showing relationships

Visualization

- Select appropriate chart or graphs
- Visualize data on selected charts and make it more interactive.
- Optimized charts to support more for the analysis

Reporting

- Analyze collected data, report findings
- Report & Document

Data Preparation

ParkCode	ParkName	CategoryName	Order	Family	TaxonRecordStatus	SciName	CommonNames	Synonyms	ParkAccepted	Sensitive	RecordStatus	Occurrence	OccurrenceTags	Nativeness	NativenessTags	Abund
ACAD	Acadia National Park	Mammal	Artiodactyla	Cervidae	Active	Alces alces	Moose	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Rare
ACAD	Acadia National Park	Mammal	Artiodactyla	Cervidae	Active	Odocoileus virginianus	Northern White-tailed Deer	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Abund
ACAD	Acadia National Park	Mammal	Carnivora	Canidae	Active	Canis latrans	Coyote	N/A	TRUE	FALSE	Approved	Present	N/A	Non-native	N/A	Comm
ACAD	Acadia National Park	Mammal	Carnivora	Canidae	Active	Canis lupus	Eastern Timber Wolf	N/A	TRUE	FALSE	Approved	Unconfirmed	N/A	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Canidae	Active	Vulpes vulpes	Black Fox	N/A	TRUE	FALSE	Approved	Present	N/A	Unknown	N/A	Comm
ACAD	Acadia National Park	Mammal	Carnivora	Felidae	Active	Lynx canadensis	Canada Lynx	N/A	TRUE	FALSE	Approved	Unconfirmed	N/A	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Felidae	Active	Lynx rufus	Bay Lynx	N/A	TRUE	FALSE	Approved	Present	N/A	Non-native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Mephitidae	Active	Mephitis mephitis	Eastern Skunk	N/A	TRUE	FALSE	Approved	Not In Park	Historical	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Active	Lutra canadensis	Otter	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Comm
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Active	Martes pennanti	Blackcat	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Rare
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Active	Mustela	Weasel	N/A	FALSE	FALSE	In Review	Unconfirmed	N/A	Unknown	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Active	Mustela erminea	Bonaparte Weasel	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Uncom
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Active	Mustela frenata	Long-tailed Weasel	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Uncom
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Inactive	Mustela macrodon	Ancient Sea Mink	N/A	TRUE	FALSE	Approved	Unconfirmed	N/A	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Mustelidae	Active	Mustela vison	American Mink	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Uncom
ACAD	Acadia National Park	Mammal	Carnivora	Odobenidae	Active	Odobenus rosmarus	Walrus	N/A	TRUE	FALSE	Approved	Unconfirmed	N/A	Non-native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Phocidae	Active	Halichoerus grypus	Gray Seal	N/A	TRUE	FALSE	Approved	Unconfirmed	N/A	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Phocidae	Active	Phoca vitulina	Common Seal	N/A	TRUE	FALSE	Approved	Unconfirmed	N/A	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Carnivora	Procyonidae	Active	Procyon lotor	Common Raccoon	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Comm
ACAD	Acadia National Park	Mammal	Carnivora	Ursidae	Active	Ursus americanus	Black Bear	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Occas
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Eptesicus fuscus	Big Brown Bat	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Uncom
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Lasionycteris noctivagans	Silver-haired Bat	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Unkno
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Lasiurus borealis	Eastern Red Bat	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Occas
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Lasiurus cinereus	Hoary Bat	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Unkno
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Myotis keenii	Keen's Myotis	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Comm
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Myotis leibii	Eastern Small-footed Myotis	N/A	TRUE	FALSE	Approved	Not In Park	Historical	Native	N/A	N/A
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Myotis lucifugus	Little Brown Bat	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Comm
ACAD	Acadia National Park	Mammal	Chiroptera	Vespertilionidae	Active	Pipistrellus subflavus	Eastern Pipistrelle	N/A	TRUE	FALSE	Approved	Not In Park	False Report	Non-native	N/A	N/A
ACAD	Acadia National Park	Mammal	Lagomorpha	Leporidae	Active	Lepus americanus	Snowshoe Hare	N/A	TRUE	FALSE	Approved	Present	N/A	Native	N/A	Comm

ParkName: National Park Full Name.
CategoryName: Species Category.
CommonNames: Common name of the species.
Abundance: How abundant is the species in the park.
References: Display the number of associated evidence records

Data Preparation

01 **Extract important data columns from the dataset**
Important columns for analysis include: 'ParkName', 'CategoryName', 'CommonNames', 'References'

02 **Handled missing values in the dataset**
E.g.: Filling null values in the "Abundance" column with "Unknown" to maintain data integrity.

