U.S. National Park Visit Data (1979-2023)

Melanie Walsh and Os Keyes

2024-06-01

Data Essay

Introduction

This dataset contains the number of visits, per year, to each of the current 63 National Parks administered by the United States National Park Service (NPS), from 1979 to 2023. The NPS also collects visitation and use data about other park units, such as national battlefields, national rivers, and national monuments. However, information about other park units is not included in this particular dataset.

Brief Survey

If you use our materials in your class or another setting, we would love to hear about it!

View Summary of Columns

```
//|echo: false
viewof selectedColumns
viewof dataSummaryView
```


The National Park datasets included here are drawn from data published by the U.S. NPS, and most (but not all) of the contextual information is drawn from material published by the NPS.

We decided to publish this version of the data, along with our own synthesized documentation and narrative, because the original data is made available in an NPS data portal that is

relatively hard to find and to use, and the documentation is distributed across many different web pages, PDFs, and other documents. (The NPS has created an interactive Microsoft Power BI dashboard to help users explore the data more easily.)

The datasets were curated and published by Melanie Walsh, and the data essay was written by Melanie Walsh and Os Keyes.

History

A national park is an area of land that a country's government deems important enough to officially protect, preserve, and make available to the public. There are thousands of national parks around the world (some of which are featured in the Netflix documentary, "Our Great National Parks," narrated by former President Barack Obama).

In the United States, the very first National Park—Yellowstone National Park, in Wyoming—was signed into law in 1872 by President Ulysses S. Grant.

Figure 1: Old Faithful, the most famous geyser of the whopping ~500 geysers at Yellowstone National Park. Photo credit: NPS/Neal Herbert.

Over the next several decades, a handful of other parks—such as Sequoia (1890), Yosemite (1890), Mount Rainier (1899), and Crater Lake (1902)—joined the system, too.

> What is the most recent National Park?
>
> The most recently added National Park is New River Gorge National Park in West Virginia. It was designated in 2020.

While the National Parks were originally created to protect precious, beautiful lands and to make them accessible to everyday people—a noble goal—it is important to remember that

Figure 2: Mount Rainier, also known by the Indigenous name Tahoma, is an active volcano and 14,411 feet tall. Mount Rainier National Park, which is 60 miles south-east of Seattle, Washington, was founded in 1899. Photo credit: NPS (public domain).

many of these lands were taken, sometimes forcibly, from Native American people who already owned, lived, and worked on them (Spence 2000; Beauchamp 2020). Today, there are still calls for the NPS to return the lands of the National Parks to Indigenous people.

In a similar vein, scholars have shown that early environmental conservation movements—movements that helped to spur the development of the National Parks—were troublingly intertwined with racism and eugenics movements (Beauchamp 2020). These prejudiced origins, combined with continuing forms of environmental racism (e.g., many parks are located far from cities, with limited public transporation options and limited community outreach), have contributed to the marginalization of people of color and other minorities in the parks. Research has shown that white people visit the parks more than other racial groups (Weber and Sultana 2013; Alba et al. 2022; Floyd and Johnson 2002). So while the National Parks are technically open to everyone, they are not equally accessible to everyone in the same way. And these exclusions shape the parks' visitation data even before it's counted.

So when and why did visit counting start at the U.S. National Parks? Well, according to the NPS, the counting of park visits started as early as 1904 (more than 10 years before the National Park Service itself was officially created). But at this time, and for the next 50 years or so, their data collection methods were mostly informal, inconsistent, and low-tech.

