

Lec 13

# Data at a Scale

APIs, Scraping, Datasets, Personal Data, etc.

*Omar Hammad*

**Write down all of your Interactions  
with digital products in your 24 hours**



.....



+5B

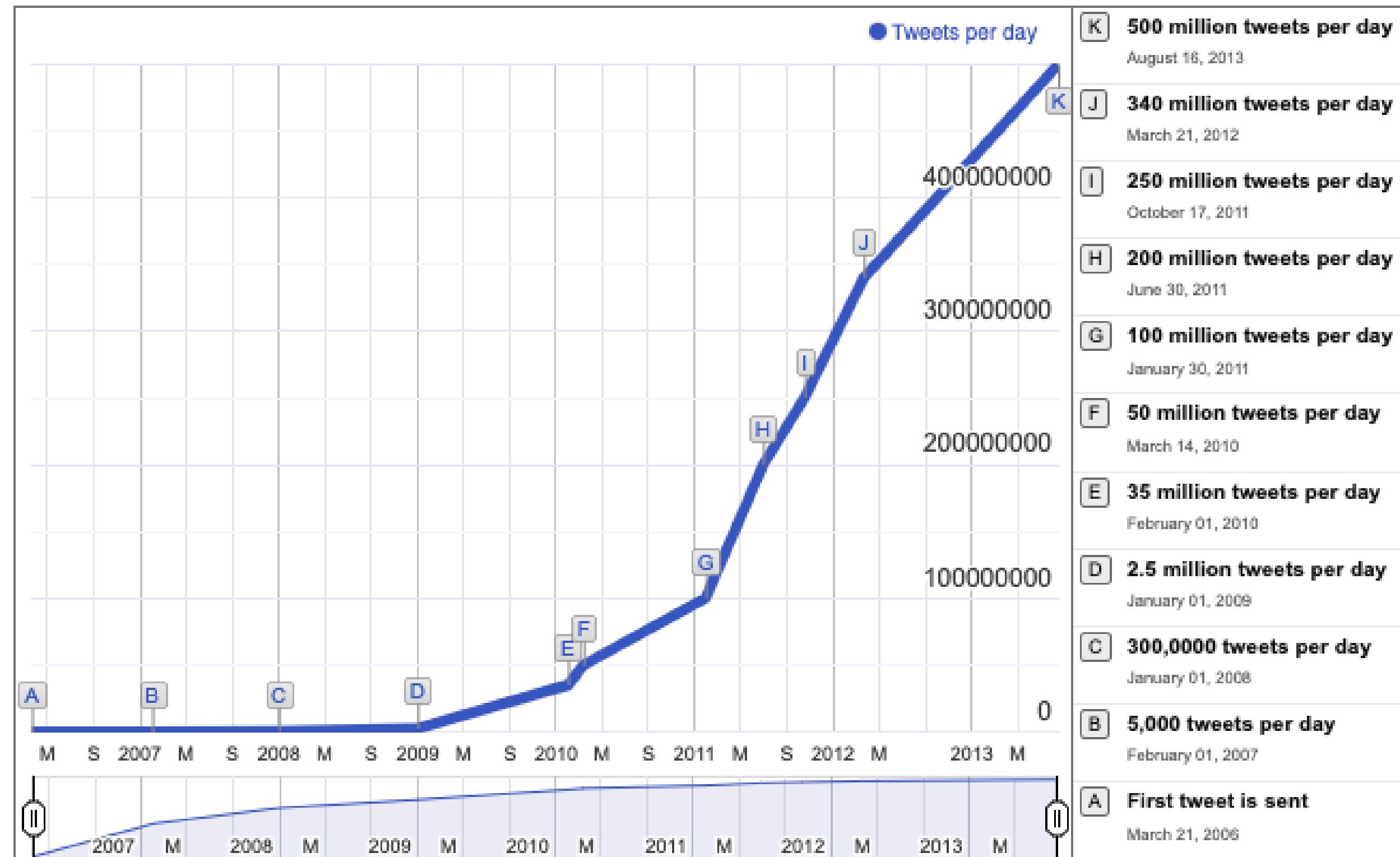
Active SM user

Data Reportal

+2h

Avg daily spend

# 500M tweets/day



.....

