



AGENCY FOR HEALTHCARE RESEARCH AND QUALITY



Traffic tracker: A gentle introduction to Python, APIs, and GitHub actions

Emily M. Mitchell

June 15, 2023

Disclaimer



The views expressed in this presentation are those of the author and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred.

Background



We track website views on our .gov sites



We post programming materials like example R code on GitHub



>> Need to track website views on GitHub repositories

Problem

GitHub traffic only available for the last 2 weeks



GitHub API
me/myrepo/traffic/views

How to store traffic data for longer?

Process



Write

Python code

that queries GitHub API  and outputs data to csv



GitHub API

me/myrepo/traffic/views

```
# Define API endpoint
traffic_views = "https://api.github.com/repos/HHS-AHRQ/MEPS/traffic/views"

# Make request
r = requests.get(traffic_views, headers = headers)

# Convert to pandas dataframe
traffic_df = pd.DataFrame(r.json()['views'])

# output to csv
traffic_df.to_csv(
    f"traffic_reports/traffic_MEPSrepo_{date.today()}.csv",
    index = False)
```




Process

1



Generate personal access token

from github.com/settings/developers

 scope: repo

2



Write Python code

that queries GitHub API  and outputs data to csv

```
# Set headers, using token
headers = {
    "Authorization": "token ghp_Iehoi9notarealtoken2908sh", 
    "User-Agent": "testing-requests-5801",
    "Accept": "application/vnd.github.v3+json",
}

# Define API endpoint
traffic_views = "https://api.github.com/repos/HHS-AHRQ/MEPS/traffic/views"

# Make request
r = requests.get(traffic_views, headers = headers)

# Convert to pandas dataframe
traffic_df = pd.DataFrame(r.json()['views'])

# output to csv
traffic_df.to_csv(
    f"traffic_reports/traffic_MEPSrepo_{date.today()}.csv",
    index = False)
```



