

NYT

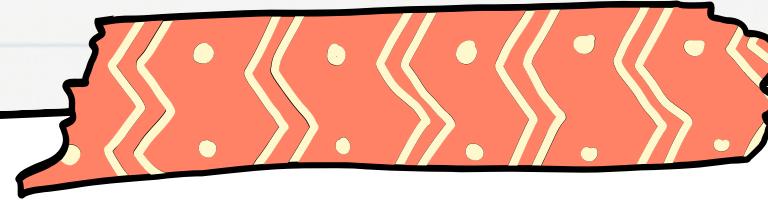
Bestsellers

Scott Macalister

Overview

- Project Overview
- Dataset
- Hypotheses
- Data Cleaning
- Challenges
- Results





Project Overview

Topic

- This project uses the New York Times' 'Books API', supplemented with Wikipedia information, to compare the bestselling works of fiction from 2013 and 2023.
- The goal is to gain insights into how the demographic makeup of America's most read authors has changed over the past decade.



Dataset



BOOKS | BEST SELLERS

Combined Print & E-Book Fiction

FICTION ▾ | NONFICTION ▾ | CHILDREN'S ▾ | MONTHLY LISTS ▾

When you purchase an independently ranked book through our site, we earn an affiliate commission.

1 2 WEEKS ON THE LIST
THE BOYFRIEND
by Freida McFadden | Poisoned Pen
A series of recent deaths causes Sydney Shaw to become suspicious of the handsome doctor she started dating.
[BUY ▾](#)

2 3 WEEKS ON THE LIST
COUNTING MIRACLES
by Nicholas Sparks | Random House
A man in search of the father he never knew encounters a single mom and rumors circulate of the nearby appearance of a white deer.
[BUY ▾](#)

3 WEEKS ON THE LIST
THE HOUSEMAID
by Elizabeth Strout | Random House
From behind closed doors, she sees everything.

October 27, 2024

When you purchase an independently ranked book through our site, we earn an affiliate commission.

BEST SELLERS METHODOLOGY
A version of this list appears in the October 27, 2024 issue of [The New York Times Book Review](#). Rankings on weekly lists reflect sales for the week ending October 12, 2024. Lists are published early online.

[More about our methodology](#)

BEST SELLERS SYMBOL KEY

↑ Up in Rank since last week

↓ Down in Rank since last week

If a book is not in a rank since the previous week, it will not have an arrow. The last few titles on the list known as the extended list, never have arrows.

* An asterisk indicates that a book's sales are barely distinguishable from those of the book above it.

PLANNER

ACCOUNTS

Dataset

The New York Times' Bestseller List

- Widely considered the most important list of best-selling books in America.
- Published the list weekly since October 12, 1931.
- It has evolved into multiple lists, including Fiction, Nonfiction, Audio, Children's Books, Graphic Novels.
- For this project, I chose to focus on Combined Print & E-Book Fiction.



Dataset

Stephen King

From Wikipedia, the free encyclopedia

(Redirected from Stephen king)

For other people with either the same or similar names, see [Stephen King \(disambiguation\)](#).

Not to be confused with [Steve King](#).

Stephen Edwin King (born September 21, 1947) is an American author. Widely known for his [horror](#) novels, he has been crowned the "King of Horror".^[2] He has also explored other genres, among them [suspense](#), [crime](#), [science-fiction](#), [fantasy](#) and [mystery](#).^[3] Though known primarily for his novels, he has written approximately 200 short stories, most of which have been published in collections.^[4]