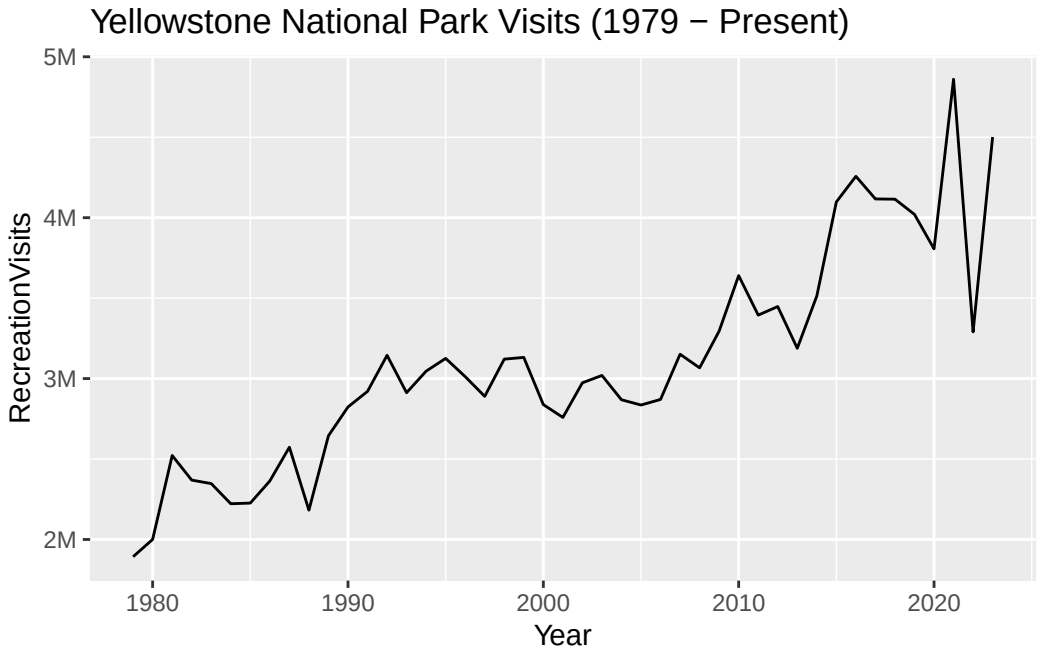
But in 1965, the NPS started getting serious about counting visits. That year, the U.S. Congress passed The Land and Water Conservation Fund Act of 1965. This act created a new source of government money specifically dedicated to protecting natural resources and

expanding outdoor recreation infrastructure. Because the act stipulated that the amount of money allocated to each area should be "proportional to visitor use," the NPS buckled down on counting visitor use. They "developed and institutionalized a formal system for collecting, compiling and reporting visitor use data."

In 1979, the NPS comprehensively changed their counting procedure, and all parks began tracking vistor use by month (as opposed to year) across 11 different statistics. This is why the datasets featured here begin in 1979.[1] **Note: We aggregated monthly counts into yearly counts for the dataset featured in this essay. A dataset with visit counts by month is available in "Explore the Data."**

```
# Note on installation: https://statsandr.com/blog/an-efficient-way-to-install-and-load-r-pac

# Load the dplyr package for data manipulation
# Load the ggplot2 package for data visualization
# Load "ggthemes", which let's us use colorblind-compatible palettes. When we've only got one
# Load "scales" for abbreviating axis labels
library(dplyr, warn = FALSE)
library(ggplot2)
library(ggthemes)
library("scales")

# Load National Park Visitation data
np_data <- read.csv("https://raw.githubusercontent.com/melaniewalsh/responsible-datasets-in-c

# Specify the colorblind palette
cb_palette <- colorblind_pal()(8)

# Turn off scientific notation
options(scipen = 999)

# Filter down to Yellowstone National Park
yellowstone_data <- np_data %>% filter(ParkName == "Yellowstone NP")

# Visualise it
ggplot(data = yellowstone_data) +
  geom_line(aes(x = Year, y = RecreationVisits), color = cb_palette[1]) +
  labs(title = "Yellowstone National Park Visits (1979 - Present)") +
  # abbreviate numbers by millions and thousands
  scale_y_continuous(labels = label_number(scale_cut = cut_short_scale()))
```

[1]The NPS also offers annual visitation information between 1904-1979, but it is a separate, less consistent dataset.



