NATIONAL PARKS

A Data Analysis by Jared Baker

MOTIVATION

I've always been deeply interested in nature and have had the opportunity to visit several national parks. These places are truly remarkable. For my capstone project, I seized the opportunity to combine my interest in the natural wonders of the United States with data analytics. After initial research, I decided to drill down into the how characteristics of the parks affect their visitation.

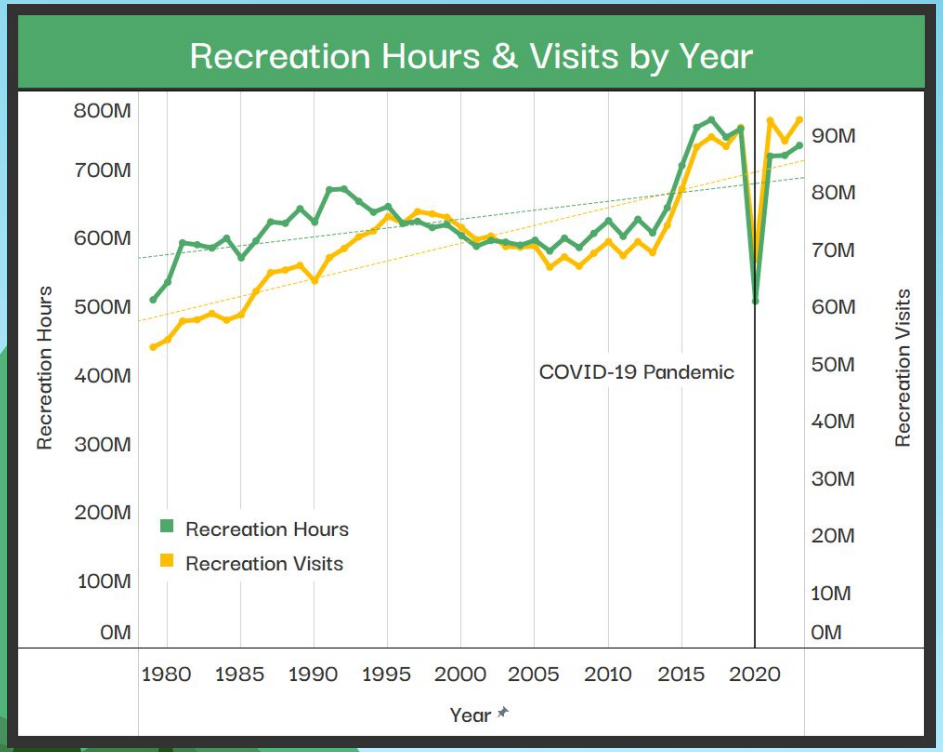
How do time, biodiversity, and climate relate to recreational visitation at national parks?

Term Definitions

- **Visit** - One entrance per individual per day may be counted.
- **Visitor Hours** - Aggregate time spent by visitors in a park.
- **Recreational** - Visits not using park territory, roads, and facilities for their own convenience or as a part of their occupation.