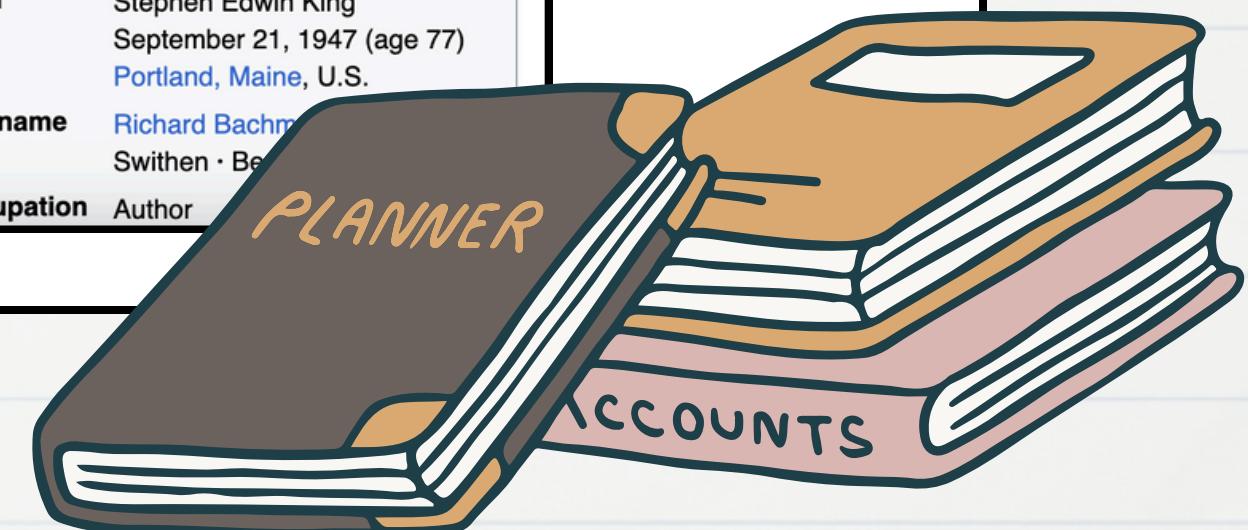
His debut, *Carrie* (1974), established him in horror. *Different Seasons* (1982), a collection of four novellas, was his first major departure from the genre. Among the successful films adapted from King's fiction are *Carrie* (1976), *The Shining* (1980), *The Dead Zone* (1983), *Christine* (1983), *Stand by Me* (1986), *Misery* (1990), *The Shawshank Redemption* (1994), *Dolores Claiborne* (1995), *The Green Mile* (1999), *The Mist* (2007) and *It* (2017), but there's also critical failures, such as *Children of the Corn* (1984), *Dreamcatcher* (2003) and *The Dark Tower* (2017). He has published under the pseudonym [Richard Bachman](#) and has co-written works with other authors, notably his friend [Peter Straub](#) and sons [Joe Hill](#) and [Owen King](#). He has also written nonfiction, notably *Dance Macabre* (1981) and *On Writing: A Memoir of the Craft*.

Stephen King in 2024

Born Stephen Edwin King
September 21, 1947 (age 77)
Portland, Maine, U.S.

Pen name Richard Bachman
Swithin · Be

Occupation Author



Dataset

Wikipedia

- Supplemented this data with Wikipedia data.
- I used regex to collect Age and Nationality data.
- My assumption was that, in the header text, the first four-digit number would show year of birth, and first country mentioned would show nationality.
- E.g. “Stephen Edwin King (born September 21, **1947**) is an **American** author.”



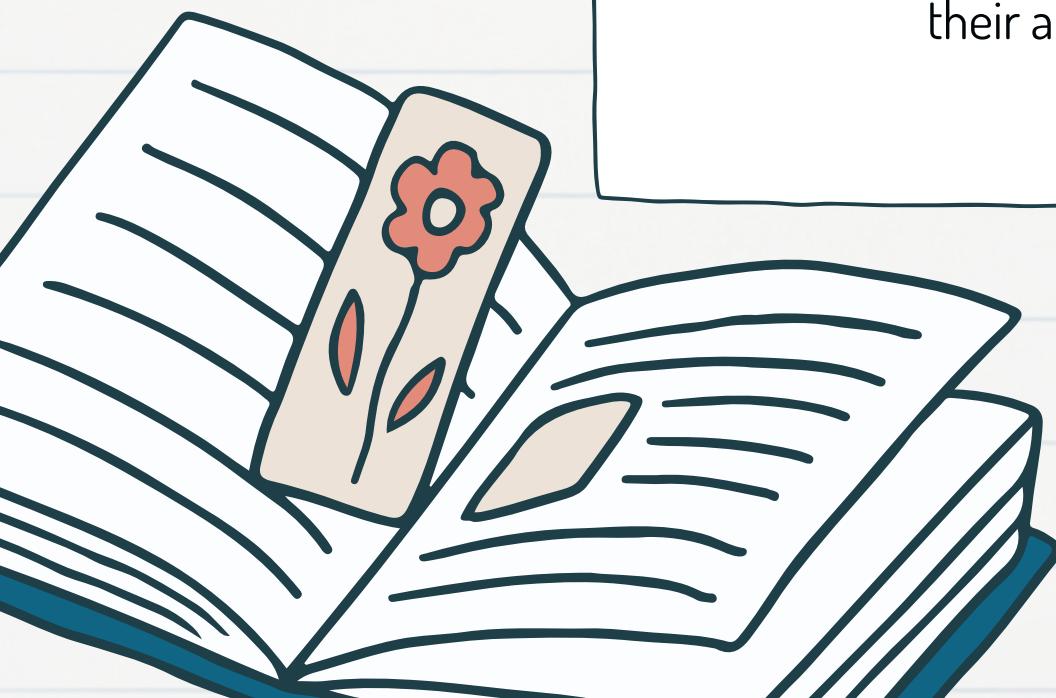
Hypotheses

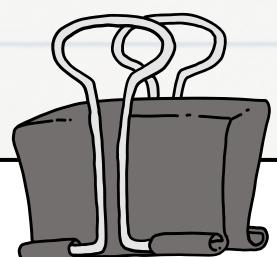
Age

I predict that the average age will be relatively high (≥ 55) for both cohorts, as older authors have had longer to hone their craft and build their audience.