**Numbers**

.....

**Videos**

.....

**emotions**

.....

.....

**Images**

**Thoughts**

.....

.....

**Heartbeats**

**Location**

**apps usage**

.....

.....

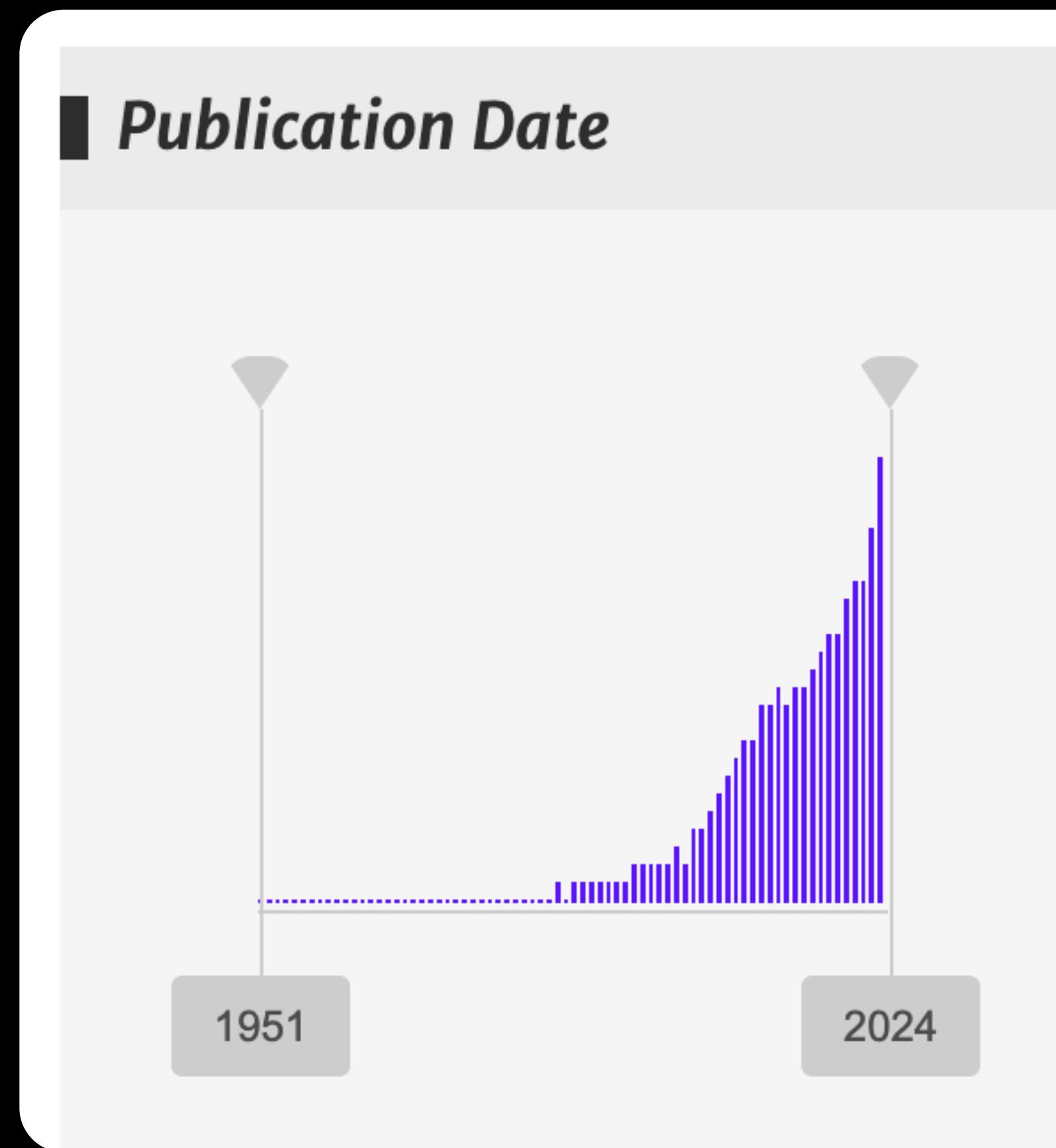
.....

