

# ACCESS HYDROLOGICAL DATASET

GLY606 Water Data Analysis & Modeling

Sep 16<sup>th</sup> 2024



# Homework #2

- It is due 1pm, Sep 18<sup>th</sup> (Wednesday)
- How to open .tar.gz file?

- `tar -xvzf gly606_hw2.tar.gz`

```
(base) jovyan@jupyter-yifancheng-5fub:~/assignment/assignment_2$ ls  
GLY606_Assignment_2_bash.pdf  gly606_hw2.tar.gz  
(base) jovyan@jupyter-yifancheng-5fub:~/assignment/assignment_2$ tar -xvzf gly606_h  
w2.tar.gz > unzip.log  
(base) jovyan@jupyter-yifancheng-5fub:~/assignment/assignment_2$ ls  
GLY606_Assignment_2_bash.pdf  gly606_hw2  gly606_hw2.tar.gz  unzip.log  
(base) jovyan@jupyter-yifancheng-5fub:~/assignment/assignment_2$
```

# Homework #2

- It is due 1pm, Sep 18<sup>th</sup> (Wednesday)
- How to open .tar.gz file?
  - `tar -xvzf gly606_hw2.tar.gz`
- What do I expect?

```
(base) jovyan@jupyter-yifancheng-5fub:~/assignment/assignment_2$ ls  
GLY606_Assignment_2_bash.pdf  gly606_hw2.tar.gz  hw2_script2.sh  model_run  
gly606_hw2                      hw2_script1.sh    hw3_script3.sh  unzip.log
```

# Homework #2

- It is due 1pm, Sep 18<sup>th</sup> (Wednesday)
- How to open **.tar.gz** file?
  - `tar -xvzf gly606_hw2.tar.gz`
- What do I expect?
- How to create folders based on txt files?
  - `for line in `cat site_list.txt`; do`
  - `!split line by comma, get site_id and year`
  - `mkdir ${site_id}`
  - `done`

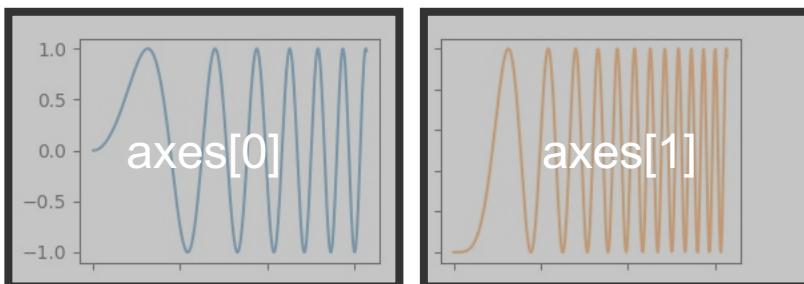
# Recap - Numpy

- How do we import numpy packages?
- Given an array `x=np.array([1,2,3])`, how do we add 3 to every element in `x`?
- What is the numpy function to generate a continuous array, such as [0,1,2,3,4,5]?
- What is the difference between `math.exp` and `np.exp`?
- What if we do not know how to use `np.eye`?
- What would happen if we convert a mixed-type list to an array, `my_list = ['Yifan', 31, 'want to be rich']`?
- Import numpy as np
- `x+3`
- `np.arange(6)` or `np.linspace(0, 5, 6)`
- `math.exp` can only be used to calculate the exponential for one value while `np.exp` can be used for an array
- Type `np.eye()` and press `<shift>+<tab>` within parenthesis.
- The array would be `['Yifan', '31', 'want to be rich']` 5

# Recap - matplotlib

- How to import the main interface to actual plotting functions?
- `plt.figure` versus `plt.subplots`

- What's the syntax if we want two subplots (1 row, 2 columns)?



- What is the syntax to do a line plot?
- What is the syntax to do a scatter plot?