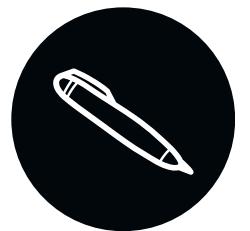
Diversity

I predict that the 2023 cohort will be more diverse (>20%) in terms of nationality, reflecting a greater interest in diversity within literary circles.



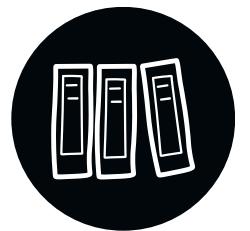


Data Cleaning



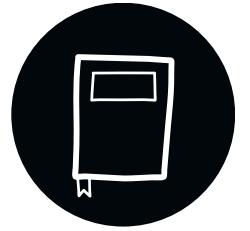
Author Names

- Some books were written by more than one author, e.g. 'James Patterson and Mark Pearson'.



Nationality

- Using regex, some authors were described by genre, instead of nationality, e.g. 'romance author' or 'thriller author',



Age

- My regex approach included obviously incorrect values, such as -20 or 700, which I excluded.



Challenges

Changing Project

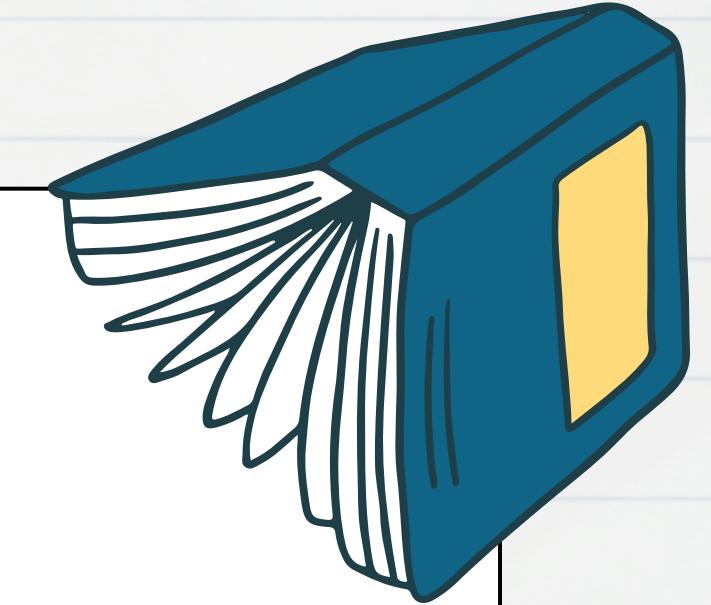
I switched my project very late, due to difficulties with the history APIs I was using. My planned project also relied to sending images to ChatGPT via API, which wasn't technically feasible.

Requests

Waiting for API requests is frustrating. Need to learn about parallel processing, selenium etc. Need to have some sort of printing function built in, so you have a rough idea of how long its taking. At one time, I accidentally turned off my wifi, losing my progress 😤😭

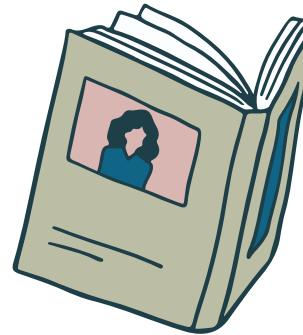
Backup Dataframe

Another painful lesson: have the dataframe your working with saved as a backup....



