

## 1. Data science books over time.

### Query:

```
SELECT year(cout) as years, count(*) as counts from spl_2016.inraw
where title like 'data science%' or title like '%data science%'
group by 1
limit 10
```

### Result:

The table below illustrates 2 things:

- People started writing about data science in 2014 (when the discipline, or rather a new name for it emerged).
- Readers' interest in data science books peaked in 2019 (as general interest in data science careers emerged).

years	counts
1970	1
2014	32
2015	64
2016	81
2017	78
2018	58
2019	164
2020	48
2021	77
2022	84

## 2. Math books over time

Now that we know when people started reading about data science, I want to see whether or not they paid equal attention to mathematical fundamentals.

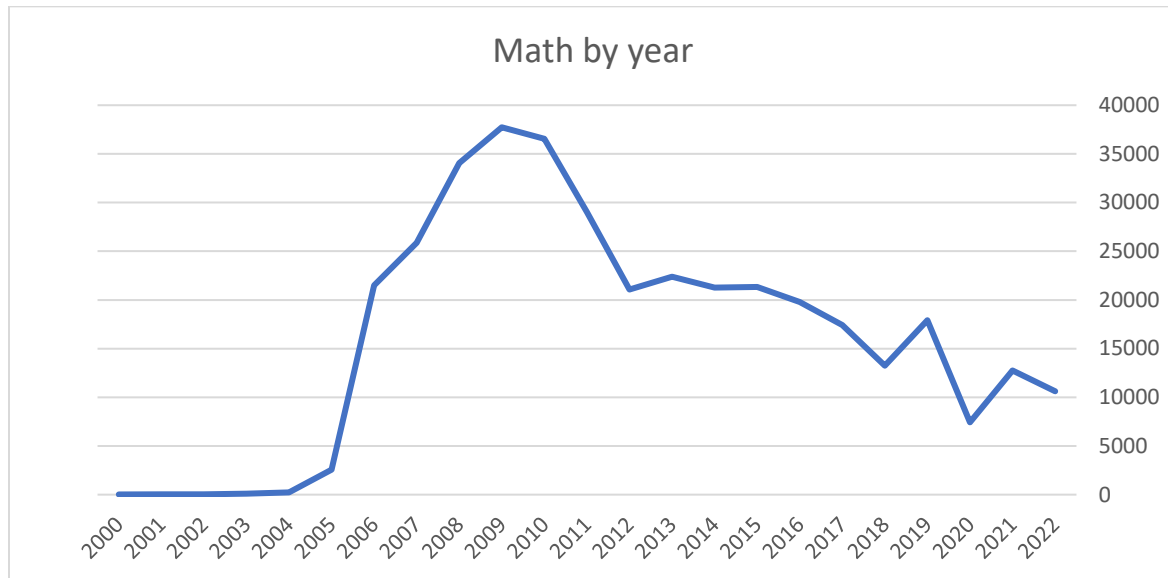
### Query:

```
SELECT year(cout) as years, count(*) as counts
from spl_2016.inraw
where deweyClass between 500 and 519 and year(cout) between 2000 and 2022
group by years
order by years desc
```

Results:

The table is too long to paste, so here is a quick Excel visualization:

- Interest in math books peaked around 2009 and decreased afterwards.
- Interest in math books increased around 2019, the same time when interest in data science books and in data science careers peaked.



### 3. Did people actually boycott J.K. Rowling after her Twitted scandal in June, 2020?

In June 2020, J.K. Rowling made a series of scandalous remarks on Twitter, which were considered transphobic by many. I want to see whether this situation influenced the number of times people checked out her books in the library.

Query:

```
select
```

```
    extract(year from cout) as years,
```

```
    extract(month from cout) as months,
```

```
    count(*) as Harry_Potter_counts
```

```
from spl_2016.inraw
```

```
where title like 'Harry Potter%' and cout > '2018%'
```

```
group by 1,2
```

```
order by 1,2
```

**Results:**

The table below illustrates the year 2020.

- April, June, and July have no data at all. Data entry issue?
- There is only 1 book checked out in May. Another data entry issue?
- Following J.K. Rowling's tweet, the number of people reading her books did seem to drop sharply in August and recovered to a certain extent in September.

years	months	Harry_Potter_counts
2020	1	581
2020	2	598
2020	3	316
2020	5	1
2020	8	10
2020	9	127
2020	10	194
2020	11	216
2020	12	249

**4. Political science books by year**

I want to know whether people's interest in political science coincides with the periods of increased political activity in the US (specifically presidential elections and primaries, to a lesser extent midterm elections).

**Query:**

```
SELECT
```

```
    year(cout) as years,
```

```
    count(*) as political_books
```

```
from spl_2016.inraw
```

```
where deweyClass between 320 and 328 and year(cout) between 2000 and 2022
```

```
group by 1
```

```
order by 1
```

**Results:**

- People definitely read more about politics during the presidential election year of 2008.
- No particular interest in 2012.

- A slight increase in 2016
- A significant drop in 2020.