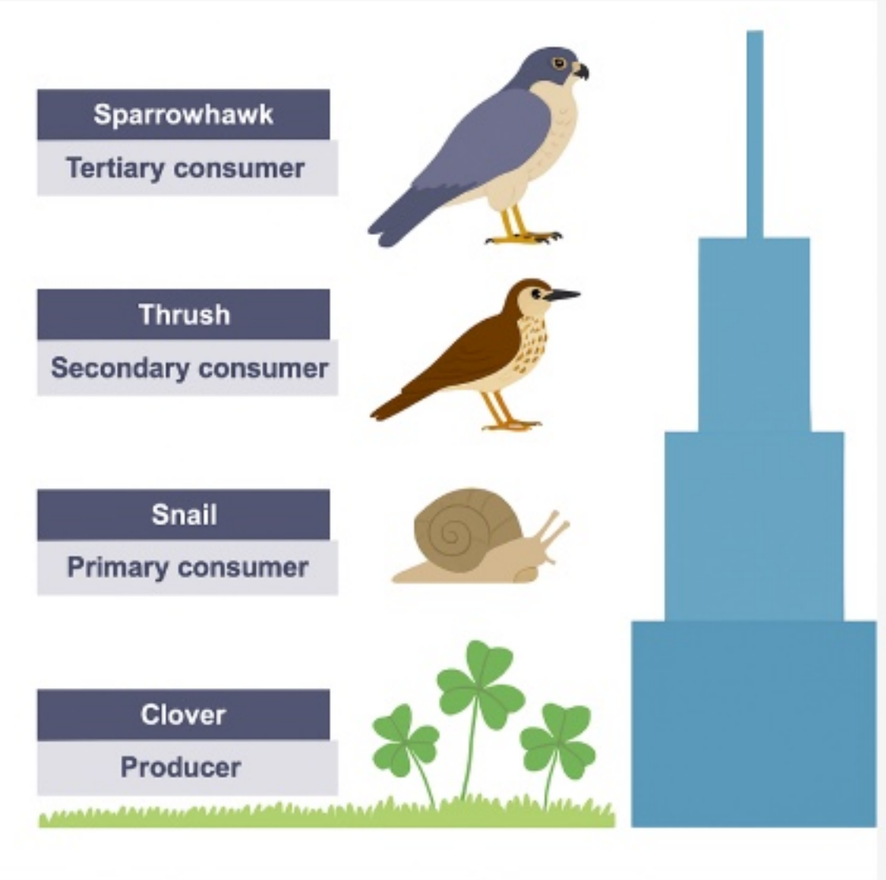
03 **Standardized data formats across all fields**
E.g.: converting all scientific names (SciName) to proper capitalization format for consistency.

04 **Match species from 'CategoryName' to food chain level**
E.g. "Vascular Plant" is level 1 (Lowest), “Mammal” is level 6 (Highest)

05 **Plot regression line showing species number trend**

06 **Normalized species counts across parks.**
E.g.: calculating percentage composition of each species category (e.g., Vascular Plants comprise 80.7% of Yosemite's species).

Question 1
Is the species data from 15 most visited park reflect pyramid of number/biomass (Food chain pyramid)?