Results

01



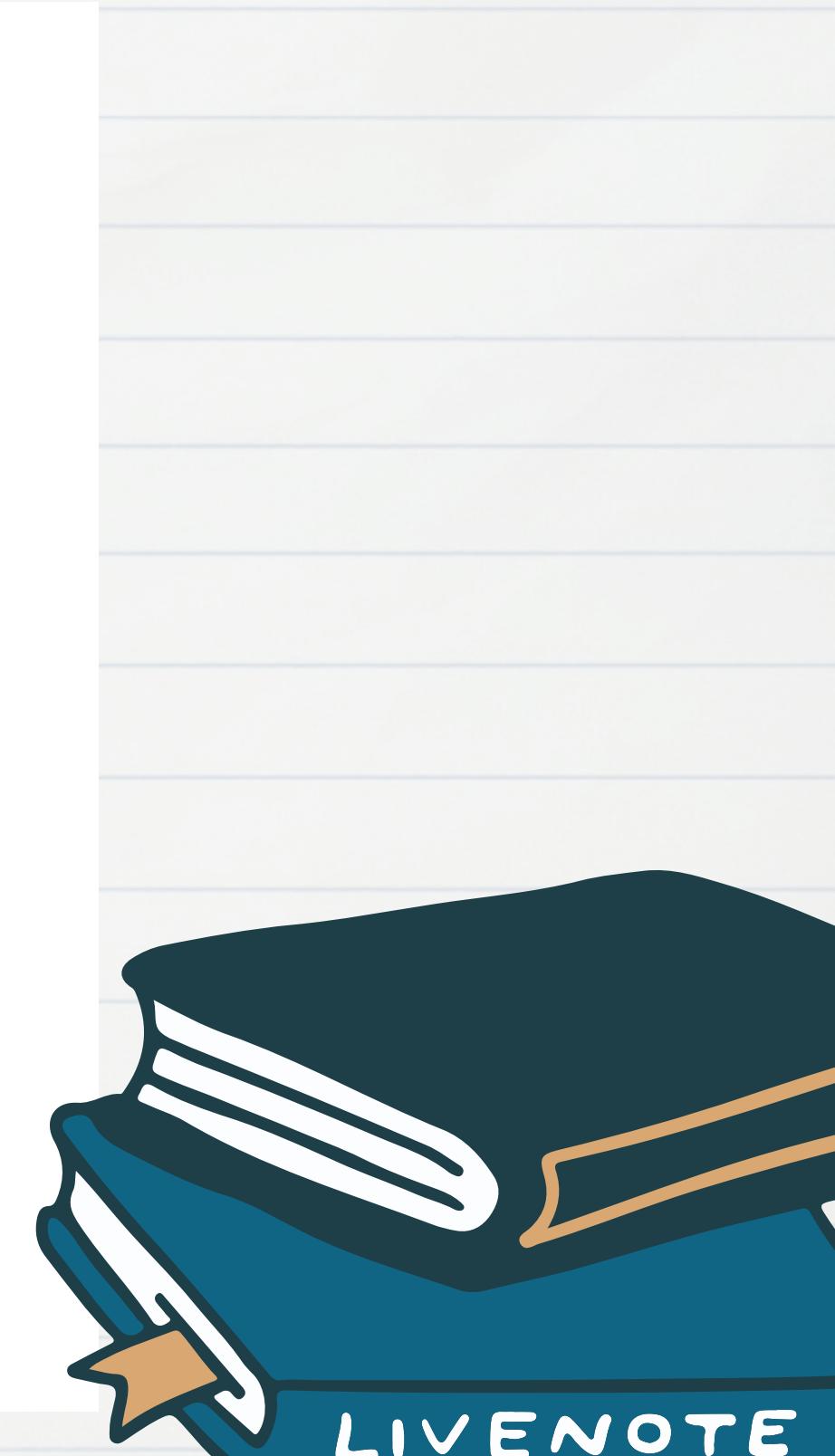
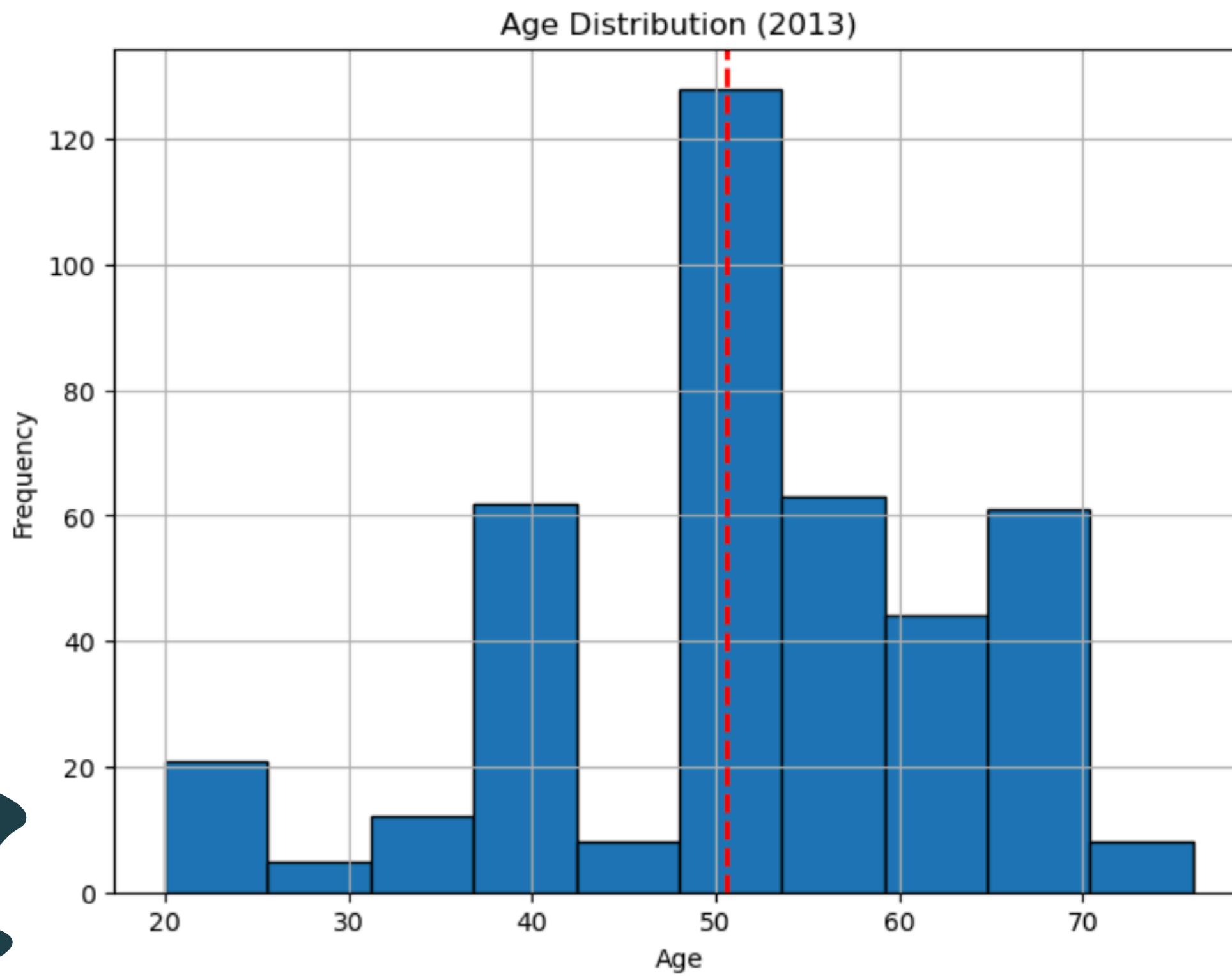
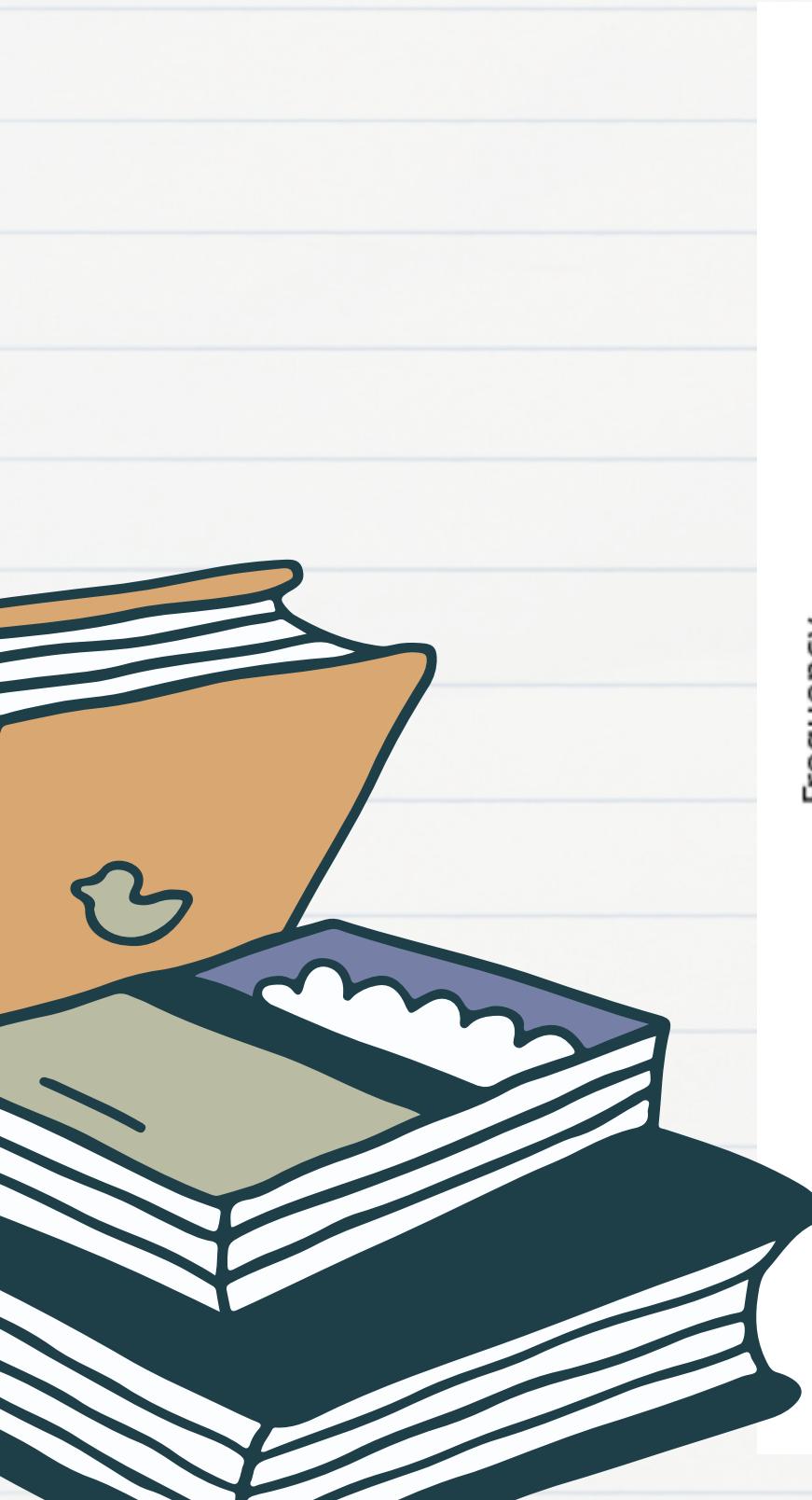
The mean age was 50.6 and 48.8 for 2013 and 2023 respectively. That was lower than I anticipated, but still higher than America general population. Perhaps the lower number in 2023 reflects trends like BookTok, grew in popularity in 2020 during COVID.