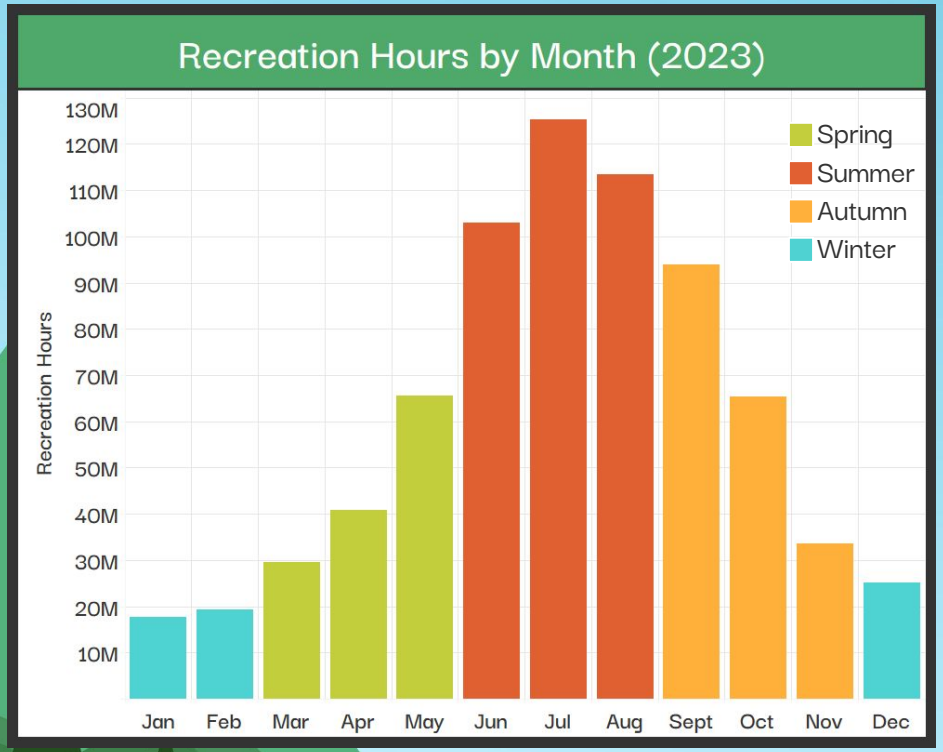
Data Sources

- **National Park Service (NPS)** - IRMA Data Portal
- **National Oceanic and Atmospheric Administration (NOAA)** - Climate Normals Station Query Tool
- **Wikipedia** - Web Scrape



TIME

- Recreation Hours **increased by ~225M (44%)** from 1979 to 2023
- Recreation Visits **increased by ~40M (76%)** from 1979 to 2023
- Over time, Recreation Visitors are **spending less time** per visit
- In 2020, Recreation Hours hit its **lowest point** in **41 years**; likely due to the COVID-19 pandemic

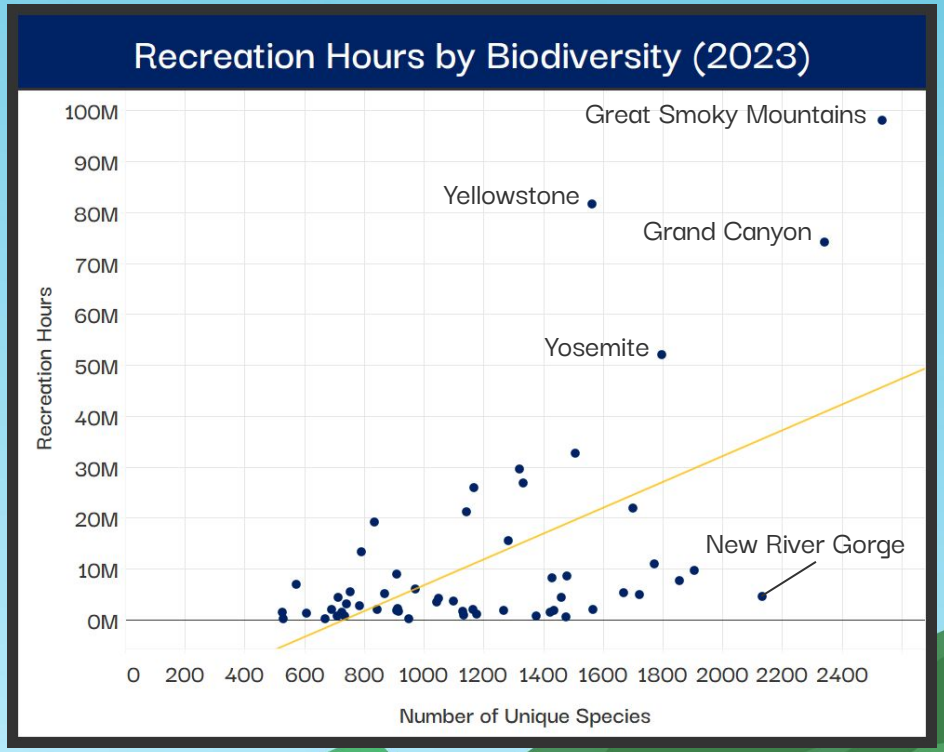


TIME

- The **most popular** month to visit a national park is **July**
- The **least popular** month to visit a national park is **January**
- In 2023, national parks were **5.5 times busier** in summer than winter