.....

# In today's lecture

- How can we benefit from this huge amount of data?
- How did research shift in caring about user data?
- What sources can we use to collect big data?
- What tools can we use to collect big user data?
- How can we visualize big data?
- What ethical concerns should we care about?

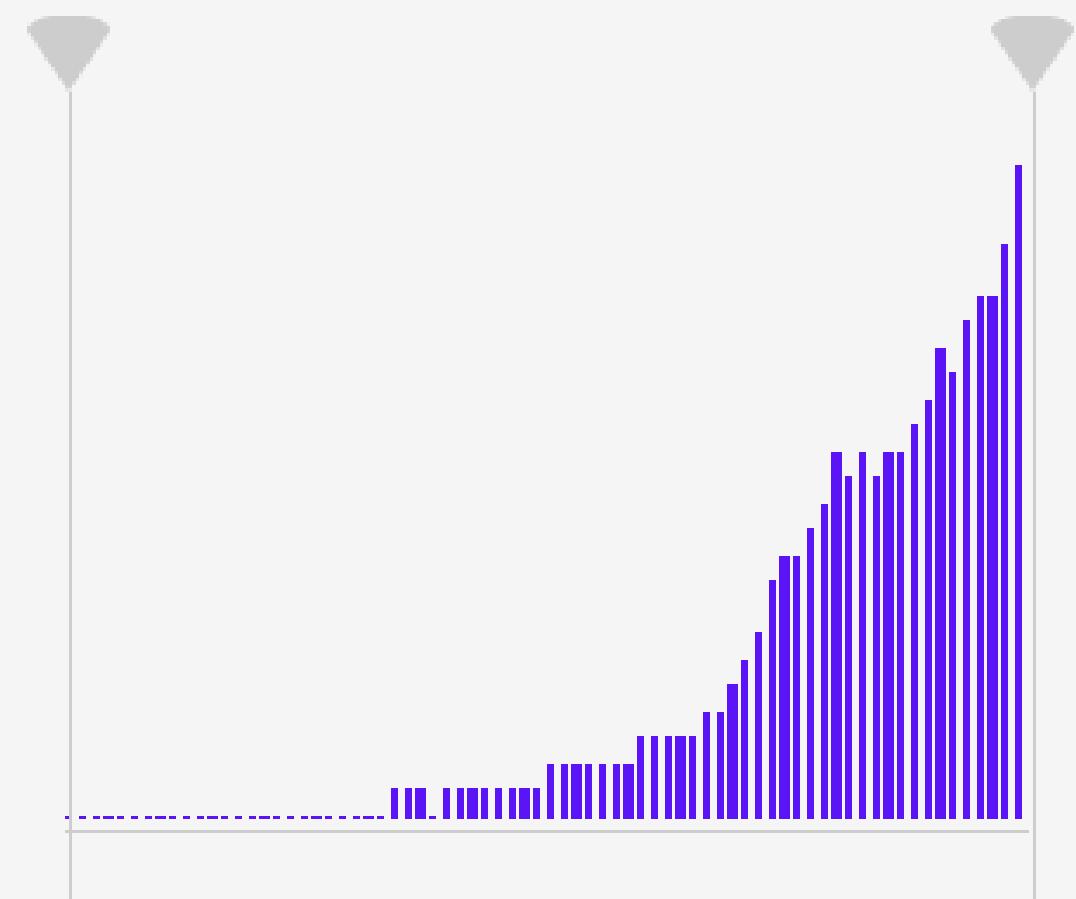
**How can we benefit from this opportunity?**



