Problem Assignment: 04

Khulan Bud

2026-02-07

RecustomVerbatimEnvironment{verbatim}{Verbatim}{ showspaces = false, showtabs = false, breaksymbolleft={}, breaklines }

Due 02/07 at 5:00PM Central.

"This submission is my work alone and complies with the 30538 integrity policy." Add your initials to indicate your agreement: K.B.

Github Classroom Assignment Setup and Submission Instructions

1. **Accepting and Setting up the PS4 Assignment Repository**
   - Each student must individually accept the repository for the problem set from Github Classroom ("ps4") – https://classroom.github.com/a/hWhtcHqH
     ‣ You will be prompted to select your cnetid from the list in order to link your Github account to your cnetid.
     ‣ If you can't find your cnetid in the link above, click "continue to next step" and accept the assignment, then add your name, cnetid, and Github account to this Google Sheet and we will manually link it: https://rb.gy/9u7fb6
   - If you authenticated and linked your Github account to your device, you should be able to clone your PS4 assignment repository locally.
   - Contents of PS4 assignment repository:
     ‣ `ps4_template.qmd`: this is the Quarto file with the template for the problem set. You will write your answers to the problem set here.
2. **Submission Process**:
   - Knit your completed solution `ps4.qmd` as a pdf `ps4.pdf`.
     ‣ Your submission does not need runnable code. Instead, you will tell us either what code you ran or what output you got.
   - To submit, push `ps4.qmd` and `ps4.pdf` to your PS4 assignment repository. Confirm on Github.com that your work was successfully pushed.

Grading

- You will be graded on what was last pushed to your PS4 assignment repository before the assignment deadline
- Problem sets will be graded for completion as: {missing (0%); ✓- (incomplete, 50%); ✓+ (excellent, 100%)}

‣ The percent values assigned to each problem denote how long we estimate the problem will take as a share of total time spent on the problem set, not the points they are associated with.

• In order for your submission to be considered complete, you need to push both your `ps4.qmd` and `ps4.pdf` to your repository. Submissions that do not include both files will automatically receive 50% credit.

Step 1: Develop initial scraper and crawler

	Title	Date	Link
0	Houston Transplant Doctor Indicted For Making …	February 5, 2026	https://oig.hhs.gov/fraud/ enforcement/houston-…
1	MultiCare Health System to Pay Millions to Set…	February 4, 2026	https://oig.hhs.gov/fraud/ enforcement/multicar…
2	Brooklyn Banker Pleads Guilty to Laundering Pr…	February 3, 2026	https://oig.hhs.gov/fraud/ enforcement/brooklyn…
3	Delafield Man Sentenced to 18 Months' Imprison…	February 3, 2026	https://oig.hhs.gov/fraud/ enforcement/delafiel…
4	Former NFL Player Convicted for $197M Medicare…	February 3, 2026	https://oig.hhs.gov/fraud/ enforcement/former-n…

Step 2: Making the scraper dynamic

1. Turning the scraper into a function

• a. Pseudo-Code

1. Define a function that takes start_year and start_month as inputs.
2. If start_year < 2013, print a warning message and exit the function.
3. Initialize the base URL, page counter, and an empty list to store results.
4. While enforcement actions are within the specified date range, scrape the current page.
5. For each enforcement card, extract the title, date, and link and append valid entries to the results list.
6. Stop the crawler once an enforcement action predates the input month and year.
7. Increment the page counter and wait one second before requesting the next page.
8. Convert the collected results into a DataFrame and save it as enforcement_actions_year_month.csv.

