Using APIs to dynamically update U.S. state maps

DOL API CoP Presentation

Ali Ruth, PhD

2023-04-17

Session overview

-Problem statement: COVID in nursing homes

-Intro to the Socrata API

-Demo: Mapping COVID-19 risk to residents and staff in U.S. nursing homes

-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

Background, continued

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

-Two biggest risk factors for SNF Covid outbreaks: Nursing home size and community-level Covid spread

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

-Download entire CDC county-level Covid dataset

-Filter most recent date of interest

-Filter state of interest

-Create map with the resulting data extract

Problem: Inefficient; requires repetitive data manipulation steps

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 is housed on Socrata

Socrata is great for learning to use APIs!

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

API demo: 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

-Changing data upload frequencies

-Data quality caveats

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