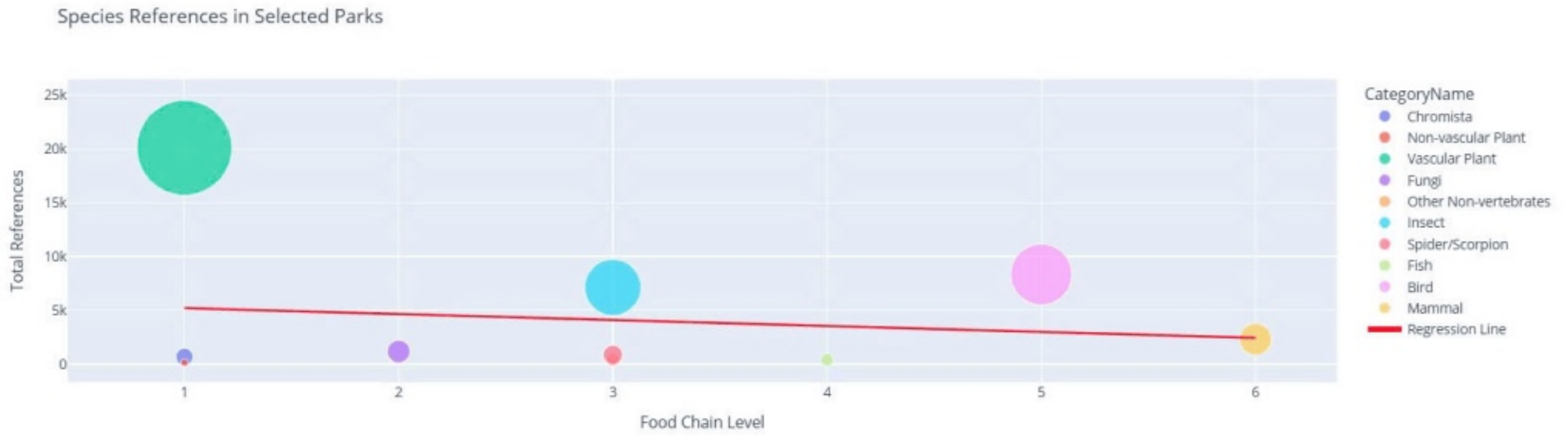


Why?

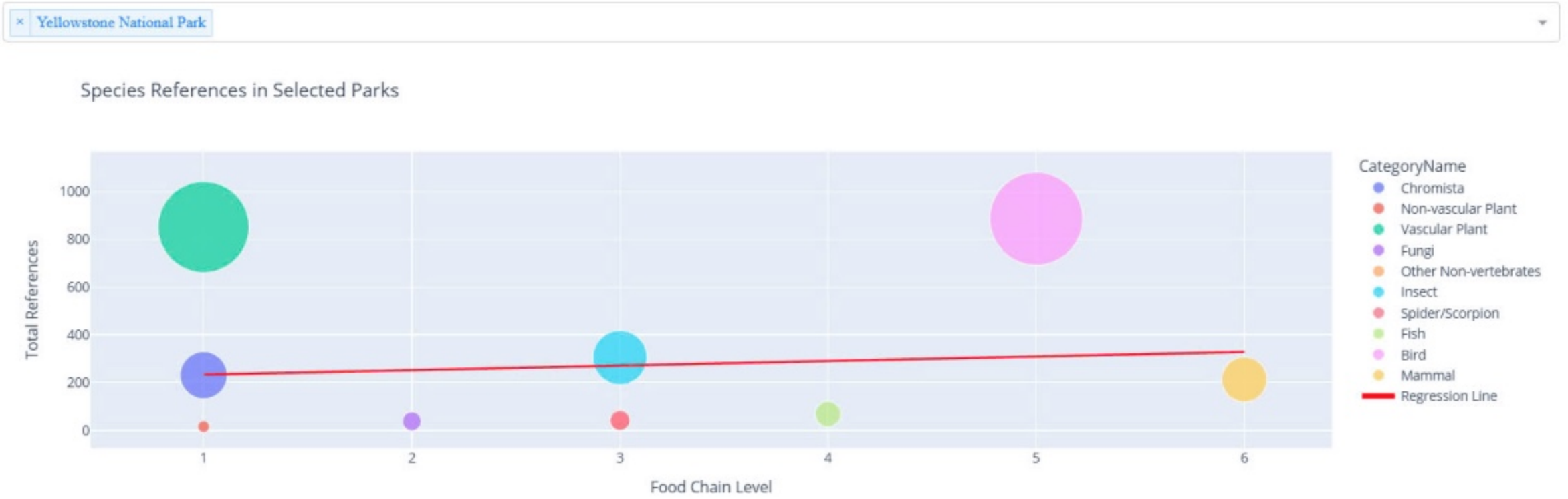
A pyramid of numbers shows the total number of individual organisms at each level in the food chain of an ecosystem. To put it simple, it means that species higher up in the food chain will usually be less abundant.