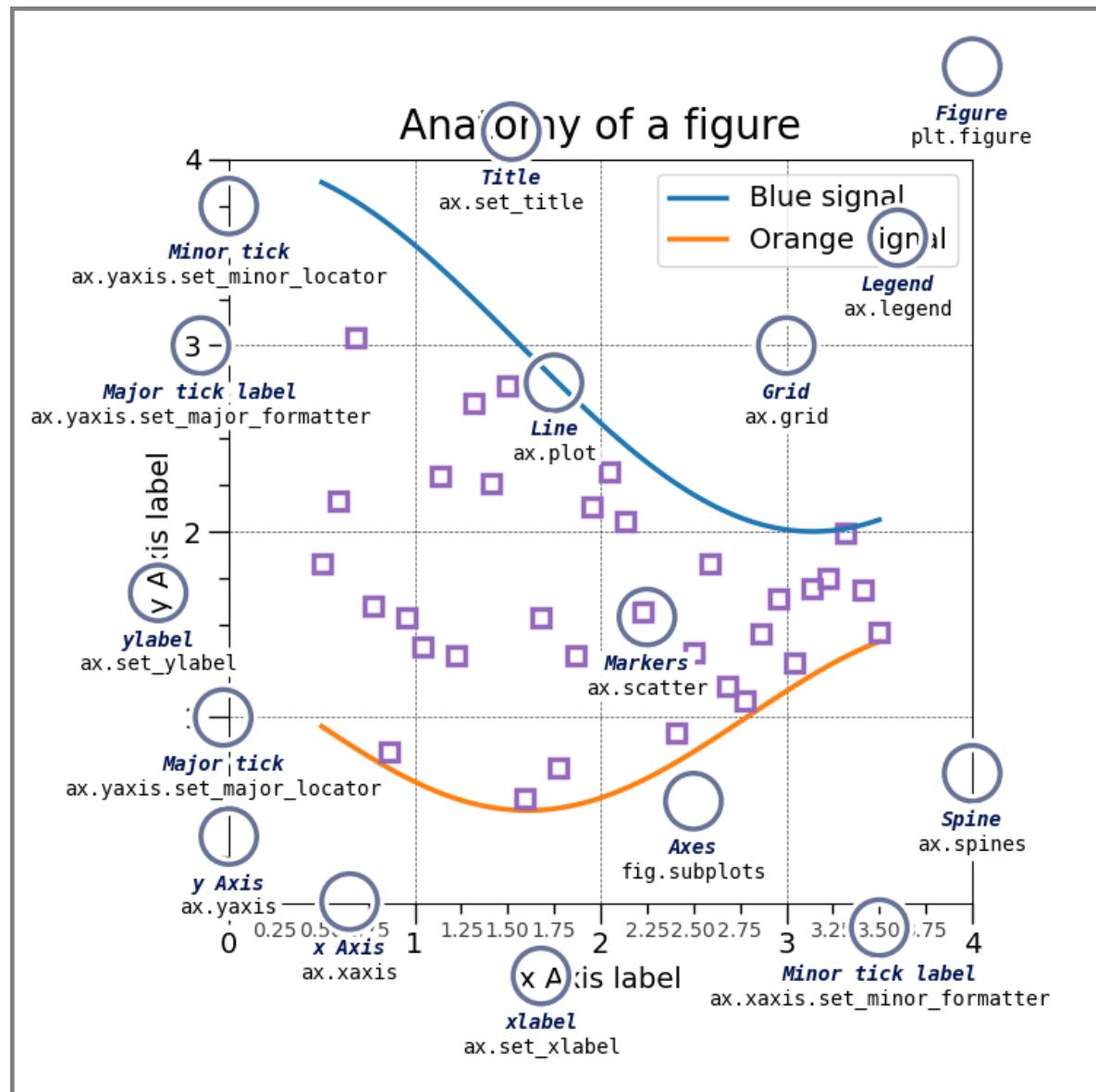
- Import `matplotlib.pyplot as plt`
- `plt.figure` only creates the figure and you manually add subplots, which is usually used to make single figure. `plt.subplots` creates both the figure and subplots in one call, making it simpler for layouts with multiple plots.
- `fig, axes=plt.subplots(1, 2)`
- `plt.plot()`
- `plt.scatter()`

# Recap - matplotlib

- Anatomy of a figure!
- Credit:

