

# Missing datasets

What they are and how you can fill information voids  
[bit.ly/missing-data-gijc23](https://bit.ly/missing-data-gijc23)







# TROPIQUES



...  
L'Actualité  
de Cochinchine  
\*  
Le G. R. I. C.  
à Toulon  
\*  
Trois espèces  
sur le Laos.  
...  
...  
...  
...  
...  
...

Tropiques, April 1949

REVUE DES TROUPES COLONIALES

They described her as “an indomitable Frenchwoman, a true Joan of Arc of good colonization, always on horseback or by car, managing her plantations better than ten men would have done. Madame Marie-Madeleine O'Connell, adored by her thousands of coolies\* and their families and throughout the region, to the point that the natives had devoted a real cult to her and had consecrated her "Tutelary Genius" protector of their country.”

*From “A resistance against the Japanese, the Caodaists and the Vietminh”, Tropiques, April 1949*

\*From [Wikipedia](#): “A coolie (also spelled koelie, kuli, cooli, cooly, or quli) is an outdated, offensive, and racist term for a low-wage laborer, typically of Asian descent.”



Publicly available  
gun trace data

Cause of June 2015 black church  
fires

Number of mosques  
surveilled by FBI

Global web user measurements that  
include VPNs

Public list of citizens on

Quantifiable affect of  
corruption in loan economies

Employment statistics that include

Accurate birth registration

Demographics of all bitcoin

Financial information for all

Number of jobs actually lost

Number of civilians killed  
during Nigerian Civil War

Accurate number of car crashes  
that occur in the U.S.

Full beliefs on nationwide

# Missing data sets (Mimu Onuoha)

*"Missing data sets" are the blank spots that exist in spaces that are otherwise data-saturated. [...] That which we ignore reveals more than what we give our attention to. It's in these things that we find hints of what is deemed important. Spots that we've left blank reveal our hidden social biases and indifferences.*



Yourself



DIY data sets

Instead of relying on others, make your own datasets

Do

IT

## Three methods of making your own data sets

- Surveys
- Scraping
- Quantified selfie

Do

IT

yourself



# Surveys



# Hello!

The Marshall Project and Slate are conducting a survey to understand the political opinions of prisoners in the United States. Your participation is voluntary. The results will be shared in an upcoming edition of News Inside.

At the end of the survey, you have the option of

**1. How often do you discuss politics with family, friends or others who are incarcerated?**

- Very frequently
- Frequently
- Occasionally
- Rarely
- Never

**2. How do you primarily get your news?**

- Newspapers
- Local radio
- National Public Radio
- Local TV news
- Cable news
- Word of mouth
- Do not read/watch/listen to news

**7. Which party do you committed to criminal**

- Democrats
- Republicans
- They're equally c
- Neither is commi
- Don't know / No

**8. Please indicate ho about these policies**

- a. Banning assault we
- Strongly support
  - Support
  - Undecided
  - Oppose
  - Strongly oppose

b. Legalizing marijuana

# Slate/Marshall Project: Prisoner Survey

Slate and Marshall Project reporters and editors brainstormed a list of questions. We also consulted with imprisoned sources. [...]

We hoped for 1,000 responses. We were surprised to receive more than 8,000.

Source: [Slate/The Marshall Project](#)

