Using APIs to dynamically update U.S. state maps

DOL API CoP Presentation

Ali Ruth, PhD

2023-04-17

Session overview

I. Problem statement: COVID in nursing homes
II. Intro to the Socrata API
III. Demo: Mapping COVID-19 risk to residents and staff in U.S. nursing homes
IV. Q&A

Summary of tools used in demo

- **Socrata Developer Platform**
  - Homepage: https://dev.socrata.com/
  - Learning resources: https://dev.socrata.com/consumers/getting-started.html

- **Tidycensus R package**
  - Homepage: https://walker-data.com/tidycensus/
  - Learning resources: https://walker-data.com/census-r/index.html

Background

COVID-19 has had a devastating impact on residents and staff in U.S. long-term care facilities (LTCFs).

- Mortality among residents in skilled nursing facilities (SNFs) increased by one-third during 2020
- SNF residents account for less than 1% of the U.S. population but accounted for 32% of all U.S. COVID deaths
- LTCF staff had one of the deadliest jobs in 2020-2021

Reference: McGarry BE, Grabowski DC. Nursing homes and COVID-19: a crisis on top of a crisis. The ANNALS of the American Academy of Political and Social Science. 2021 Nov;698(1):137-62. https://journals.sagepub.com/doi/full/10.1177/0002716221106150 9

Background

- Long-term care settings were prioritized in Tier 1A for nationwide vaccination in late 2020

- Two biggest risk factors for SNF Covid outbreaks:
  1) Nursing home size
  2) Community-level Covid spread

Reference: Konetzka RT, White EM, Pralea A, Grabowski DC, Mor V. A systematic review of long-term care facility characteristics associated with COVID-19 outcomes. Journal of the American Geriatrics Society. 2021 Oct;69(10):2766-77. https://agsjournals.onlinelibrary.wiley.com/doi/abs/10.1111/jgs.17434

Background

- Long-term care settings were prioritized in Tier 1A for nationwide vaccination in late 2020

- Two biggest risk factors for SNF Covid outbreaks:
  1) Nursing home size
  2) Community-level Covid spread

Reference: Konetzka RT, White EM, Pralea A, Grabowski DC, Mor V. A systematic review of long-term care facility characteristics associated with COVID-19 outcomes. Journal of the American Geriatrics Society. 2021 Oct;69(10):2766-77. https://agsjournals.onlinelibrary.wiley.com/doi/abs/10.1111/jgs.17434

Motivating question

In a given state, what nursing homes are/were at elevated risk for Covid-19 outbreaks at a specified time point, based on the latest community-level transmission rates?

Initial approach

1. Download entire CDC county-level Covid dataset
2. Filter most recent date of interest
3. Filter state of interest
4. Create map with the resulting data extract

Initial approach

1. Download entire CDC county-level Covid dataset
2. Filter most recent date of interest
3. Filter state of interest
4. Create map with the resulting data extract

Problem: Inefficient; large file sizes & repetitive operations

Initial approach

1. Download entire CDC county-level Covid dataset
2. Filter most recent date of interest
3. Filter state of interest
4. Create map with the resulting data extract