While today's National Park data collection system is more formal and sophisticated than the one that the NPS used in 1904, there are still many inconsistencies, flaws, and limitations (as the NPS openly acknowledges). This data does *not* represent the *exact* number of people who visited the parks in the last 50 years—hardly! Think about how difficult it would be to count every single one of the millions of people who walked, hiked, backpacked, drove, shuttled, canoed, biked, or skied into each of the 63 different parks since 1979. These parks are located in dozens of different geographic areas, including mountains, volcanoes, deserts, canyons, wetlands, forests, and islands; the parks have experienced countless different weather conditions during this time, including blizzards, hurricanes, wildfires, avalanches, and extreme heat; and the parks have also been allocated varying amounts of money and staff members to do the counting. Given all this variability, it is simply not possible to count every single visit to every single National Park ever.

We believe the National Park visit data is useful to study and consider precisely for this reason: because it helps demonstrate that **data never reflects reality precisely**. It also demonstrates that collecting and analyzing data, even when it is flawed and approximate, is sometimes worthwhile—but only if you fully understand the data's flaws, limitations, and history, and only if you incorporate these considerations into all subsequent analyses, interpretations, and takeaways.

Where did the data come from? Who collected it?

The National Park data on this website was originally organized and published by the NPS Social Science Program, which in turn runs the NPS Visitor Use Statistics program, an initia-

tive that coordinates visitor use statistics across the parks. Thousands of staff members across all 63 parks were also involved in the data collection process.

According to the NPS, the Visitor Use Statistics program aims to:

- Provide a statistically valid, reliable, and uniform method of collecting and reporting visitor use data for each independent unit administered by the NPS
- Support regular collection, and timely publication, analysis and interpretation of these data
- Enact quality control checks, verify measurements, and ensure consistency and comparability of data among areas of the NPS

We accessed the original data through the NPS's Visitor Use Statistics data portal, which publishes visit use data in alignment with the program's stated goals. Through this portal, anyone can generate reports and download data for different visit use categories and time periods—at both national and individual park levels.

To download the data included here, we first selected "National Reports" in the data portal, and we then selected the "Query Builder for Public Use Statistics (1979 - Last Calendar Year)" report type. Here are the selections we made:

- For "Park Types," we selected only "National Parks."
- For "Years," we selected all possible years (1979-2023).
- For "Regions," we selected all possible regions.
- For "Field Type," we selected only "Recreation Visits" (excluding the other 10 possible statistics: "NonRecreation Visits," "Recreation Hours," "NonRecreation Hours," "Concessioner Lodging," "Concessioner Camping," "Tent Campers," "RV Campers," "Backcountry Campers," "NonRecreation Overnight Stays," and "Miscellaneous Overnight Stays").
- For "Additional Fields," we selected "State" and "Region."
- We also selected the option of viewing the report as an annual summary of visit counts (as opposed to monthly visit counts).

If you choose to download this report as a CSV file, it will unfortunately not look exactly like the report pictured in Figure **??**; instead, the CSV will include all visit and use types, and it will include visit and use information by month rather than by year. When I (Melanie Walsh) have compiled this data to share with my students in the past, I have sometimes downloaded the CSV file, removed the columns that I'm not interested in, and aggregated the data by year programatically. In other cases, I have simply copied and pasted the annual summary report into a CSV file.

In either case, it is usually necessary to explicitly transform the format of the "RecreationVisits" column into a number and to remove the commas that separate the numbers by thousands (a transformation that you can do with spreadsheet applications like Excel or Google Sheets,