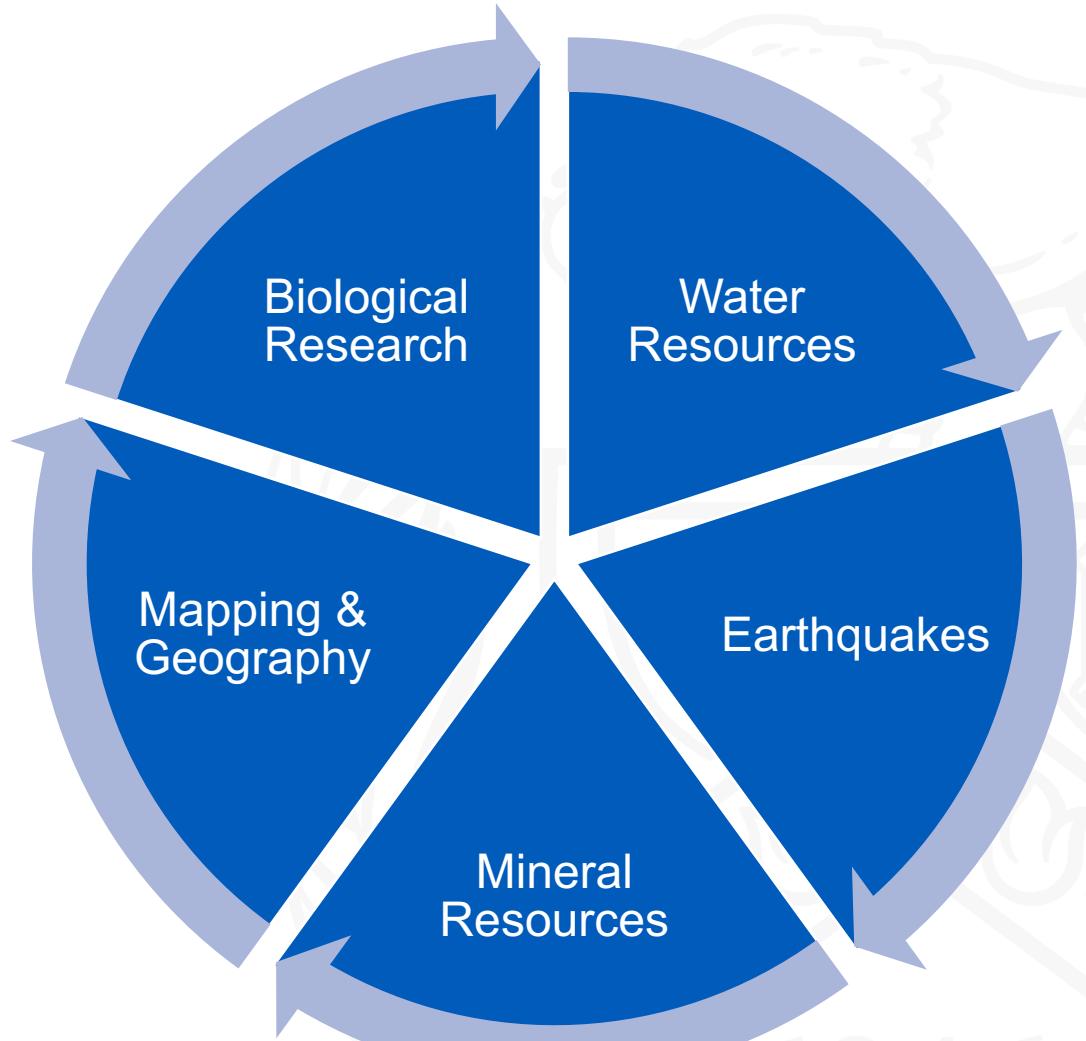
<https://matplotlib.org/stable/gallery/showcase/anatomy.html>

```
fig, ax=plt.subplots(1,1)
```