National Park Species References

× Yellowstone National Park × Great Smoky Mountains National Park × Acadia National Park × Glacier National Park × Bryce Canyon National Park × Rocky Mountain National Park × Joshua Tree National Park

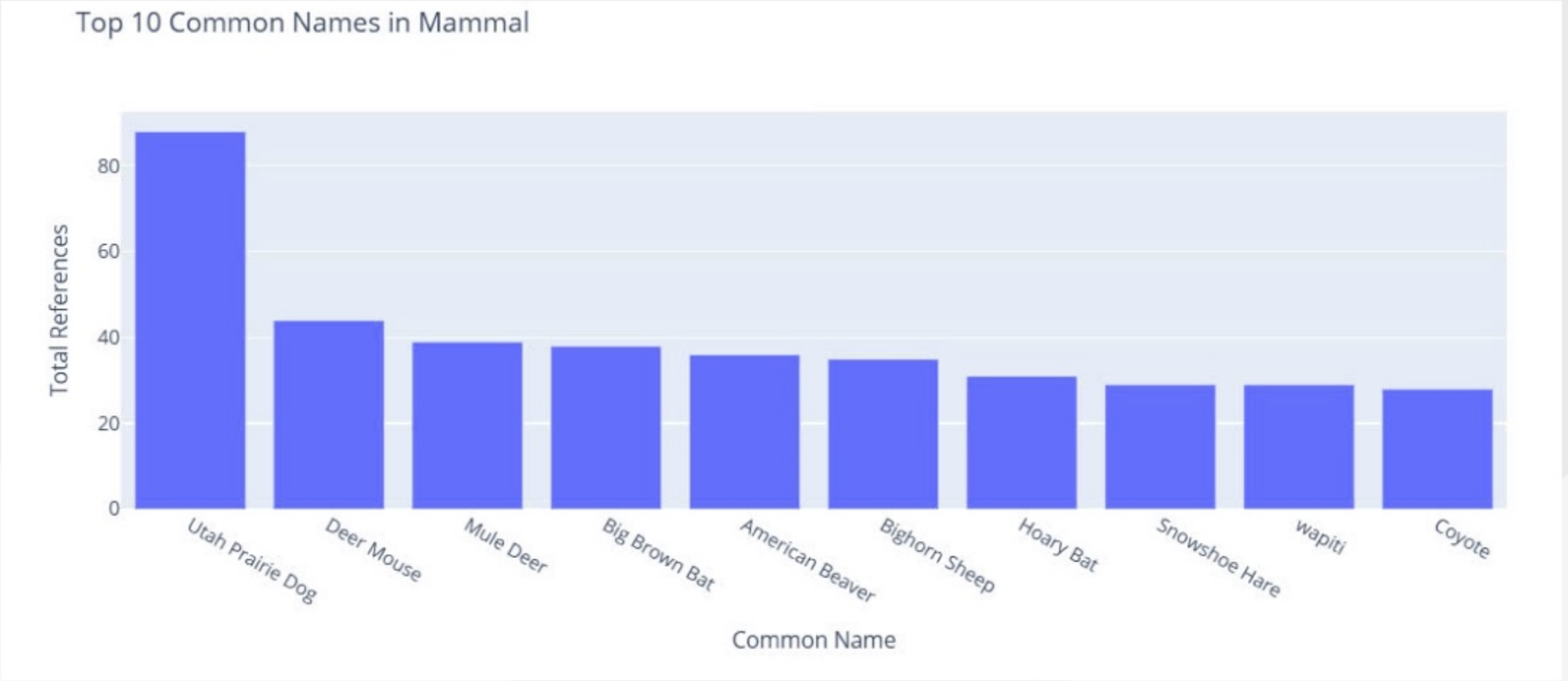


National Park Species References



Scatter Plot

- Shows trends in species numbers according to pyramid of numbers, the line should be downward to match with the Pyramid
- Some outlier park like Yellowstone national Park show the trend upward, but when filter multiple park, the regression line prone to be downward like the pyramid of number
- The number of Bird and Vascular Plant usually higher than other Category, showing bias in recording species numbers



Bar chart of specific Category

- Show the top 10 most appeared species Names for specific Category
- Some outliers such as mule dear or snowshoe hare are listed as Mammal and are at the top of the food chain incorrectly, suggesting that a more precise classification is needed to properly order species in the food chain.