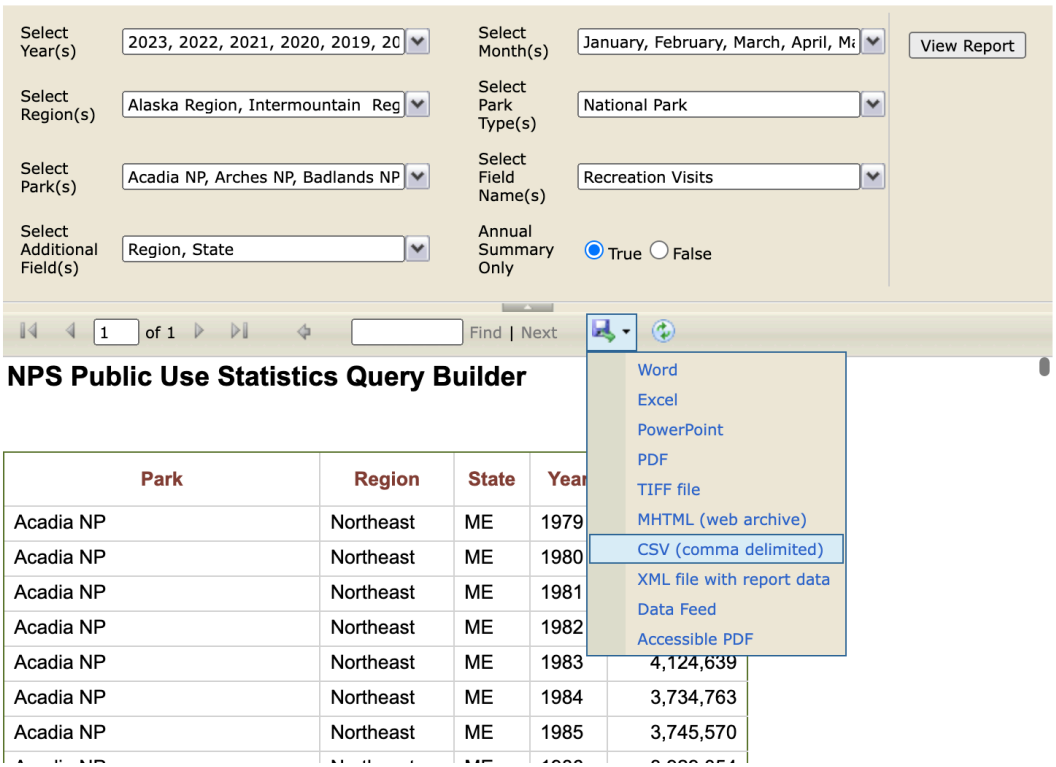


Figure 3: Selections for National Park visit data generated with "Query Builder for Public Use Statistics (1979 - Last Calendar Year)".

or with a programming language like Python or R). Finally, we published the data to this project's GitHub repository for easier storage and access.

Why was the data collected? How is the data used?

The NPS collects visit data partly because the government requires it, as we've already discussed. But the NPS also uses the visit data for other internal purposes—to help determine which parks might need more staff members and programming, which hiking trails might need more maintenance, which natural areas might need more protection, or which visitor centers might need more bathrooms.

The visit data also helps the communities and businesses surrounding the parks understand how they can best provide and share resources, like emergency vehicles, sanitation, and water. For example, if there's been a large influx of hikers to Mount Rainier National Park in recent years, that would be an important thing for the surrounding community to know. Because those hikers would probably need more ambulance trips and rescue helicopters (unfortunately but inevitably), and the surrounding towns wouldn't want visitors to the National Park booking up all the available emergency vehicles in town.

The visitation data also helps the NPS estimate the beneficial impact—economic and otherwise—that the parks have on nearby communities and the nation at large (Figure **??**). For example, in 2021, an NPS report showed that "4.5 million visitors to Grand Canyon National Park…spent an estimated $710 million in gateway regions near the park," which "supported 9,390 jobs in the local area." These estimations are important because they help the parks advocate for more funding, support, and attention.

The data is also frequently reported on by journalists, who use it to highlight the most popular parks and noteworthy visitation records, and to point their readers to parks where they might be able to find some peace and quiet (see articles in Thrillist, Smithsonian, and CNN).

> Discussion Question 1
>
> How else might the National Park visit data be used? How might it be used by artists, historians, literary scholars, sociologists, or librarians?
> For more, see Discussion Q 1.

What's in the data? What "counts" as a visit?

Now that we know how the data is used, let's dive into the data itself. What's actually in this dataset? What "counts" as a visit?

To get started, let's load the dataset and examine a random sample of rows.