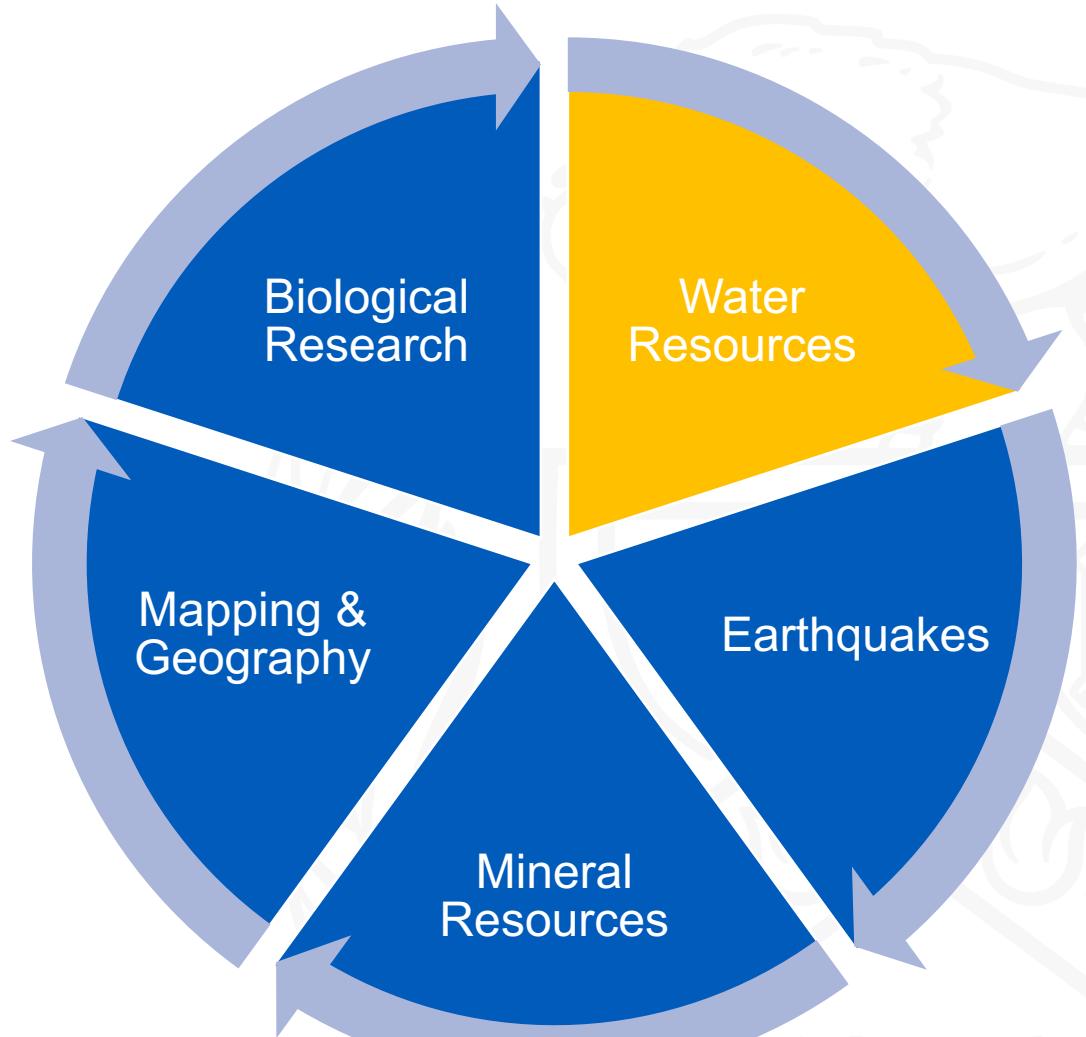
# U.S. Geological Survey

- A science bureau within the United States Department of the Interior
- Mission: *to provide reliable, impartial information about the Earth's natural systems, enabling sound decisions regarding land use, resource management, and hazard mitigation.*



# U.S. Geological Survey

- A science bureau within the United States Department of the Interior
- Mission: *to provide reliable, impartial information about the Earth's natural systems, enabling sound decisions regarding land use, resource management, and hazard mitigation.*



# National Water Information System (NWIS)

- A database of real-time and historical water data
- <https://waterdata.usgs.gov/nwis/sw>

## How can we download streamflow data?!



# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

## USGS Surface-Water Data for the Nation

### Current Conditions

(11,234 sites)

Current conditions at selected sites based on the most recent data from on-site automated recording equipment. Measurements are commonly recorded at a fixed interval of 15- to 60-minutes and transmitted to the USGS every hour. Values may include "Approved" (quality-assured data that may be published) and/or more recent "Provisional" data (of unverified accuracy and subject to revision). Most current data are provisional.

### Historical Observations

(19,248 sites)

The same data accessed by the Current Conditions link above but including both active and discontinued sites with data for any part of the period October 1, 2007, through the present. Values may include "Approved" (quality-assured data that may be published) and/or more recent "Provisional" data (of unverified accuracy and subject to revision).

### Daily Data

(29,803 sites)

Summary of all data for each day for the period of record and may represent the daily mean, median, maximum, minimum, and/or other derived value. Values may include "Approved" (quality-assured data that may be published) and/or more recent "Provisional" data (of unverified accuracy and subject to revision).

### Statistics

(27,794 sites)

Monthly

Annual

Daily

### Introduction

The U.S. Geological Survey's (USGS) National Water Information System (NWIS) is a comprehensive and distributed application that supports the acquisition, processing, and long-term storage of water data. Water Data for the Nation serves as the publicly available portal to a geographically seamless set of much of the water data maintained within NWIS.

