Accessing data via APIs

HAS Tools - Sept 27

Next forecasting assignment announcement!

- Due at midnight a week from today
- You will need to produce forecasts for the 2 weeks starting at 10/3 and 10/10 respectively
- Your forecasts will need to use data pulled in directly from the USGS Streamflow Database (you will learn how to do this today)
- Your forecasts will need to be based on a regression model (you will learn how to do this on Thursday)
- I will add a starter script to your homework repos after class today which gives you a baseline and walks you through the steps I am expecting

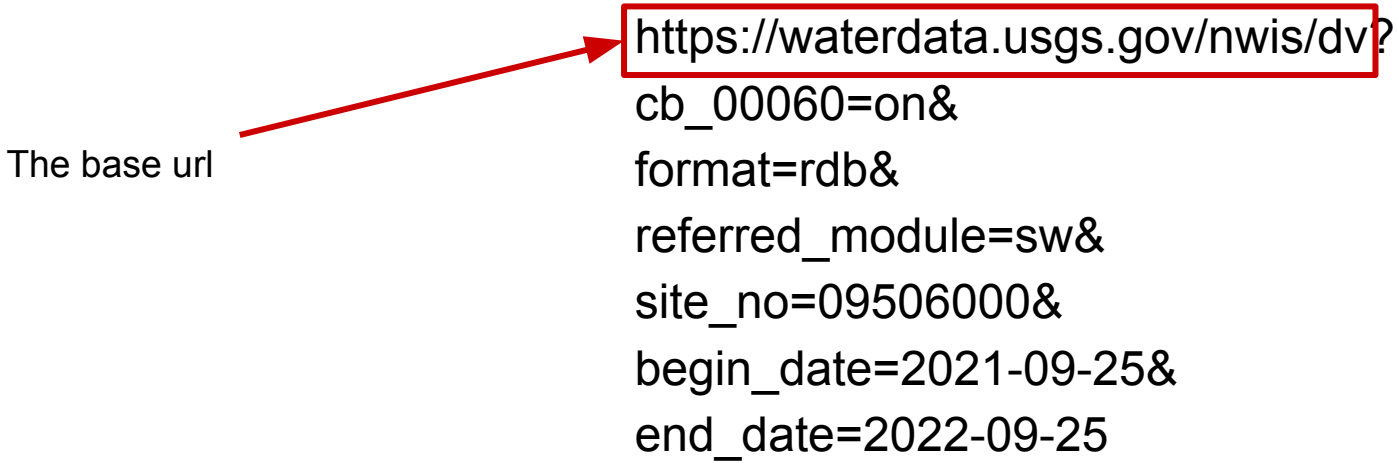
API background

- API means “Application programming interface”
- Basically it’s just a specification of how a piece of software can be used by other pieces of software.
- You’ve already seen some basic forms of API - for instance the `np.loadtxt` function. It has very rigid parameters that it can take in to produce some output that your programs used.
- Let’s see how the USGS exposes a web API for downloading streamflow data

Verde river data

- Previously I just downloaded the data and put it on GitHub for you.
- However, I got it from the USGS website: https://waterdata.usgs.gov/nwis/uv?site_no=09506000

What's in a url?



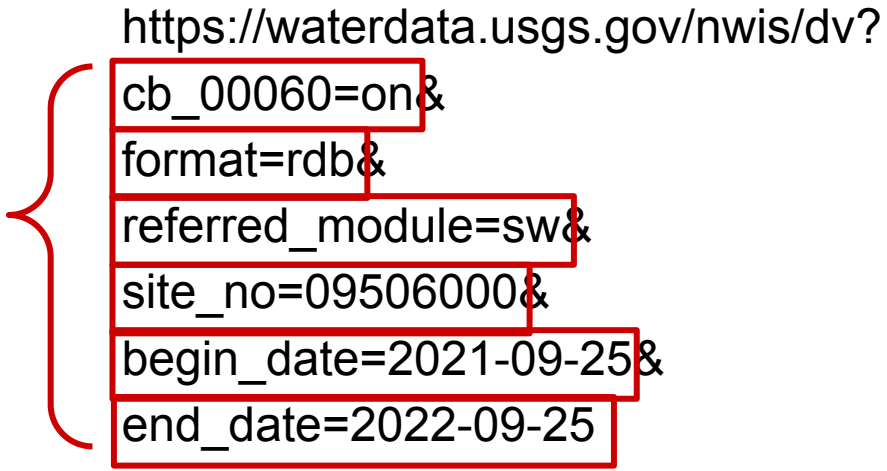
What's in a url?

"?" indicates the beginning of a query

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no=09506000&
begin_date=2021-09-25&
end_date=2022-09-25

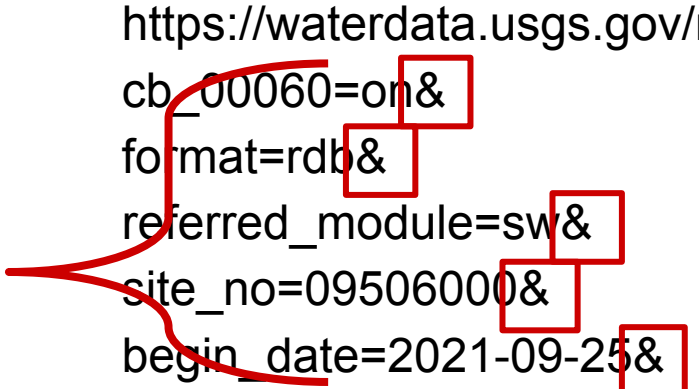
What's in a url?