Figure 4: 2021 report on NPS economic impact. Graphic credit: NPS.

```r
# https://statsandr.com/blog/an-efficient-way-to-install-and-load-r-packages/

# Load the dplyr package
library(dplyr, warn = FALSE)

# Load National Park Visitation data
np_data <- read.csv("https://raw.githubusercontent.com/melaniewalsh/responsible-datasets-in-

## Look at the structure of the dataset, randomly sample 10 rows
np_data %>% slice_sample(n = 10)
```

ParkName	Region	State	Year	RecreationVisits
Petrified Forest NP	Intermountain	AZ	2009	631613
Carlsbad Caverns NP	Intermountain	NM	1991	679450
Arches NP	Intermountain	UT	1992	799831
Carlsbad Caverns NP	Intermountain	NM	2016	466773
White Sands NP	Intermountain	NM	1987	567613
Black Canyon of the Gunnison NP	Intermountain	CO	1985	266012
Glacier Bay NP & PRES	Alaska	AK	2008	418911
North Cascades NP	Pacific West	WA	1990	456444
Big Bend NP	Intermountain	TX	2019	463832
Mammoth Cave NP	Southeast	KY	2001	1883580

Here we see five columns – "ParkName", "Region", "State", "Year", and "RecreationVisits." The first four are pretty self-explanatory, but why is the fifth labelled "RecreationVisits" rather than "Visits" or "Visitors"?

It turns out that the NPS counts visits, not visitors (which would be more difficult to track), and they distinguish between different *kinds* of visits to their parks. First, there are *reportable* and *non-reportable* visits. When NPS employees or their families go to the parks, these visits are *non-reportable*. But pretty much everything else is *reportable*. Within *reportable* visits, there are two more types of visits: *recreation* and *non-recreation* visits. Recreation visits are when people are visiting the parks for fun, vacation, exercise, school trips, etc., and non-recreation visits are when people are visiting the parks for other reasons. For example, some people need to travel *through* the parks, either because a highway runs through the park, or because they live on "inholdings" (private property that is surrounded by a National Park on all sides). Other people need to visit the parks because they have business to conduct.

Here's a full list of the "reportable non-recreation" visits, according to the NPS:

- Persons going to and from inholdings across significant parts of park land;

- Commuter and other traffic using NPS-administered roads or waterways through a park for their convenience;
- Trades-people with business in the park;
- Any civilian activity a part of or incidental to the pursuit of a gainful occupation (e.g., guides);
- Government personnel (other than NPS employees) with business in the park;
- Citizens using NPS buildings for civic or local government business, or attending public hearings;
- Outside research activities (visits and overnights) if independent of NPS legislated interests (e.g. meteorological research).

Carefully reviewing this list reveals that the term "recreation visit" excludes a significant number of visits and individuals. It also raises important questions about how the NPS distinguishes between different types of visits, which we will explore further below.

> Discussion Question 2
>
> What are the potential consequences of considering these visits to be *non-recreation* vs. *recreation* visits?
> For more, see Discussion Q 2.

The list also prompts us to consider those whose presence in the parks doesn't fit neatly into the "visit" category at all. For instance, a portion of Badlands National Park in South Dakota overlaps with the Pine Ridge Indian Reservation, which is "owned by the Oglala Sioux Tribe and managed by the National Park Service under an agreement with the Tribe." According to the NPS, when traveling through this area, visitors might encounter "signs of religious worship" from Tribal members, such as "prayer sticks" or pieces of "brightly colored fabric tied to a shrub," and they are advised to "respect [the Tribal members'] beliefs and practices and leave these objects." These symbols woven into the landscape underscore that members of the Oglala Sioux Tribe are not visitors to the Badlands but stewards and residents with deep ancestral connections. It reveals that the National Park data's focus on "visits"—whether reportable or non-reportable, recreational or non-recreational—fails to account for those who are not visitors, those who own and live on the land, and those whose ancestors lived on the land before the NPS even existed.

How was the data collected?

At this point, we know *what* counts as visit, but *how* does the NPS actually count these visits and collect data? And how do they differentiate between the different types of visits? Take a moment and see if you come up with a few guesses.

It turns out that each park counts visits differently. At many parks, *each entrance* at each park even counts visits differently.

Figure 5: Badlands National Park sign, gesturing to the South Unit's co-management between the Oglala Sioux Tribe and the NPS. Photo credit: NPS.