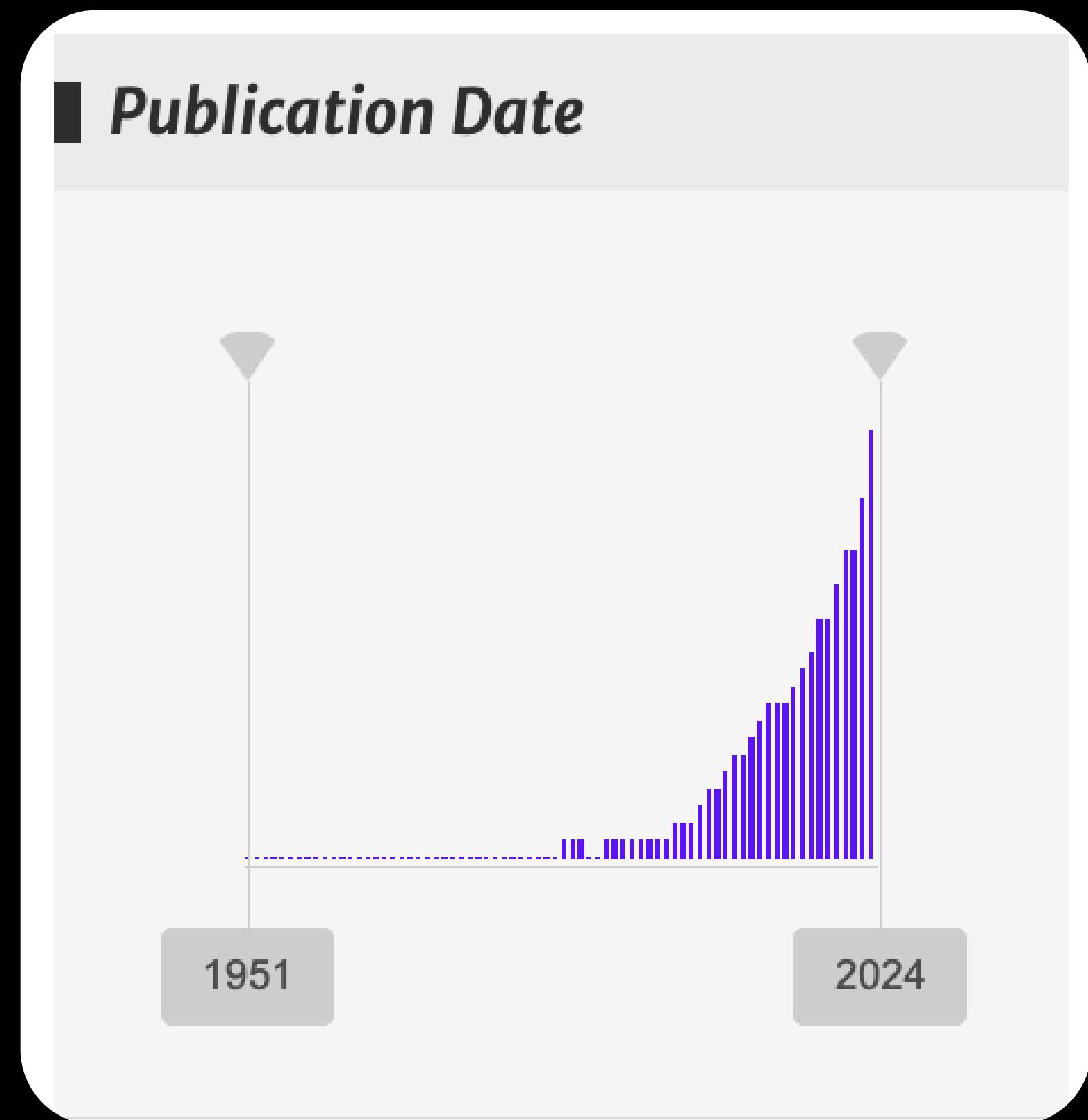
# “behaviour”