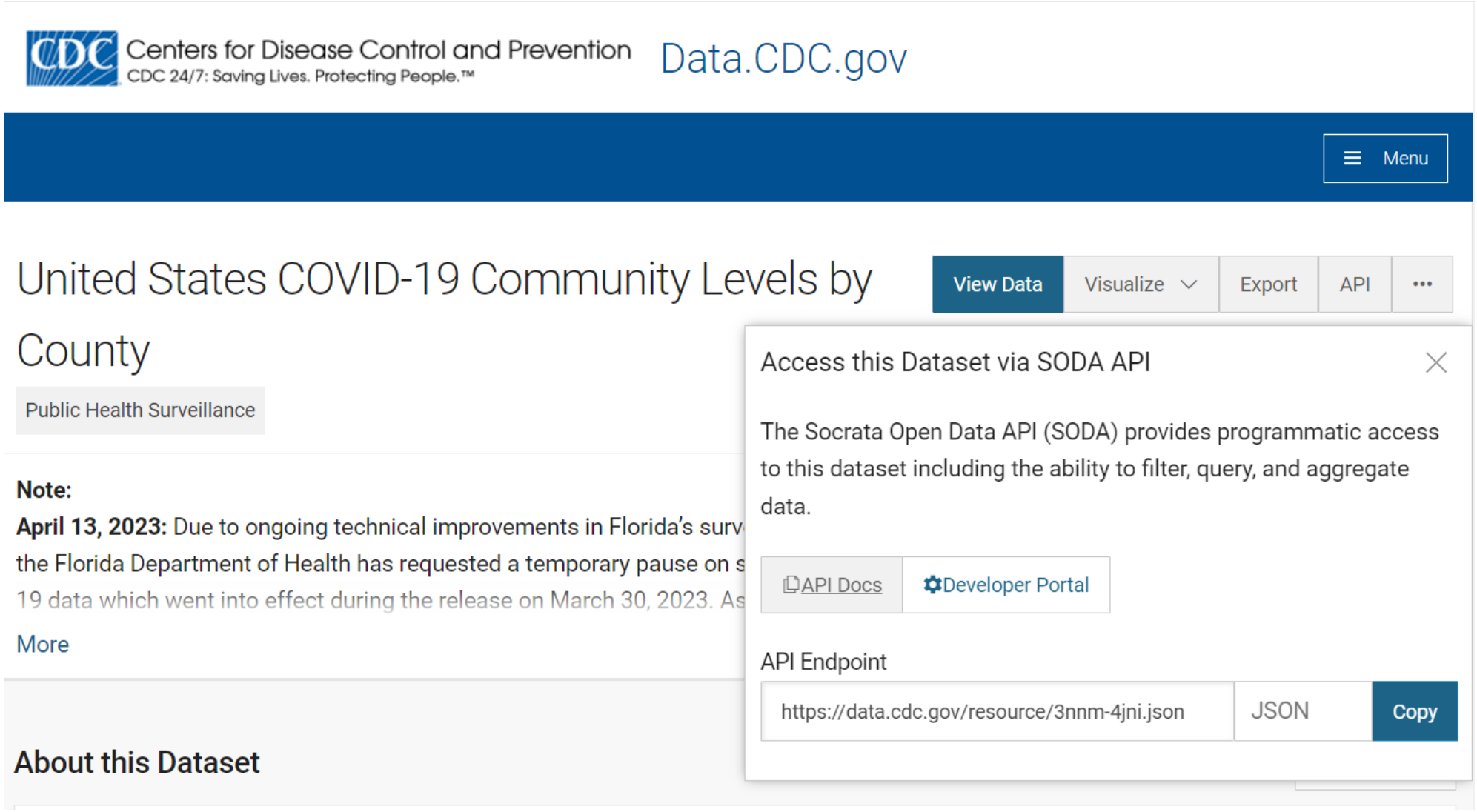
Problem: Inefficient; large file sizes & repetitive operations

✓ **Solution: Use an API!**

Why use an API for this?

- File size/manageability
  - Entire Covid dataset was quite large & only one time point was needed
- Automation
  - Didn’t want to manually import & clean the same file each time
- Consistency and standardization
  - Same approach can be applied across ALL states of interest

CDC COVID-19 county data



Socrata is great for learning to use APIs!

Code Snippets

The following are grab-and-go code code samples you can use with popular programming languages and data science tools.

jQuery | Python Pandas | PowerShell | RSocrata | SAS | soda-ruby | SODA.NET | Stata

The City of Chicago and community maintains a great `RSocrata` package on Github.

```
## Install the required package with:
## install.packages("RSocrata")

library("RSocrata")

df <- read.socrata(
  "https://data.cdc.gov/resource/3nnm-4jni.json",
  app_token = "YOURAPPTOKENHERE",
  email     = "user@example.com",
  password  = "fakepassword"
)
```

Fun with Socrata!

- Great entry point for API learning
- Easy to grab code snippets
- Lots of federal & state datasets to explore

- Check out: Central Park Squirrel Census Dataset

Building an API query in Socrata

https://data.cdc.gov/resource/3nnm-4jni.json?

date_updated=2023-01-05T00:00:00.000&

state=Maryland

…

Demo time: Let’s go!

CDC page for dataset:

https://data.cdc.gov/Public-Health-Surveillance/United-States-COVID-19-Community-Levels-by-County/3nnm-4jni

Socrata page:

https://dev.socrata.com/foundry/data.cdc.gov/3nnm-4jni

Dataset-specific things to watch out for…

- Changing variable names over time

community_covid_transmission → covid_19_community_level

Dataset-specific things to watch out for…

- Changing data upload frequencies

Daily → Weekly

Dataset-specific things to watch out for…

- Data quality caveats

See extensive notes about state-level heterogeneity at top of Socrata page with CDC COVID data!

Takeaways and general API thoughts

- **API-accessible federal datasets that include geographic variables are a very powerful tool!**
- Good to include all potentially useful geographic variables in a dataset (state name, state abbrev, fips code, coordinates)
- Consistent & logical variable naming is key
- User-friendly API guidance with many types of code samples is very helpful

Questions?

Thank you!

Ali Ruth

Email: ali.r.ruth@gmail.com

Twitter: @AliRRuth

GitHub: @alexandraruth

LinkedIn: https://www.linkedin.com/in/aliruth