Question 2
What proportion of species are endangered, threatened, or of special concern?

How?

We'll examine if rare species proportions relate to park type or geographic location.

Explore the Dataset

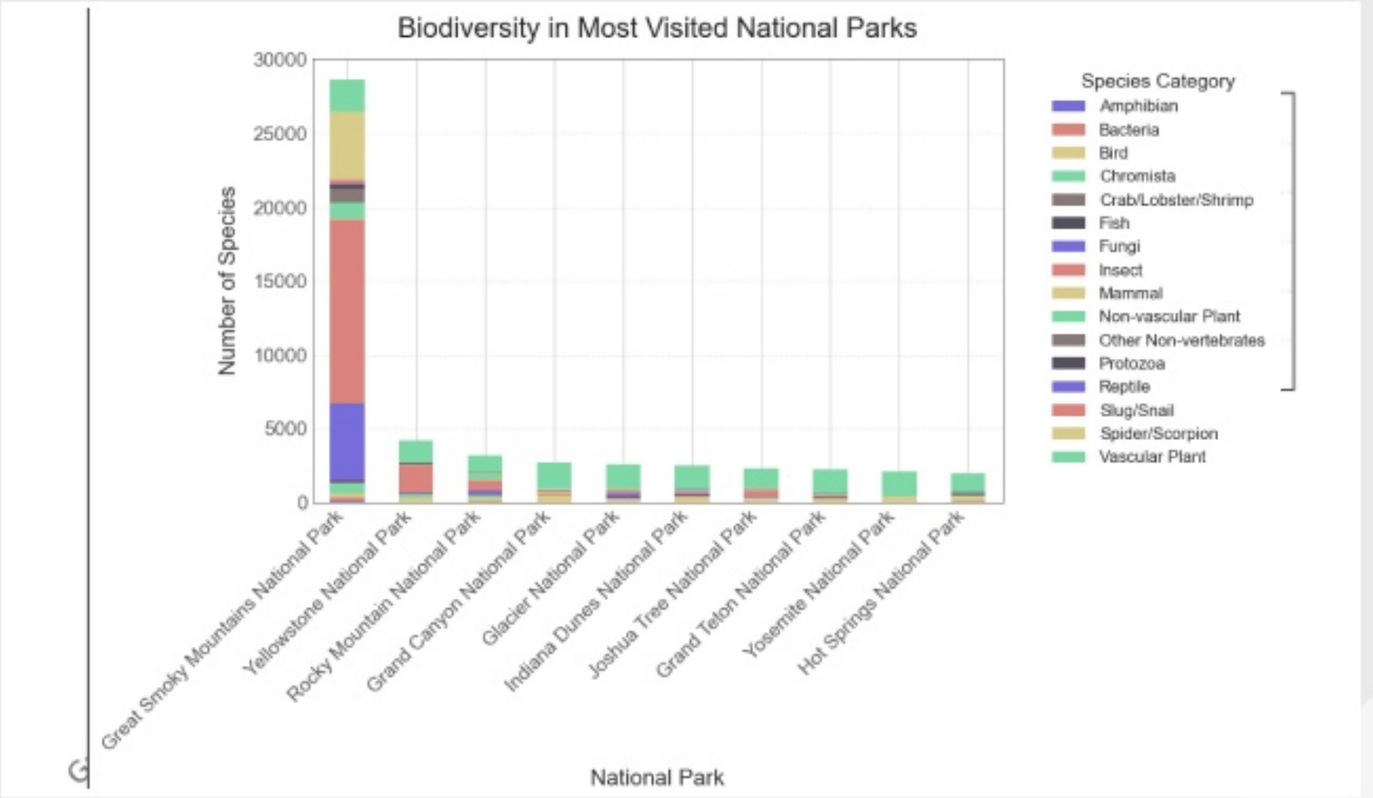
Dataset:
61,119 rows, 28 columns.

Key Columns:
Conservation status (TEStatus, StateStatus), park info (ParkCode, ParkName), species info (SciName).

Findings:

- Many missing values in conservation status.
- Common statuses: SC (Species of Concern), RT (Rare/Threatened).
- Vascular Plants and Insects are the most common species.