Nationally, USGS surface-water data includes more than 850,000 station years of time-series data that describe stream levels, streamflow (discharge), reservoir and lake levels, surface-water quality, and rainfall. The data are collected by automatic recorders and manual [field measurements](#) at installations across the Nation.

Data are collected by field personnel or relayed through telephones or satellites to offices where it is stored and processed. The data relayed through the Geostationary Operational Environmental Satellite (GOES) system are processed automatically in near real time, and in many cases, [current data](#) are available online within minutes.

Once a complete day of readings are received from a site, [daily summary data](#) are generated and made available online. USGS finalizes data at individual sites on a continuous basis as environmental conditions and hydrologic characteristics permit.

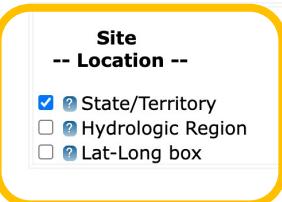
# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

## USGS Surface-Water Daily Data for the Nation

### Choose Site Selection Criteria

Choose from the following criteria to constrain the number of sites selected. By default, the "Site-type" criterion is preselected below and will default to surface water on the following page. If no additional site-selection criteria are chosen and no additional specifications are defined on the following page then output will be for all 29,803 surface water sites that have daily values data.

1   
Site  
-- Location --  
 [State/Territory](#)  
 [Hydrologic Region](#)  
 [Lat-Long box](#)

2 

[Questions or Comments](#)  
[Help](#)

Site  
-- Identifier --  
 [Site Name](#)  
 [Site Number](#)  
 [Multiple Site Numbers](#)  
 [Agency Code](#)  
 [File of Site Numbers](#)

Site  
-- Attribute --  
 [Site type](#)  
 [Drainage area](#)

Data  
-- Attribute --  
 [Number of observations](#)

[Data Tips](#)  
[Explanation of terms](#)  
[Subscribe for system changes](#)

# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

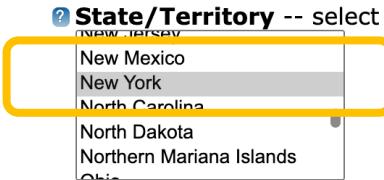
**2. Select “streamflow”**  
**Note: usually streamflow is available at ft<sup>3</sup>/s, not m<sup>3</sup>/s**

## USGS Surface-Water Daily Data for the Nation

### Select sites which meet all of the following criteria:

Define one or more values for each of the following site-selection criteria: --- or select [new criteria](#)

State/Territory -- select one or more



New Mexico  
New York  
North Carolina  
North Dakota  
Northern Mariana Islands  
Ohio

**1. Scroll to find New York**

Available parameters -- select sites that have data for the following parameters:  
 Select one or more parameters --or-- leave blank to select all:

#### Water Level/Flow Parameters

- Reservoir storage, acre-ft
- Stream velocity, ft/s
- Flow rate of well, gal/d
- Flow rate of well, gal/min
- Flow rate, instantaneous, gal/min
- Streamflow, ft<sup>3</sup>/s
- Discharge, instantaneous, ft<sup>3</sup>/s
- Elevation of reservoir water surface above datum, ft
- Mean depth of stream, ft
- Gage height, ft

#### Meteorological Parameters

- Temperature, air, °C
- Temperature, air, °F
- Barometric pressure, mmHg
- Incident solar radiation intensity, (cal/cm<sup>2</sup>)/d
- Wind speed, mph
- Wind direction, degrees clockwise from north
- Precipitation, total, in
- Precipitation, total, inches/week
- Relative humidity, percent
- Precipitation, duration of storm event, minutes

# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

**Choose Output Format**  
Display Summary of Selected Sites  
Choose one of the following options for displaying descriptions of the sites meeting the criteria above:

Show sites on a map NEW

Table of sites grouped by ▼

Scroll list of sites -- allows selection of data for multiple sites

Brief descriptions -- allows selection of data for multiple sites

Site-description information displayed in XML format ▼

(Select fields to include in site-description output)

Agency  
Site identification number  
Site name  
Site type

Save file of selected sites to local disk for future upload

**Retrieve USGS Surface-Water Daily Data for Selected Sites**  
Choose one of the following options for displaying data for the sites meeting the criteria above

Retrieve data for:

the previous 365 days (1 - 365) \*\*OR\*\*

for the date range: First date: 2024-09-14 Last date: 2024-09-15 (1838-01-01 through 2024-09-15) **2**