Choice of loop: while loop, as the number of pages is unknown and the stopping condition depends on enforcement action dates.
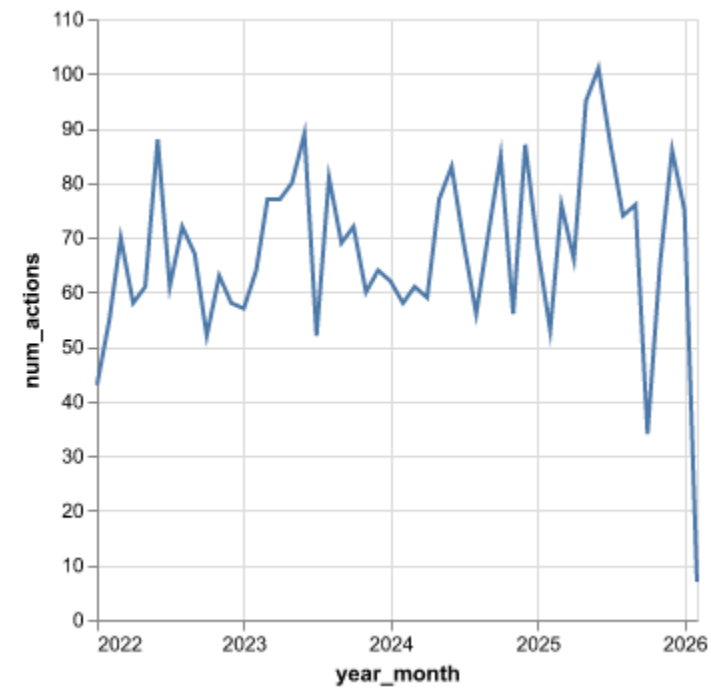
• b. Create Dynamic Scraper
• c. Test Your Code

The final dataframe contains 3,377 enforcement actions. The earliest enforcement action scraped is dated January 4, 2022.

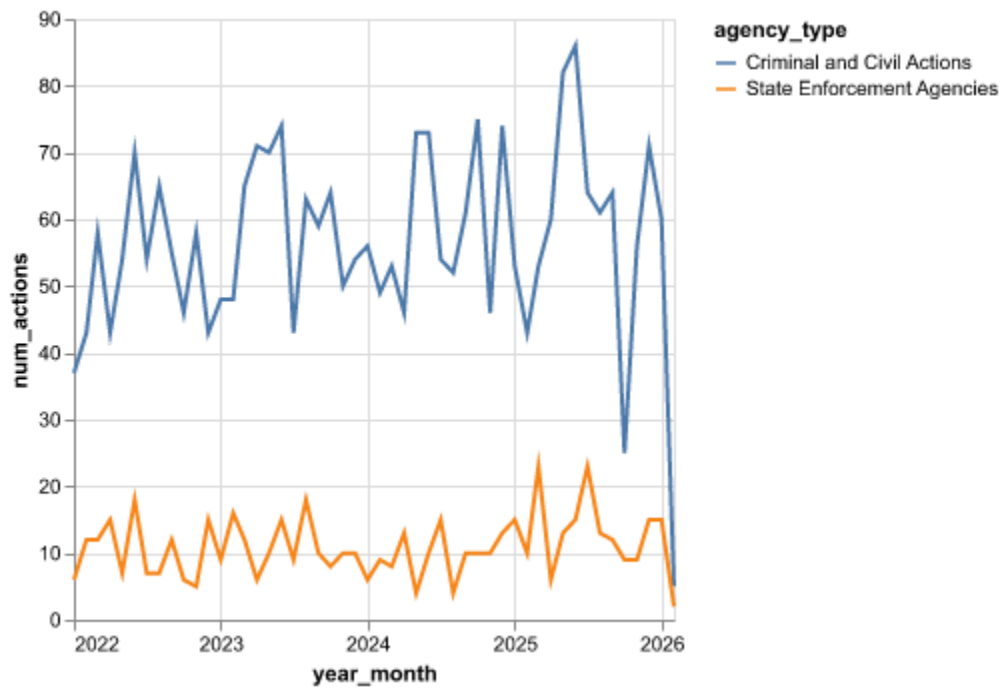|	Title	Date	Link
3376	Integrated Pain Management Medical Group Agree...	2022-01-04	https://oig.hhs.gov/fraud/ enforcement/integrat...

Step 3: Plot data based on scraped data

1. Plot the number of enforcement actions over time

2. Plot the number of enforcement actions categorized:

- based on "Criminal and Civil Actions" vs. "State Enforcement Agencies"

- based on five topics