ACM Digital Library

### ■ *Publication Date*



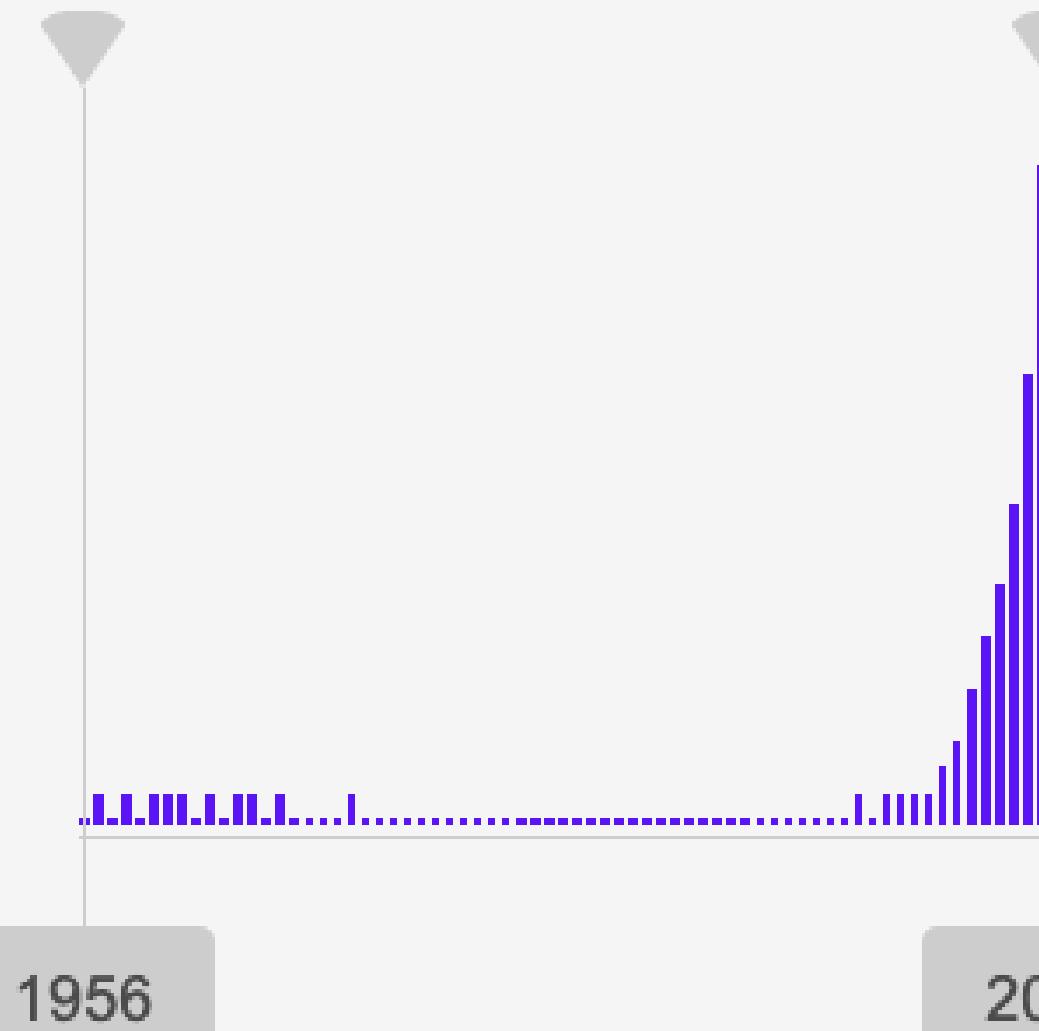
**“user model”**  
ACM Digital Library



# “recommender”

## ACM Digital Library

### **Publication Date**



