Major Literary Prize Winners

Juliana Spahr

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.

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).

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 many of these lands were taken, sometimes forcibly, from Native American people who already owned, lived, and worked on them [@spence_dispossessing_2000; @beauchamp_beyond_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_beyond_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_why_2013; @alba_covid-19s_2022; @floyd_coming_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.

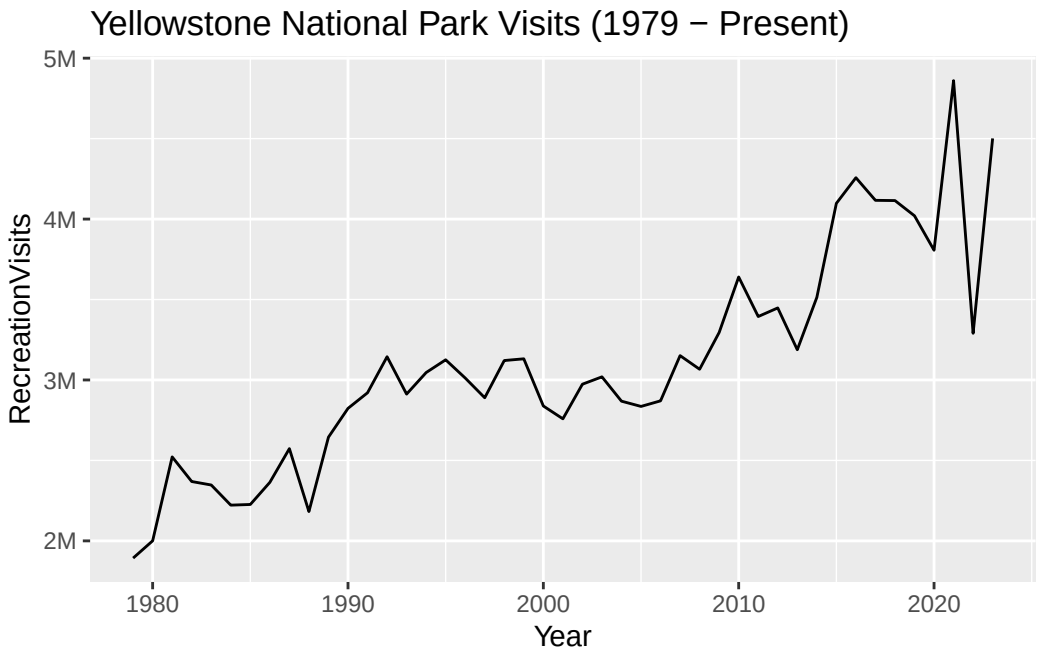
But in 1965, the NPS started getting serious about counting visits. That year, the U.S. Congress passed The Land and Water Conservation Fund Act of 1965. This act created a new source of government money specifically dedicated to protecting natural resources and expanding outdoor recreation infrastructure.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

# 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)") +
```

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

```
# abbreviate numbers by millions and thousands
scale_y_continuous(labels = label_number(scale_cut = cut_short_scale()))
```



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.