02

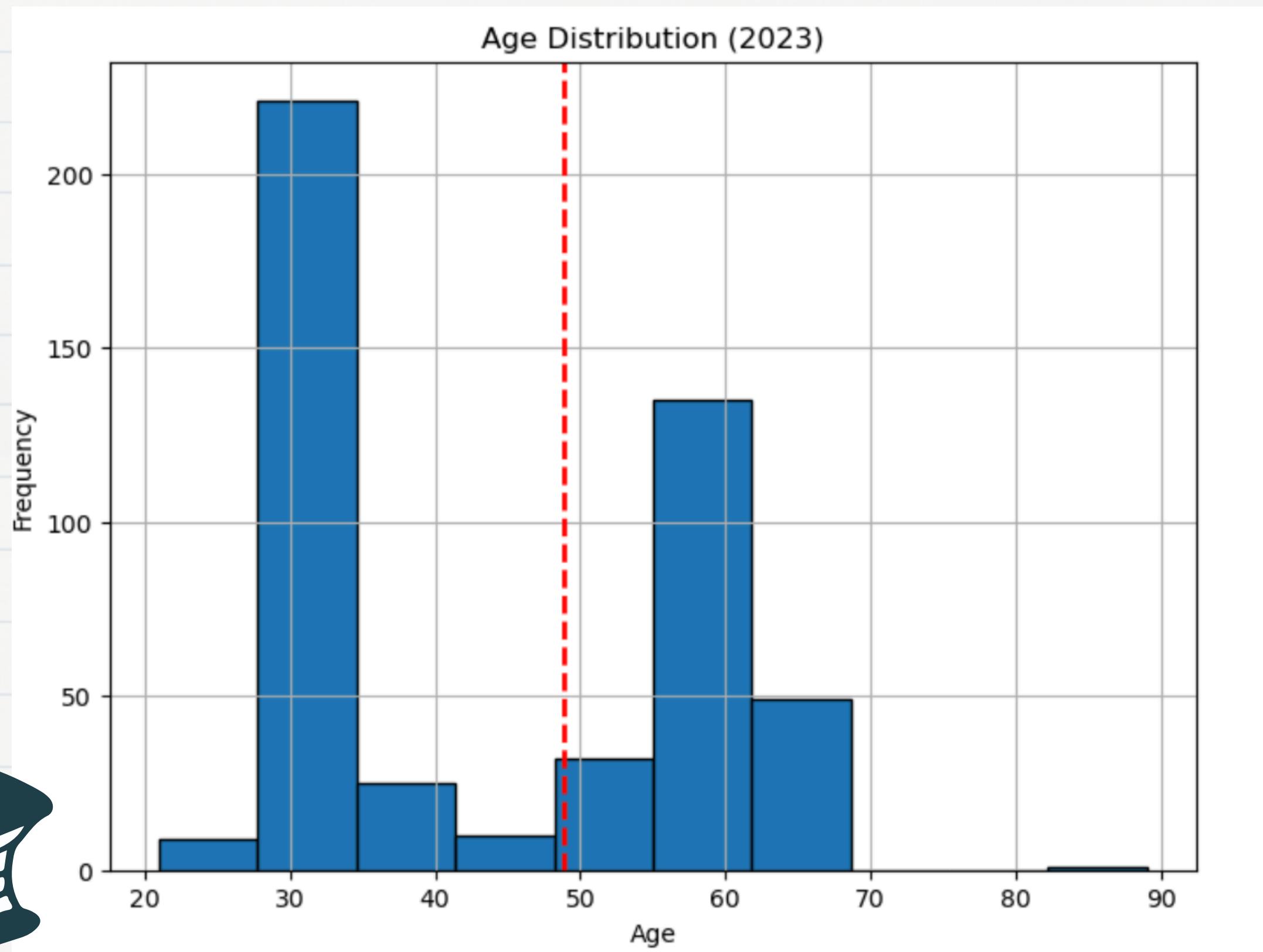


My second hypothesis proved completely wrong. Geographic diversity actually went down, from 80.9% to 93.5%. Perhaps other measures of diversity increased over this time, this would require further research.