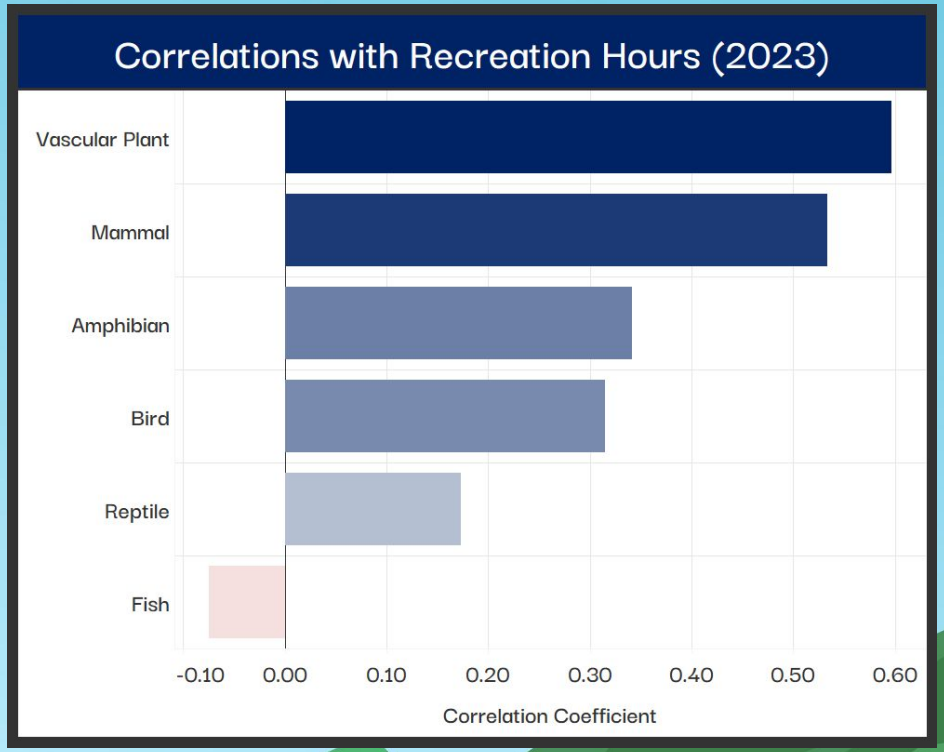
BIODIVERSITY

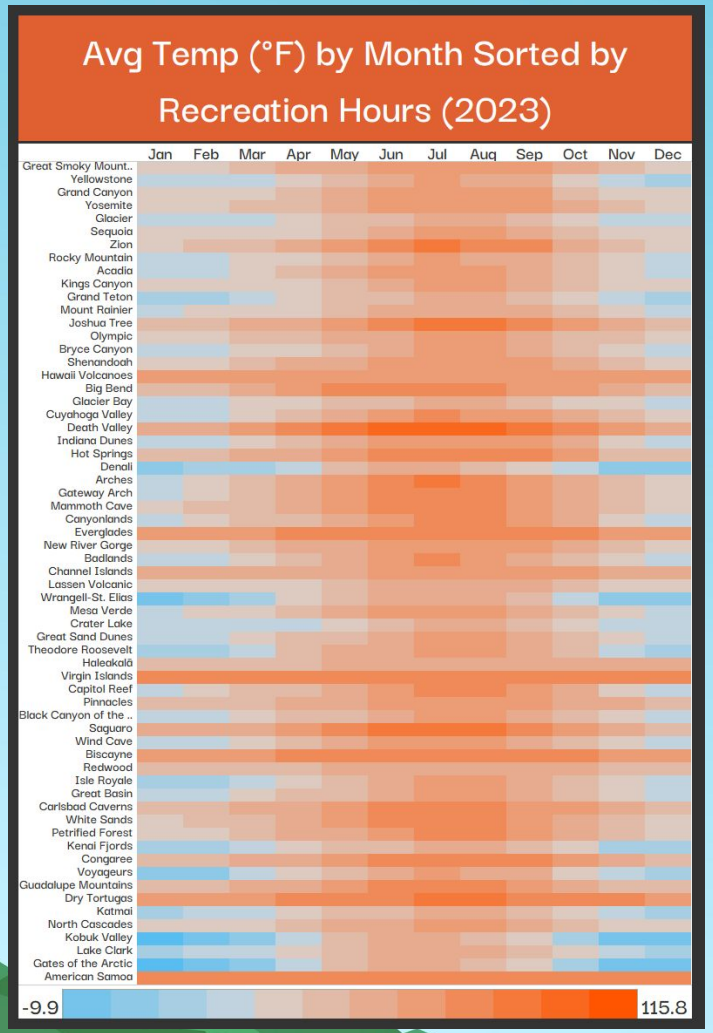
- Parks with with **more biodiversity** tend to have **more visitation**
- Correlation coefficient between Recreation Hours and Number of Unique Species: **0.58 (Moderate)**
- The Great Smoky Mountains had **2,535 species** recorded as present in 2023