Query parameters are key-value pairs, separated by the "=" sign



What’s in a url?

Query parameters are separated by the “&” sign



What's in a url?

cb_00060 is USGS jargon code for discharge in CFS

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no=09506000&
begin_date=2021-09-25&
end_date=2022-09-25

What's in a url?

"rdb" format is the USGS tab-separated table format

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no=09506000&
begin_date=2021-09-25&
end_date=2022-09-25

What's in a url?

I *think* this one means the surface water module, but not 100% sure on that

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no=09506000&
begin_date=2021-09-25&
end_date=2022-09-25

What's in a url?

The USGS site ID for Verde River near Camp Verde, AZ

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no=09506000&
begin_date=2021-09-25&
end_date=2022-09-25

What's in a url?

Start and end dates to retrieve data for

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no=09506000&
begin_date=2021-09-25&
end_date=2022-09-25

You can check the url here

https://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=rdb&referred_module=sw&site_no=09506000&begin_date=2021-09-25&&end_date=2022-09-25

This is a very rigid data format, so you might be thinking we can write some python code to automate this

https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no={VERDE_SITE_ID}&
begin_date={FORECAST_BEGIN_PERIOD}&
end_date={FORECAST_END_PERIOD}

This is a very rigid data format, so you might be thinking we can write some python code to automate this

Programmable parameter values!

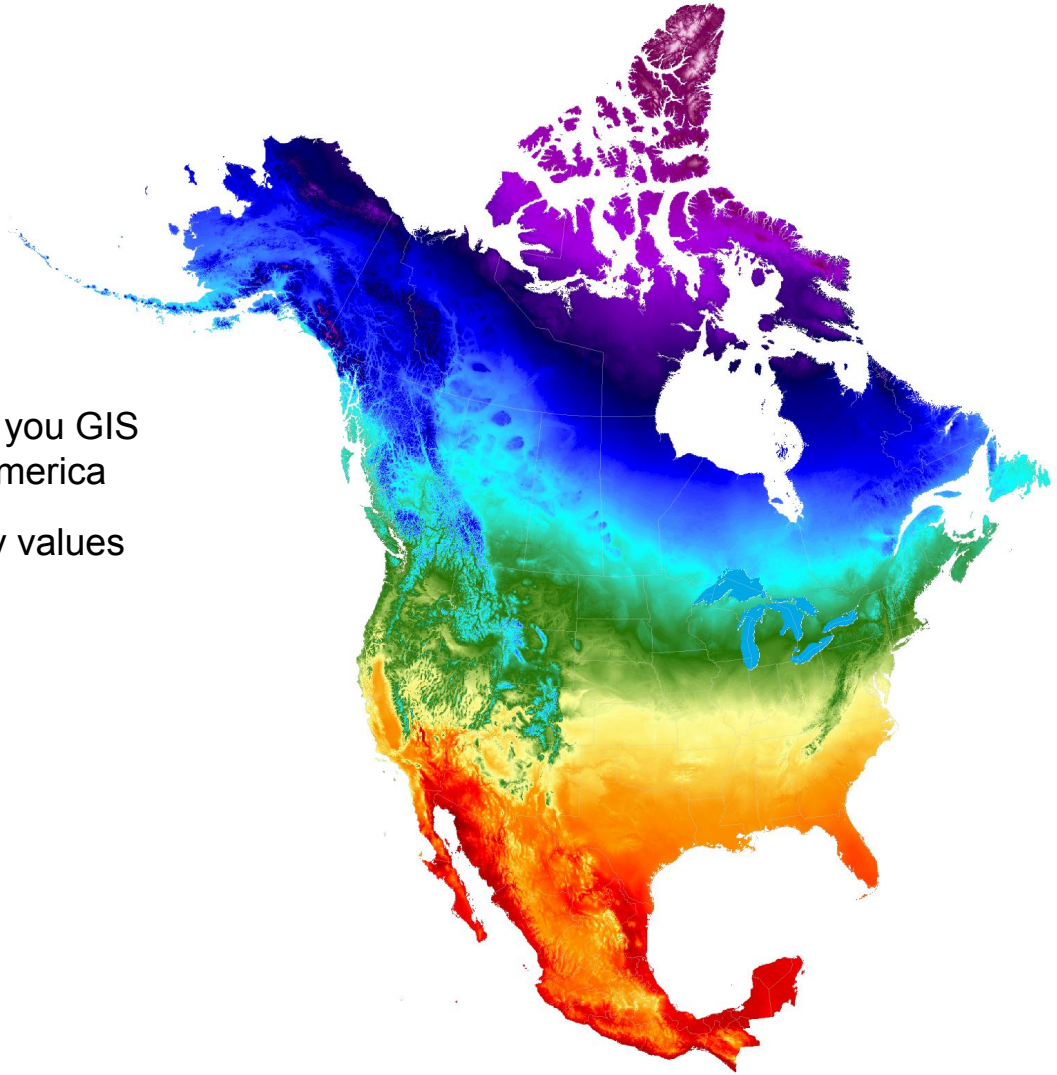
https://waterdata.usgs.gov/nwis/dv?
cb_00060=on&
format=rdb&
referred_module=sw&
site_no={VERDE_SITE_ID}&
begin_date={FORECAST_BEGIN_PERIOD}&
end_date={FORECAST_END_PERIOD}


Let's jump to GitKraken
SICKOS

Break

DayMet overview

- Daymet is a gridded (that is, raster for you GIS heads) product covering all of North America
- It contains long-term estimates of daily values for weather and climate variables
- More details can be found at https://daymet.ornl.gov/




BACK TO VSCODE
SICKOS