1. How often do you discuss politics with family, friends or others who

Very frequently

Frequently

Occasionally

Rarely

Never

2. How do you primarily get your news?

Newspapers

Local radio

National Public Radio

Local TV news

Cable news

Word of mouth

Do not read/watch/listen to news

7. Which party do you committed to criminal

Democrats

Republicans

They're equally c

Neither is commi

Don't know / No

8. Please indicate ho about these policies

a. Banning assault we

Strongly support

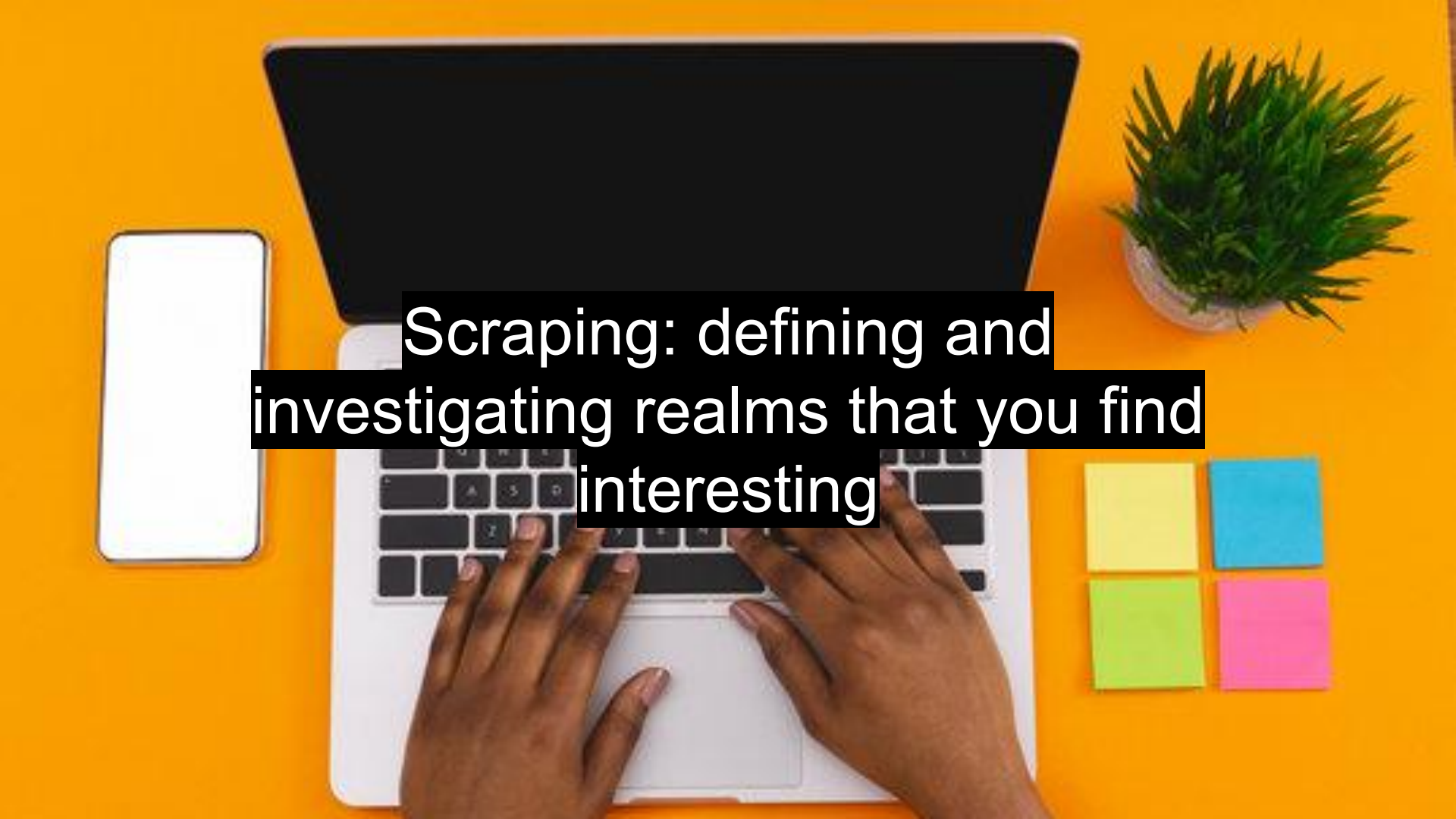
Support

Undecided

Oppose

Strongly oppose

b. Legalizing marijuan

A top-down view of a workspace on a bright yellow background. In the center is a silver laptop with a black keyboard and a black screen. A person's hands are visible, typing on the keyboard. To the left of the laptop is a white smartphone. To the right is a small green succulent in a white pot. Below the succulent are four sticky notes: two yellow, one light blue, and one pink. The text "Scraping: defining and investigating realms that you find interesting" is overlaid in white on a black background across the laptop keyboard.

Scraping: defining and  
investigating realms that you find  
interesting

# BuzzFeed News: Social Media Surveillance of Teens

BuzzFeed News submitted public records requests to more than 40 school districts, asking for the alerts they received from Social Sentinel.

Many of them denied our requests, but we received more than 1,800 alerts sent to administrators from eight school districts between May 2017 and September 2019 for which we were able to read the text or view the images in posts that were flagged.

Source: [BuzzFeed News](#)



The quantified selfie

**The Markup:  
Misinformation in the  
Vietnamese  
Community**



# Archive for [REDACTED]@gmail.com

Mar 21, 2023, 7:07:39 PM PDT • 23.9 MB • [Learn more](#) about Google archives and your [deleted data](#)

## PRODUCTS IN ARCHIVE (1)



YouTube and YouTube Music  
9 files, 23.9 MB

YouTube

e



## YouTube and YouTube Music

9 files exported successfully

[About this export](#)

[Exported files](#)

[File formats](#)

Some file formats (e.g. html, png) will be viewable in your browser, and will appear as clickable links below that will open the file you've fully extracted your archive. Other file types can be opened by their appropriate applications by navigating through the archive directory. You can click on any folders that may be present to see their contents.

Data that are still in the process of being deleted are not included in your archive. [Learn more](#) about your deleted data.

 history

[search-history.html](#)

[watch-history.html](#)

 music-library-songs

 my-comments

 my-live-chat-messages

 playlists

 subscriptions

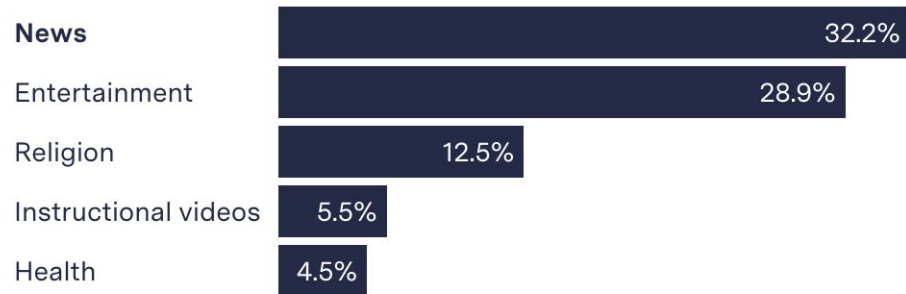
# The Markup

From Hoang's watch and search history on YouTube, I can tell that Hoang uses YouTube to help her make better decisions about her life: from spiritual guidance to instructional cooking videos to getting medical advice.

Source: [The Markup](#)

## News is Hoang's most watched category on YouTube, over three years

Percent of YouTube videos, by category



Notes: The data is a randomized sample of 1,000 videos viewed by Hoang; each was reviewed and classified by The Markup. The percentages do not add up to 100 percent because 16.4 percent of all the videos reviewed in the sample had been removed from the platform since Hoang watched them.

Chart: Lam Thuy Vo • Source: YouTube Archive provided to The Markup



Creative problem  
solving in service of  
underrepresented  
communities, help  
us tell stories like  
hers

Stay in touch!

Twitter: @lamthuyvo

[lam@themarkup.org](mailto:lam@themarkup.org)

<https://bit.ly/missing-data-giic>