Park Representation:
Multiple national parks with unique codes, many missing ParkTags - 30.773 out of 61.119 rows (50.35%)

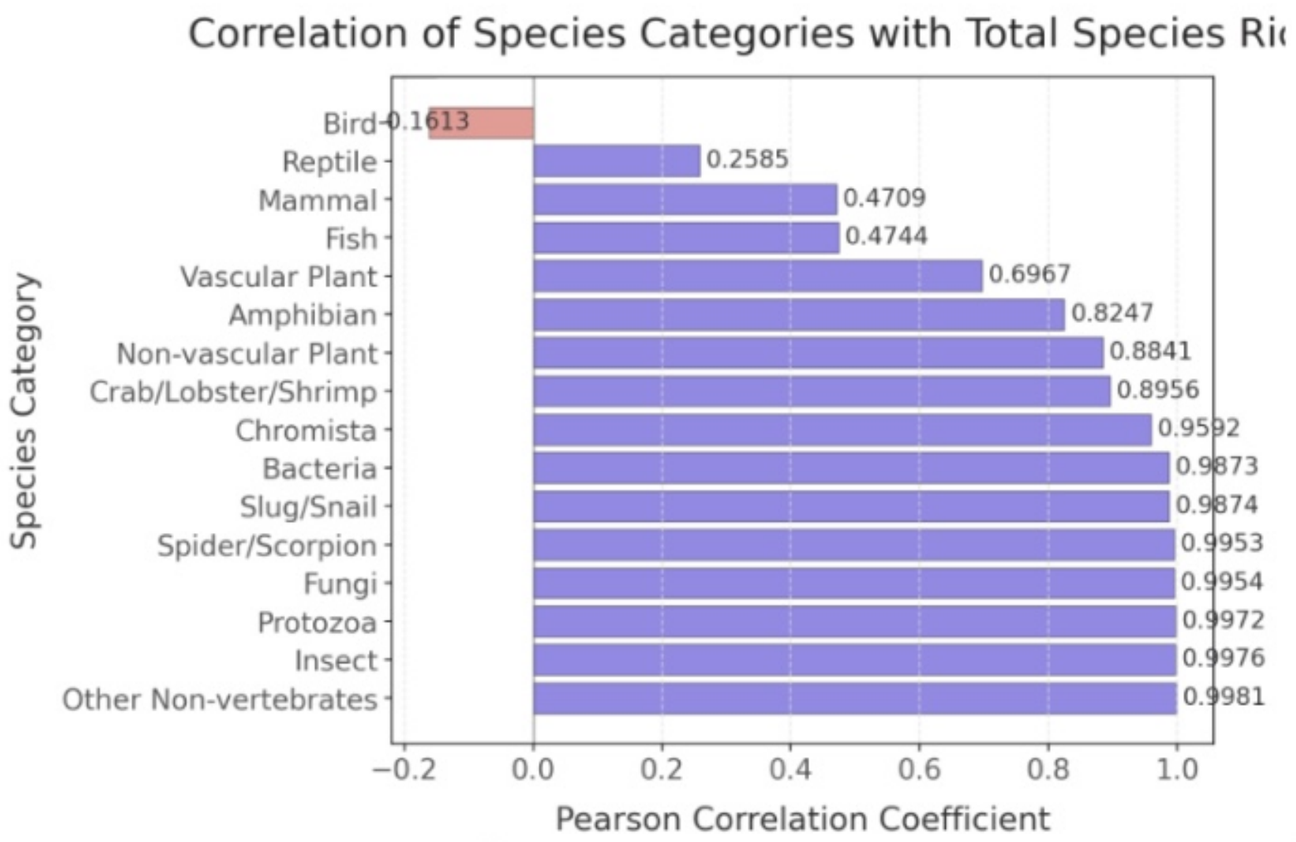


Normalization
Use percentages for species groups instead of absolute numbers.

Conservation Value
Highlights the importance of Great Smoky Mountains in U.S. biodiversity preservation.