# “misinformation”

ACM Digital Library

Common  
approaches  
to collect  
large amount  
of data

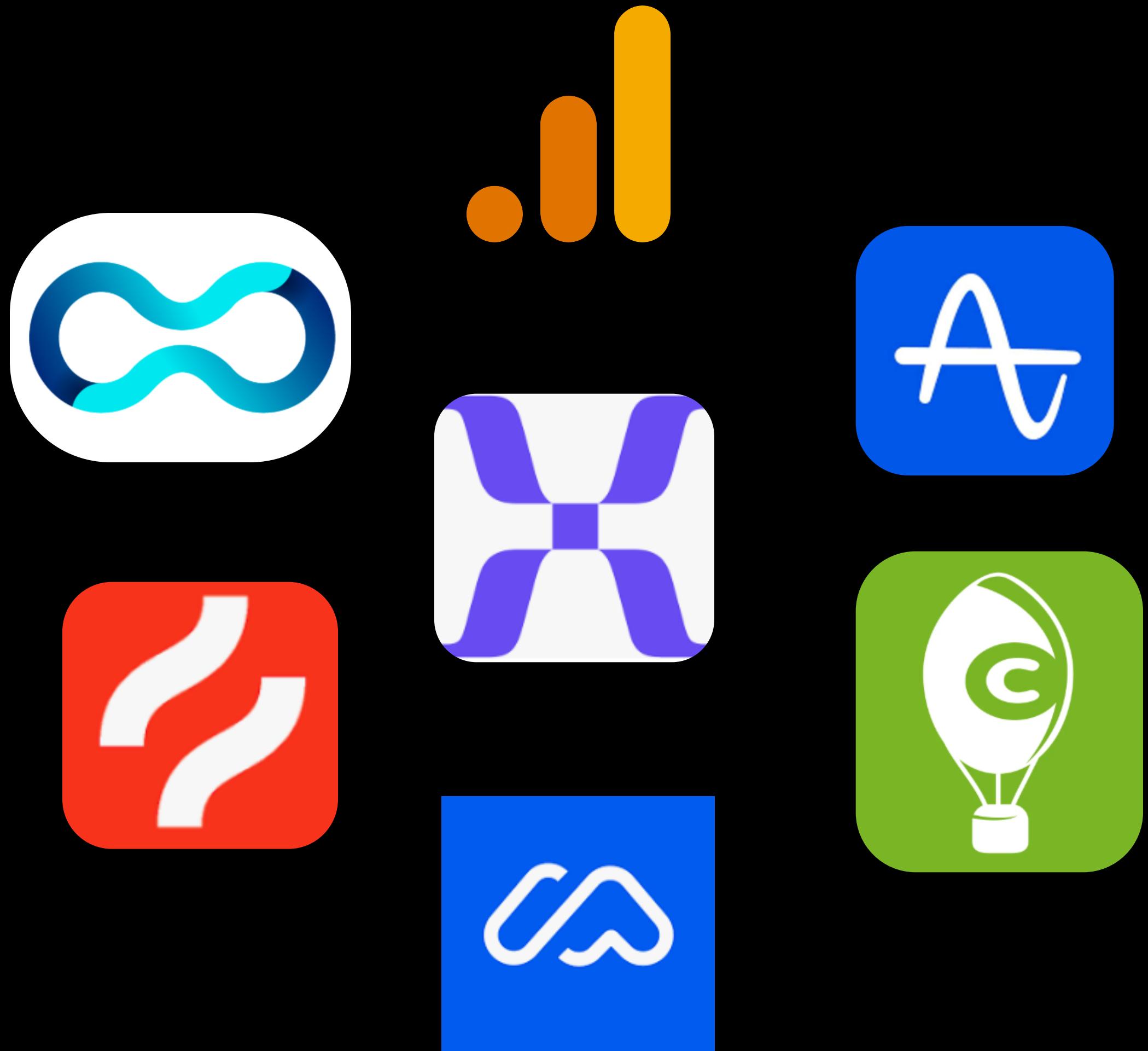
APIs   Sensors   Logs  
Scraping   Datasets  
Crowdsourcing

