National Park Visitation Data

2024-02-26

Table of Contents

National Park Visitation Data

Data Essay

Introduction

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

```
//| echo: false

/*Inputs.table(search, data)*/

Inputs.table(search, {
  layout: "fixed",
  rows: 10,
  sort: "Year",
  reverse: true,
  format: {
    /*RecreationVisits: x => d3.format('.2s')(x),*/
    Year: x => d3.timeFormat(x),
    RecreationVisits: x => html`<div style='background:${color(x)}'>${d3.format('.2s')(x)}</
```

```
  }
})
```

This particular dataset is drawn from data published by the NPS. Most (but not all) of the contextual information included here draws from material published by the NPS, as well. However, the original data is made available in an NPS data portal that is relatively hard to find, and its documentation is scattered across many different web pages, which is why we believe it is valuable to curate and publish it in a single place here.

This dataset was curated and published by Melanie Walsh, and the data essay was written by Os Keyes and Melanie Walsh.

History

The National Park Service actually began recording information about park visits in 1904 (more than 100 years ago!). However, at this time, their visit collection methods were mostly informal, inconsistent, and low-tech. But over the next century, the NPS worked hard to make their data collection methods more reliable, consistent, and (in some but not all cases) high-tech.

A big catalyst for the NPS getting serious about data collection was a new law. In 1965, the U.S. Congress passed a federal law that was very important for the NPS and for anybody who loves the outdoors: The Land and Water Conservation Fund Act of 1965. This act created a new source of government money specifically dedicated to protecting natural resources (i.e. to buying up land and water so that condo developers couldn't do it first) and building up outdoor recreation infrastructure in the U.S.

One of the clauses in this act stipulated that the amount of money allocated to each recreation area should be "proportional to visitor use." Because the NPS can't function without money, they buckled down on counting visitor use. According to the NPS, over the next twenty years, they "developed and institutionalized a formal system for collecting, compiling and reporting visitor use data."

While today's visit data collection system is far more formal and sophisticated than the one that the NPS used in 1904, there are still many inconsistencies, flaws, and limitations in this system. These shortcomings are largely unavoidable. Trying to record every single visit to a National Park — across dozens of different parks and geographic regions, many decades of time, countless different weather conditions and funding situations, and hundreds of millions of people — is pretty much impossible. In fact, one of the reasons that this dataset is so useful and illuminating is because it does a good job of communicating an important point: data can *never* reflect reality precisely.
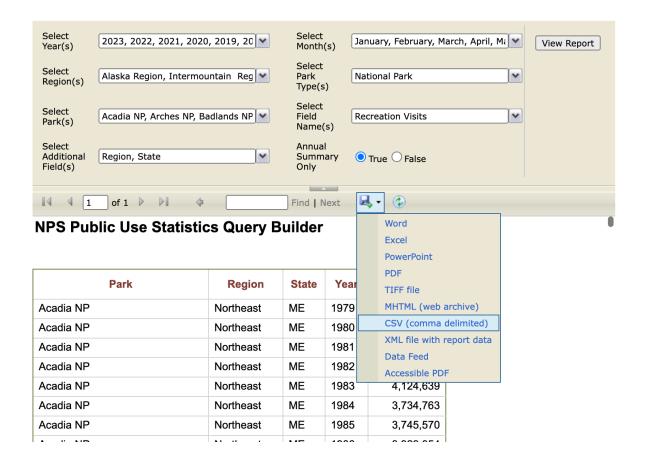
However, the NPS visitation data also does a good job of communicating why we might be interested in collecting and analyzing even flawed and approximate data, as we will dig into below.

Where did the data come from? Who collected it?

This National Park visitation data was originally organized and published by the NPS Social Science Program, a specific program tasked with coordinating visitor statistics across the parks. Thousands of staff members were also involved in the data collection process for individual parks, as we will elaborate below.

The original data was made available through the NPS Visitor Use Statistics data portal. Through this portal, you can generate reports and download data for many different park visitation categories and time periods— at both the national and individual park levels.