Process


1



Generate
personal access token
from github.com/settings/developers

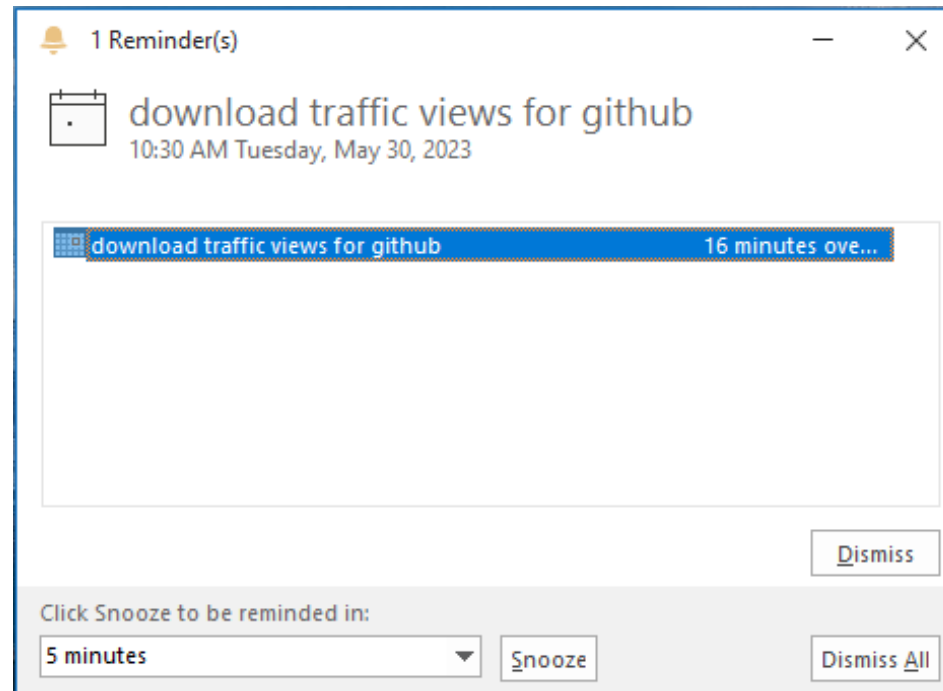
2



Write
Python code
that queries GitHub API 
and outputs data to csv

3

Automate



Process


1



Generate
personal access token
from github.com/settings/developers

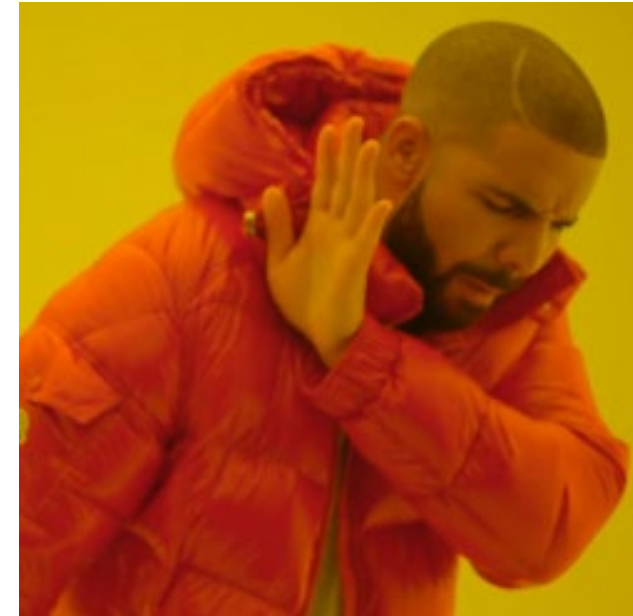
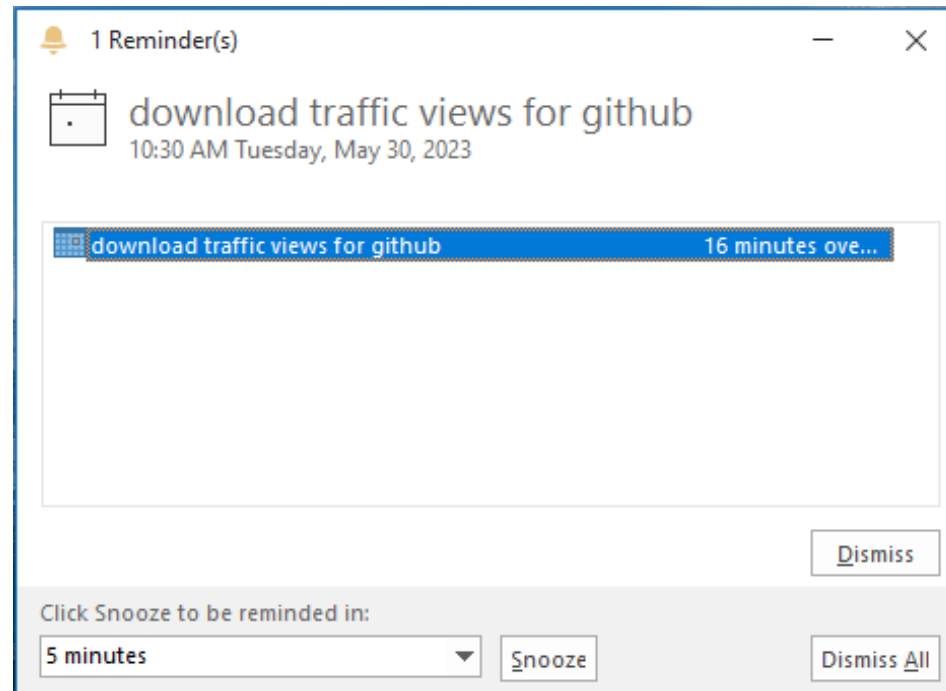
2



Write
Python code
that queries GitHub API 
and outputs data to csv

3

Automate



Process

1



Generate

personal access token

from github.com/settings/developers

2



Write

Python code

that queries GitHub API 
and outputs data to csv

3



Automate

with **GitHub Actions**

to run every day
and commit csv's to repository


```
name: traffic tracker

on:
  # Cron job should run every day at 6pm
  schedule:
    - cron: "0 6 * * *"

jobs:
  build:
    runs-on: ubuntu-latest

    steps:
      - name: Checkout repository
        uses: actions/checkout@v2

      ...

      - name: Run Python script
        env:
          GH_TOKEN: ${{ secrets.GH_TOKEN }} 
        run: python github_api_traffic.py

      - name: Commit changes
        run: |
          git add .
          git commit -m "Add generated CSV traffic files"
          git push
```