**Use of the  
following methods  
to get data for your  
project**

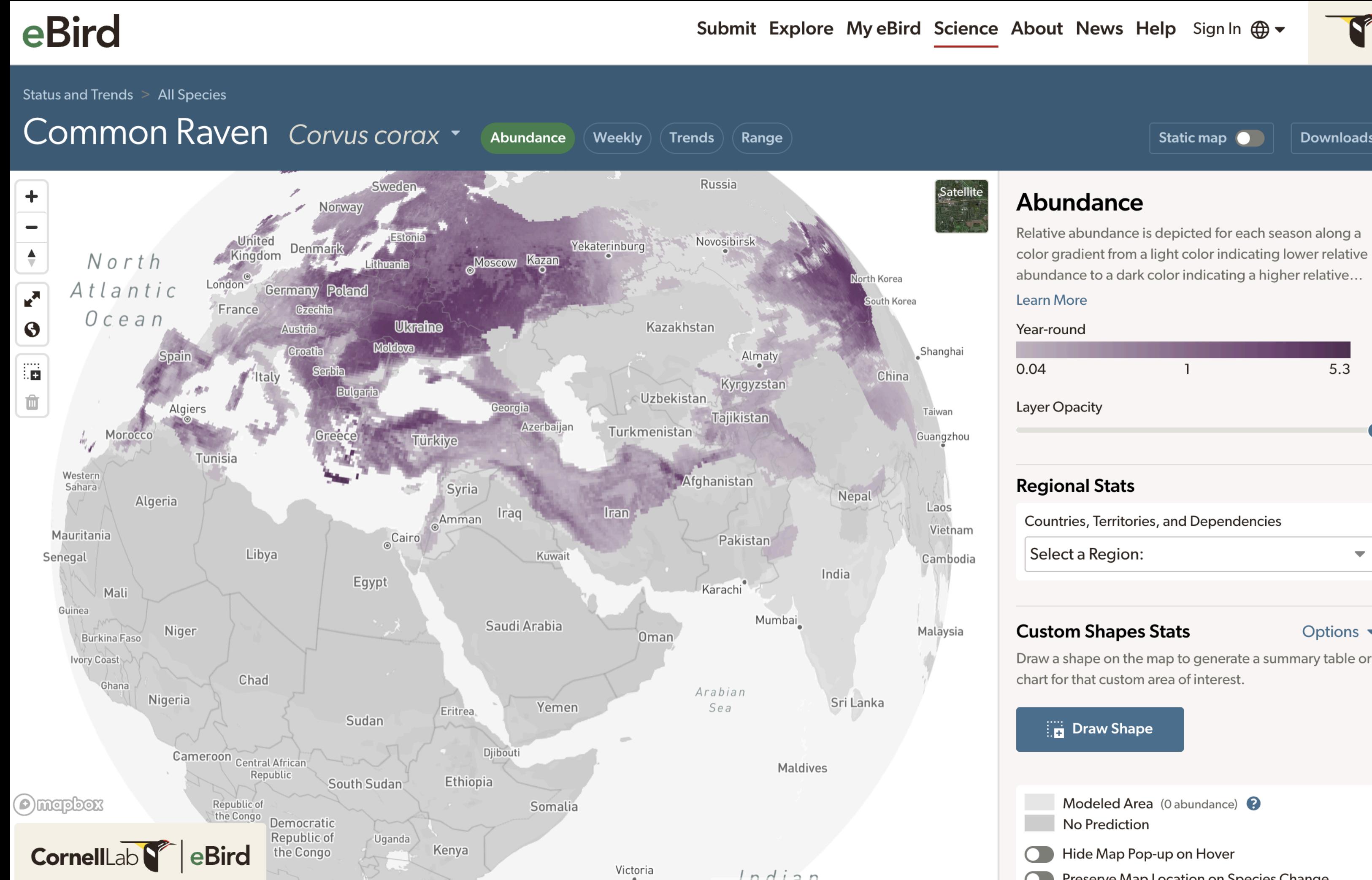
**APIs  
Scraping  
Datasets**

# User Data Tools

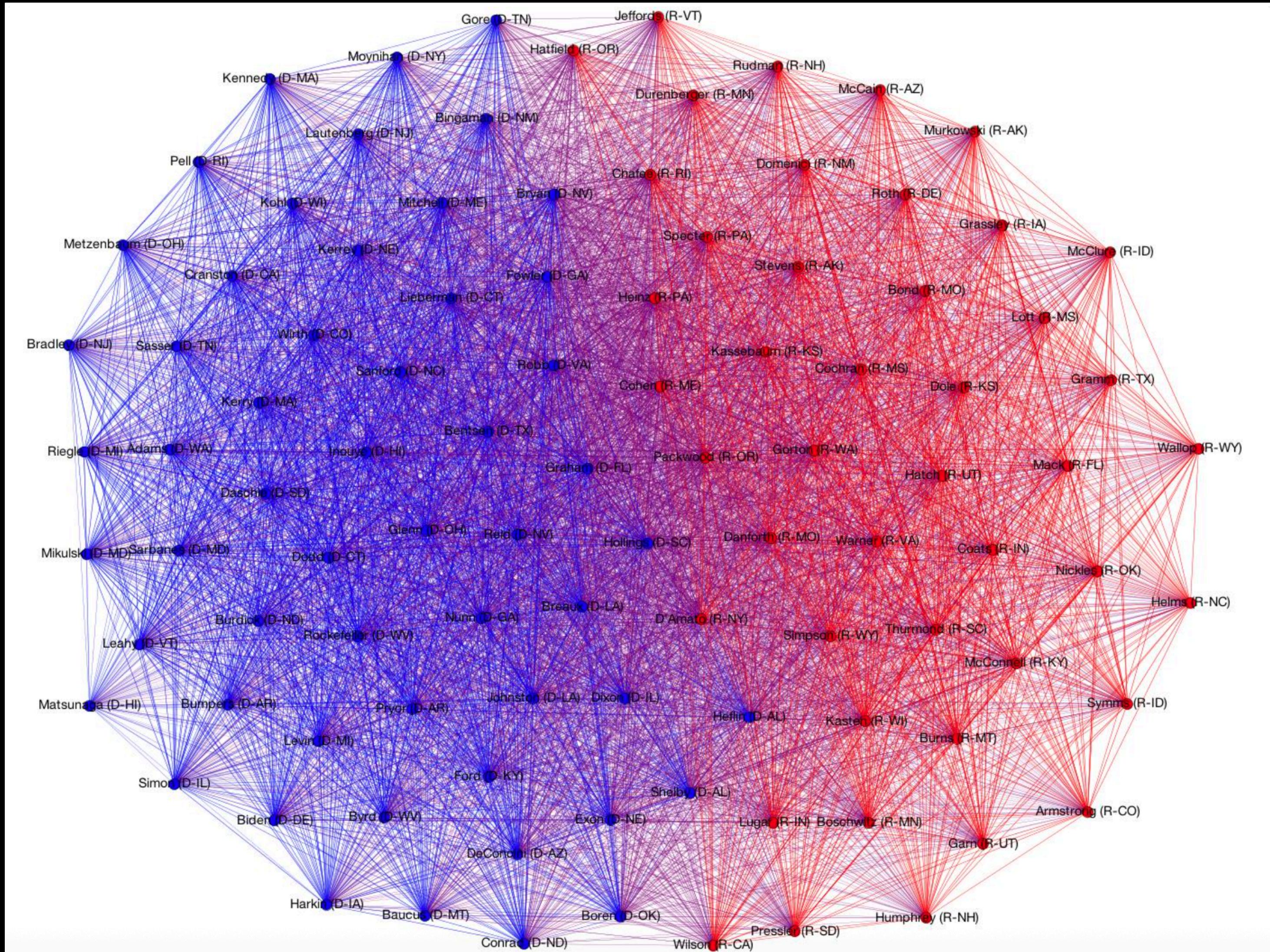
- Hotjar for heatmaps & recordings
- Mixpanel for analytics
- Smartlook combines both
- Crazy egg for A/B testing
- Amplitude for deep product analysis
- Google Analytics for marketing data



# **Big data visualization examples**



<https://science.ebird.org/en/status-and-trends/species/comrav/abundance-map>



<https://imgur.com/a/senate-voting-relationships-Wmoex>



<https://finviz.com/map.ashx?t=sec>

**Share some examples of big data  
visualization**

# Data ethics principles (FATE)

## **Fairness:**

Is the treatment just and without favoritism or discrimination?

## **Accountability:**

Is the data accurate and correct?

## **Transparency:**

Are the decisions being made by a system visible?

## **Explainability**

Can people understand the explanations provided by the system?