To download the data included here, we selected the "Query Builder for Public Use Statistics (1979 - Last Calendar Year)" report type. We then selected only National Parks; all possible years (1979-2022); all possible regions; only "Recreation Visits"; the additional fields of "State" and "Region"; as well as the option of an annual summary of visit counts (as opposed to monthly visit counts). We then downloaded this report as a CSV and published it to GitHub for easier access.

NPS Public Use Statistics Query Builder

Park	Region	State	Year	
Acadia NP	Northeast	ME	1979	
Acadia NP	Northeast	ME	1980	
Acadia NP	Northeast	ME	1981	
Acadia NP	Northeast	ME	1982	
Acadia NP	Northeast	ME	1983	4,124,639
Acadia NP	Northeast	ME	1984	3,734,763
Acadia NP	Northeast	ME	1985	3,745,570

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

As we've already discussed, one of the reasons that the NPS collects visit data is because the government basically requires it. But there are a lot of other reasons that the NPS collects this information.

As the NPS writes on their website, they use visit data to determine which facilities might need more or less attention, which parks might need more or less staff members and programs, and which hiking trails or bathrooms might need more or less maintenance. This information also helps the communities and businesses surrounding the parks understand how they can best share and support resources in a given area — services like emergency vehicles, sanitation, and water. If there are millions more people going on hikes in a particular area, and thus, inevitably, many more people requiring ambulance trips or rescue helicopters, that would be a very important thing for a community to know. It would be dangerous if visitors to National Parks suddenly and unexpectedly called all the emergency vehicles in town.

This 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. These estimations are

important because they help the parks advocate for more funding, support, attention, and collaboration.

The data can also be used for a variety of other purposes.... (such as?)

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

If we open the dataset and look at the first few rows, we will find five columns – "ParkName", "Region", "State", "Year", and "RecreationVisits":

```
# 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/Neat-Datasets/main/1979-2

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

ParkName	Region	State	Year	RecreationVisits
Grand Canyon NP	Intermountain	AZ	2001	4104809
Black Canyon of the Gunnison NP	Intermountain	CO	2000	191506
Virgin Islands NP	Southeast	VI	1984	666435
White Sands NP	Intermountain	NM	2002	498964
Congaree NP	Southeast	SC	2016	143843
Denali NP & PRES	Alaska	AK	1999	386867
Shenandoah NP	Northeast	VA	1982	1751972
Yellowstone NP	Intermountain	WY	2003	3019375
Channel Islands NP	Pacific West	CA	2014	342161
Great Basin NP	Pacific West	NV	1996	86635

The first four are self-explanatory: but why is the fifth labelled "RecreationVisits" rather than "Visits", or "Visitors"?

The answer is that what this dataset is tracking is more complicated and nuanced than "people who go to NPS properties". People go to the national parks for a lot of reasons. While many are there for recreation, some travel *through* the parks, either because a highway runs through or because they live on "inholdings" (private property that is surrounded by a national park on all sides). Because of this, the NPS defines "Recreation Visits" as visits made by people who are *not*:

> using park territory, roads, and facilities for their own convenience or as a part of their occupation. > Reportable non-recreation visits include:

- 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).

What this means is that the counts leave out a lot of people. This is worth thinking about when we evaluate what the numbers mean, and how the NPS achieves them (which we'll discuss more below)

Data and data collection

So now we know what is being collected. But let's try to understand *how* it's being collected. We can do this, in part, by exploring and visualising the data.

For example: let's visualise the visits to Crater Lakes National Park, from 1979 to the present:

```r
# Load the "ggplot2" package (which we'll be using a lot more)
library(ggplot2)

# Let's also load "ggthemes", which let's us use colorblind-compatible palettes. When we've
library(ggthemes)

# And specify the colorblind palette
cb_palette <- colorblind_pal()(8)

# Turn off scientific notation
options(scipen = 999)

# Filter down to Crater Lake National Park
crater_lake <- np_data %>% filter(ParkName == "Crater Lake NP")

# Visualize it
ggplot(data = crater_lake) +
  geom_line(aes(x =
  Year, y = RecreationVisits),
  color = cb_palette[1]) +
  labs(title = "Crater Lake National Park Visits (1979 - Present)")
```