Process


1



Generate
personal access token
from github.com/settings/developers

2



Write
Python code
that queries GitHub API 
and outputs data to csv

3

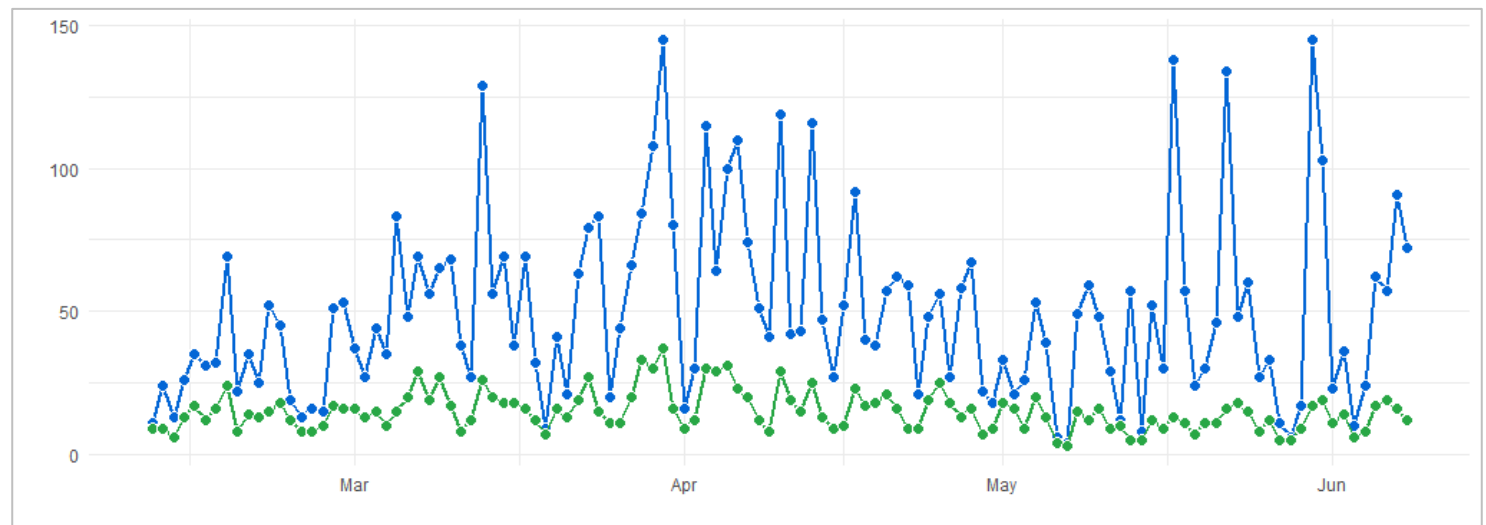
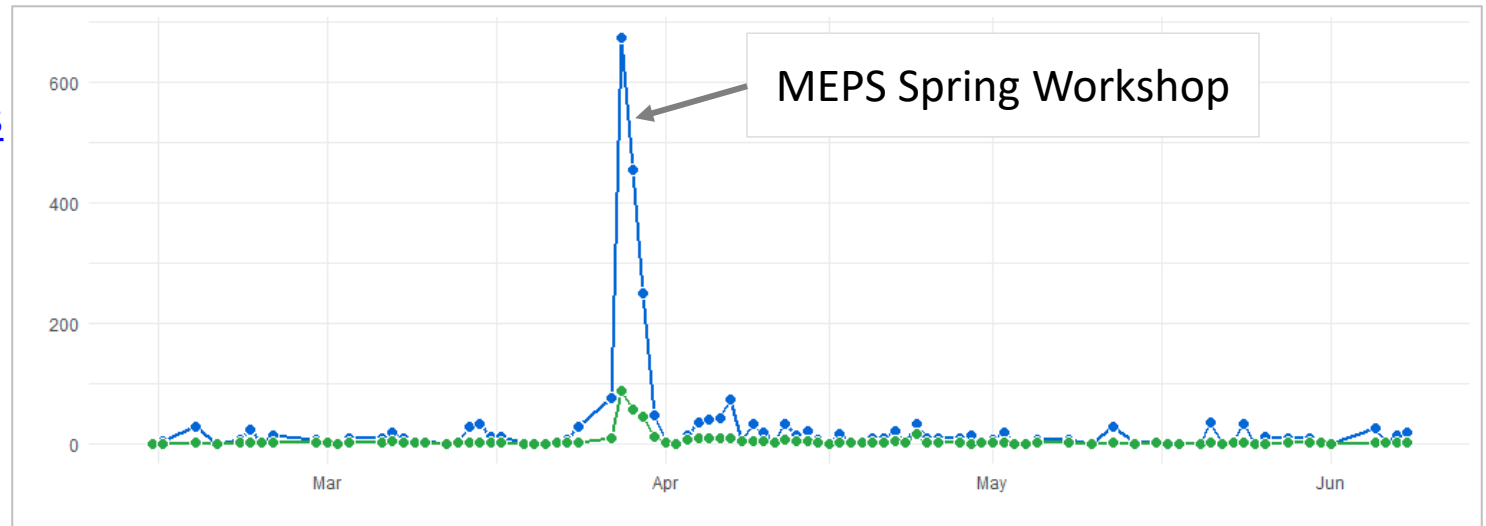


Automate
with **GitHub Actions**
to run every day
and commit csv's to repository

4



Visualize
make some nice graphs!



Known Issues / hurdles

GitHub connections may be subject to firewall constraints

R code to hit GitHub API failed, but Python code worked

If you are getting a permissions error, make sure your workflow permissions are set to 'Read and Write Permissions'

Workflow permissions

Choose the default permissions granted to the GITHUB_TOKEN when running workflows in this repository. You can specify more granular permissions in the workflow using YAML. [Learn more.](#)

Read and write permissions

Workflows have read and write permissions in the repository for all scopes.

Read repository contents and packages permissions

Workflows have read permissions in the repository for the contents and packages scopes only.

Choose whether GitHub Actions can create pull requests or submit approving pull request reviews.

Allow GitHub Actions to create and approve pull requests

Known Issues / hurdles



GitHub connections may be subject to firewall constraints

R code to hit GitHub API failed, but Python code worked

If you are getting a permissions error, make sure your workflow permissions are set to 'Read and Write Permissions'

GitHub is always changing things, so this talk will likely be obsolete in the future

Thank you!

Check it out



<https://github.com/e-mitchell/traffic-tracker>

emily.mitchell@ahrq.hhs.gov