BIODIVERSITY

- Vascular Plants and Mammals have a **moderate correlation** with Recreation Hours
- While **negligible** in strength, fish are the only species category to have a **negative correlation** with Recreation Hours





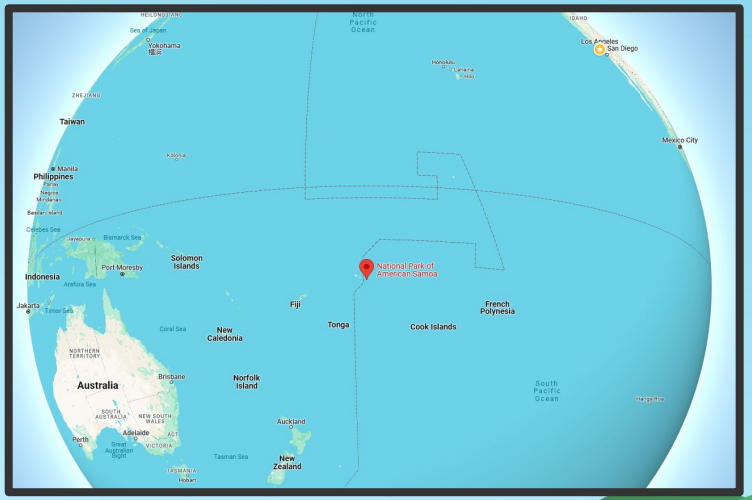
Most

Recreation Hours

Least

CLIMATE

- Climate has **little** relationship with Recreation Hours
- The coldest parks tend to have the least Recreation Hours, but that is likely due to remote location



How do time, biodiversity, and climate relate to recreational visitation at national parks?

Recreational visitation is:

- **Increasing over time,** with summer being the busiest season in the year
- **Higher** in parks with more biodiversity, with plants and mammals having the strongest correlation
- **Not** correlated with the local climate of the parks

Future questions

- Why are recreational visits shortening over time?
- Which park and time of year combination is the best option for an individual?
    - Activities
    - Weather
    - Location

CHALLENGES

- **Climate data**
  - Many parks did not have a weather station with both temperature and precipitation data; some had no station at all. For these parks, I located the nearest weather station, still easily representative of the local climate, and used that station's data. This data collection had to be done manually, one park at a time, but I was able to automate the cleaning and merge the data easily.
  - Snow data wasn't collected on many weather stations. Some obviously never get snow, others do but were missing the sensor. Snow data was excluded from the analysis
- **NPS Data**
  - **Visitation**
    - I spent a considerable amount of time manually cleaning excel sheets of data unnecessarily, and later realized there was a separate query building tool in the data portal that I used to create the exact CSV files I needed. I learned to focus more time on understanding an unfamiliar database fully before beginning work with it.
  - **Entrance Fees**
    - This data was not readily downloadable or web-scrapable without spending a full day on it, so I manually entered the information over about 15 minutes. I believe this was the right choice, and prioritized meeting the deadline without overcomplicating things. I am here to learn, but a deadline is a deadline, and the juice wasn't worth the squeeze.
    - Unfortunately this information did not yield any significant findings, and was ultimately excluded for time.
  - **Acreage**
    - This data had a some interesting facts, such as Alaska having more than half of all national park land, but ultimately did not lead to significant findings. It was excluded from the presentation for time.
  - **Biodiversity**
    - Certain parks lacked any species data, and were excluded from biodiversity-related analysis.

THANK YOU