Results

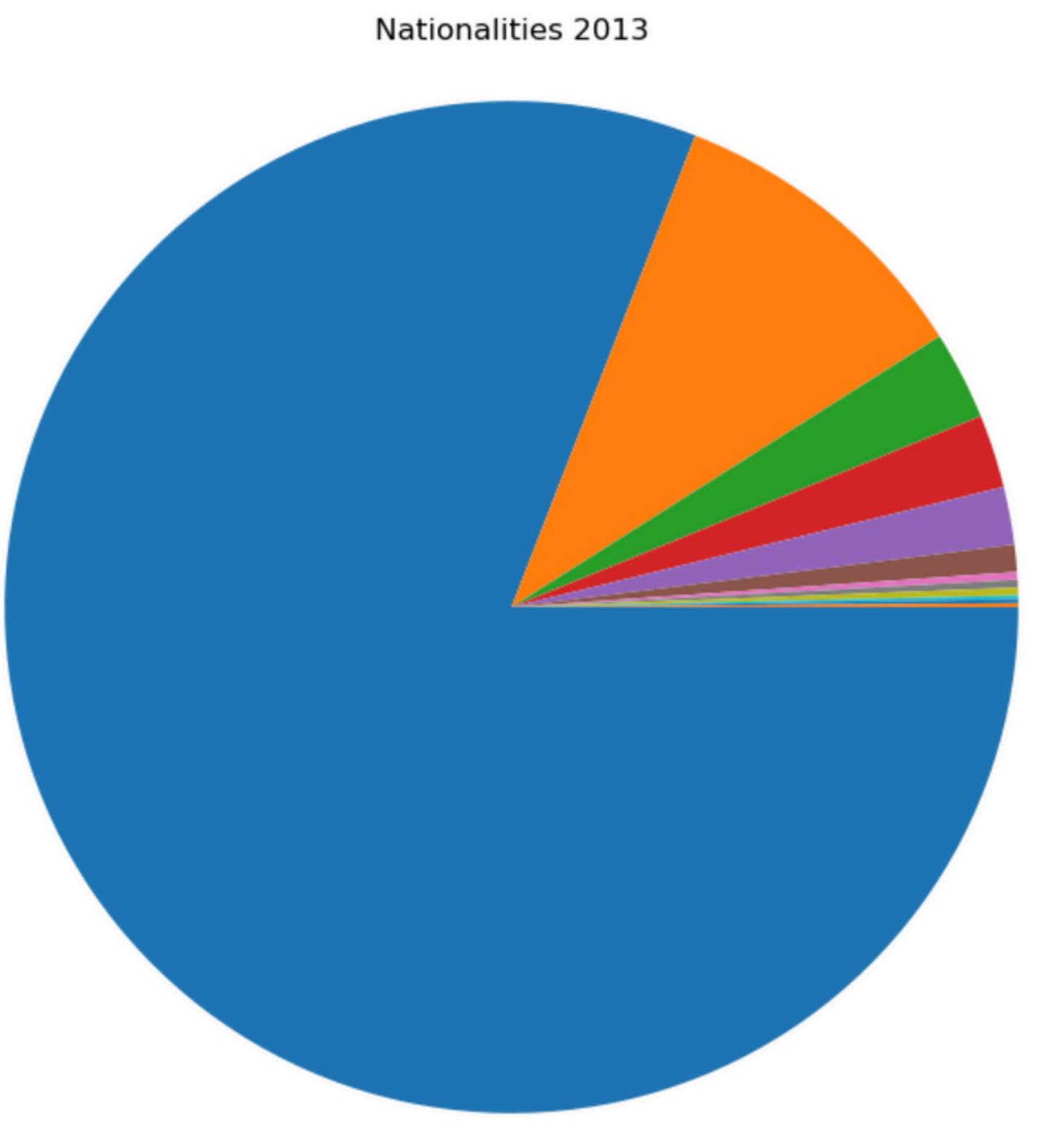


Results



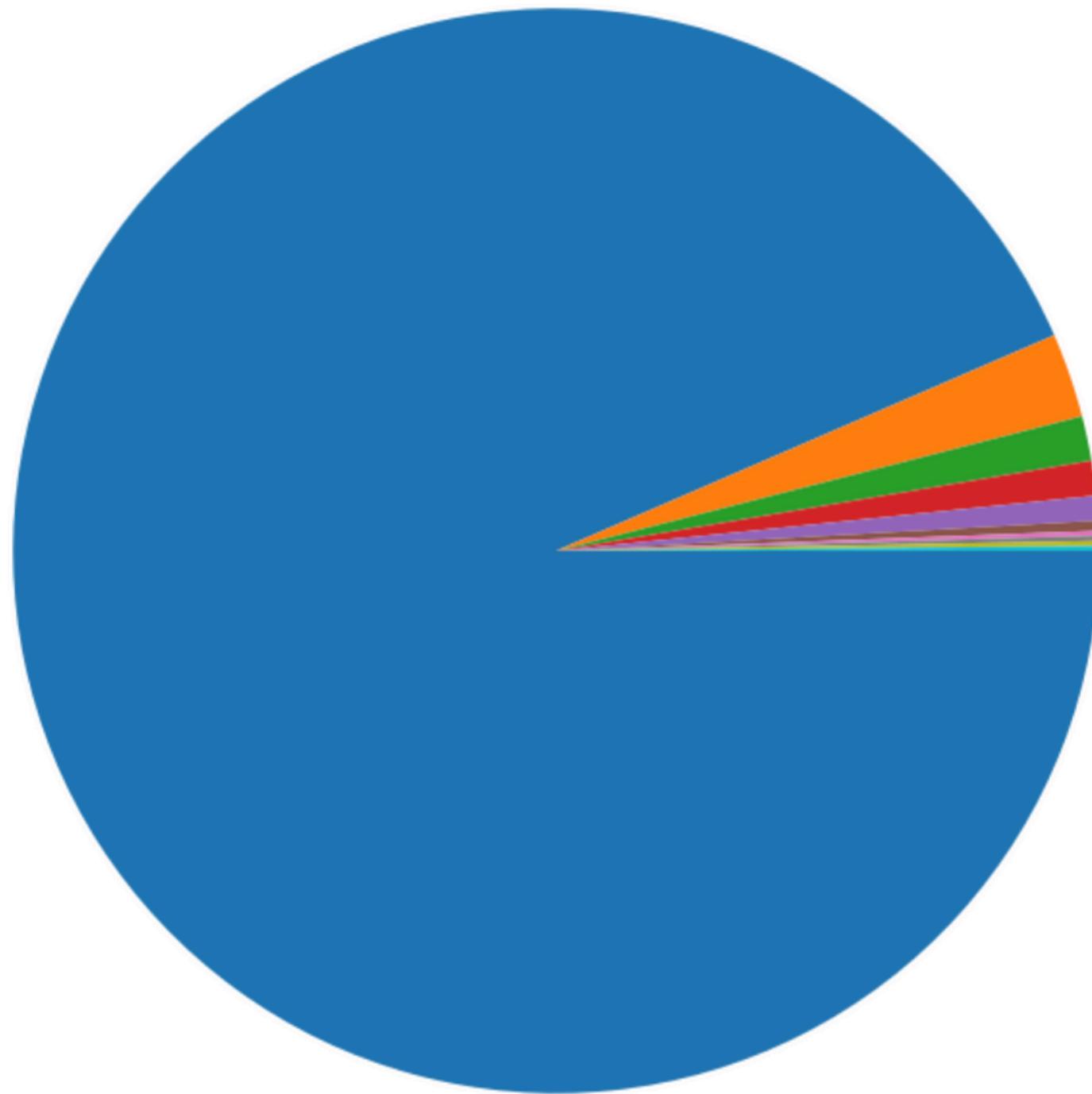
LIVENOTE

Results



Results

Nationalities 2023



LIVENOTE

Thank You