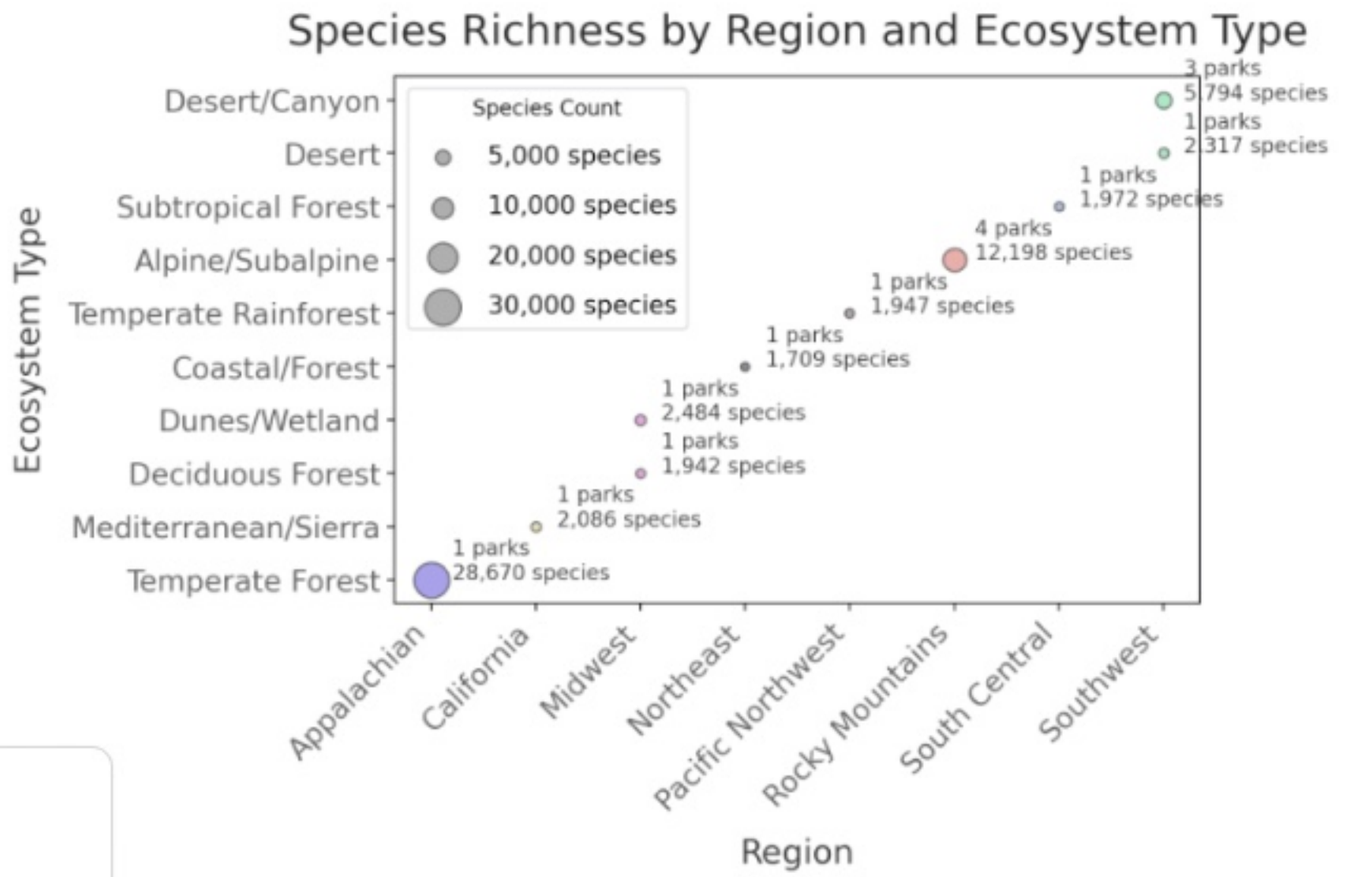
Geographic/Ecological Factors
Consider park location and ecosystem (e.g., Great Smoky Mountains' biodiversity in the Appalachian region).

2. Correlation of Species Categories with Total Biodiversity



Correlation Insights
Insects: Strongest positive correlation with biodiversity (0.9976).
Bird: Negative correlation (-0.1613), indicating high bird diversity may coincide with lower overall species counts.

3. Species Richness by Region and Ecosystem Type



Bubble Chart Summary
Purpose: Shows biodiversity across regions and ecosystems.
Highlights:
Appalachian (Great Smoky Mountains): High species richness in temperate forests.
Rocky Mountains: Notable biodiversity in Alpine/Subalpine ecosystems.
Visual: Bubble size indicates total species count.


Thank you!