**Output Options:**

Graphs of data --  use arithmetic Y-axis for streamflow

Graphs of data with long-term statistics --  use arithmetic Y-axis for streamflow

Graphs of data without long-term statistics --  use arithmetic Y-axis for streamflow

Graphs of data with field measurements --  use arithmetic Y-axis for streamflow

Table of data

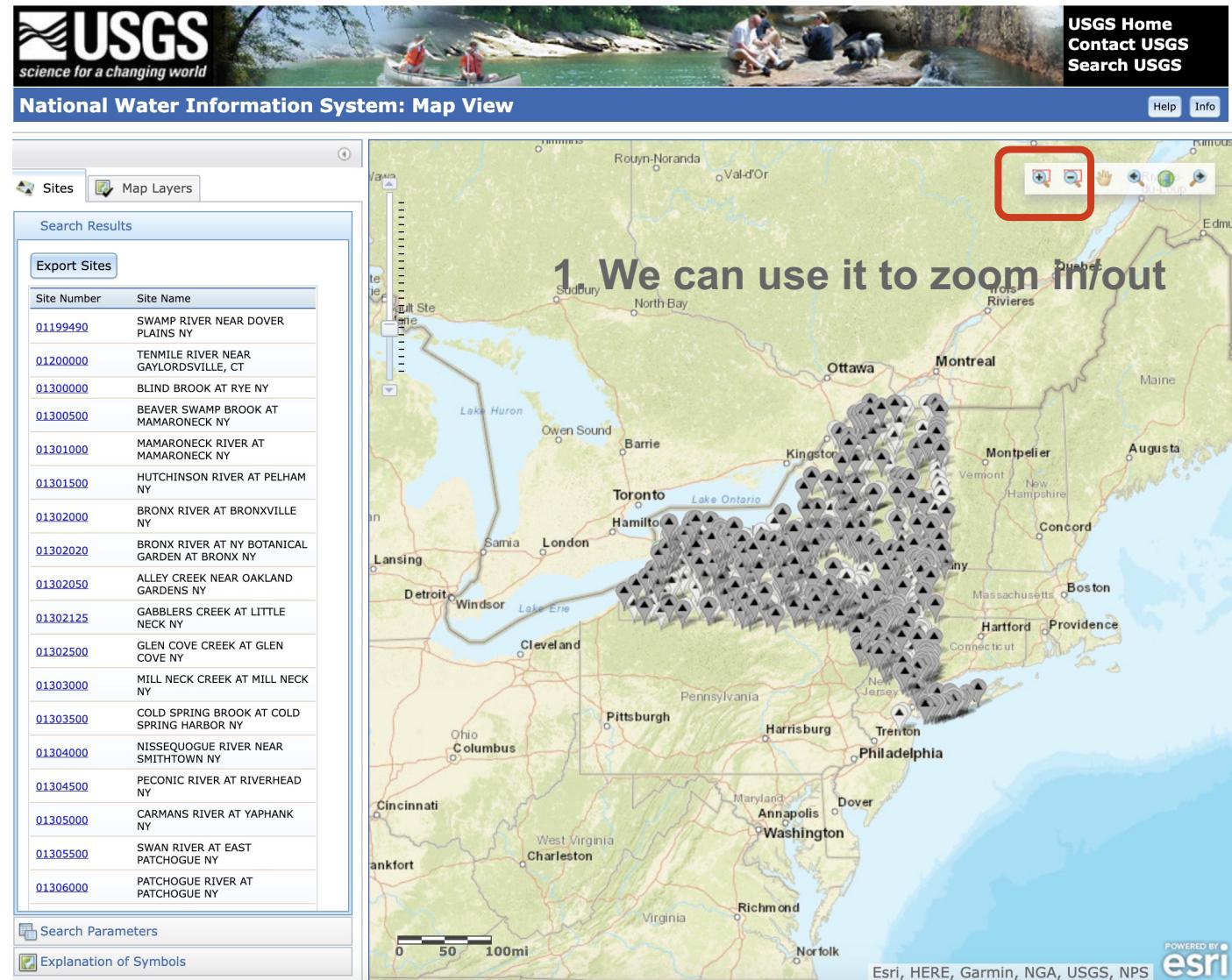
Tab-separated data YYYY-MM-DD  Save to file ▼ \*

\* Save compressed files with a .gz file extension.

**3. Click "submit"**

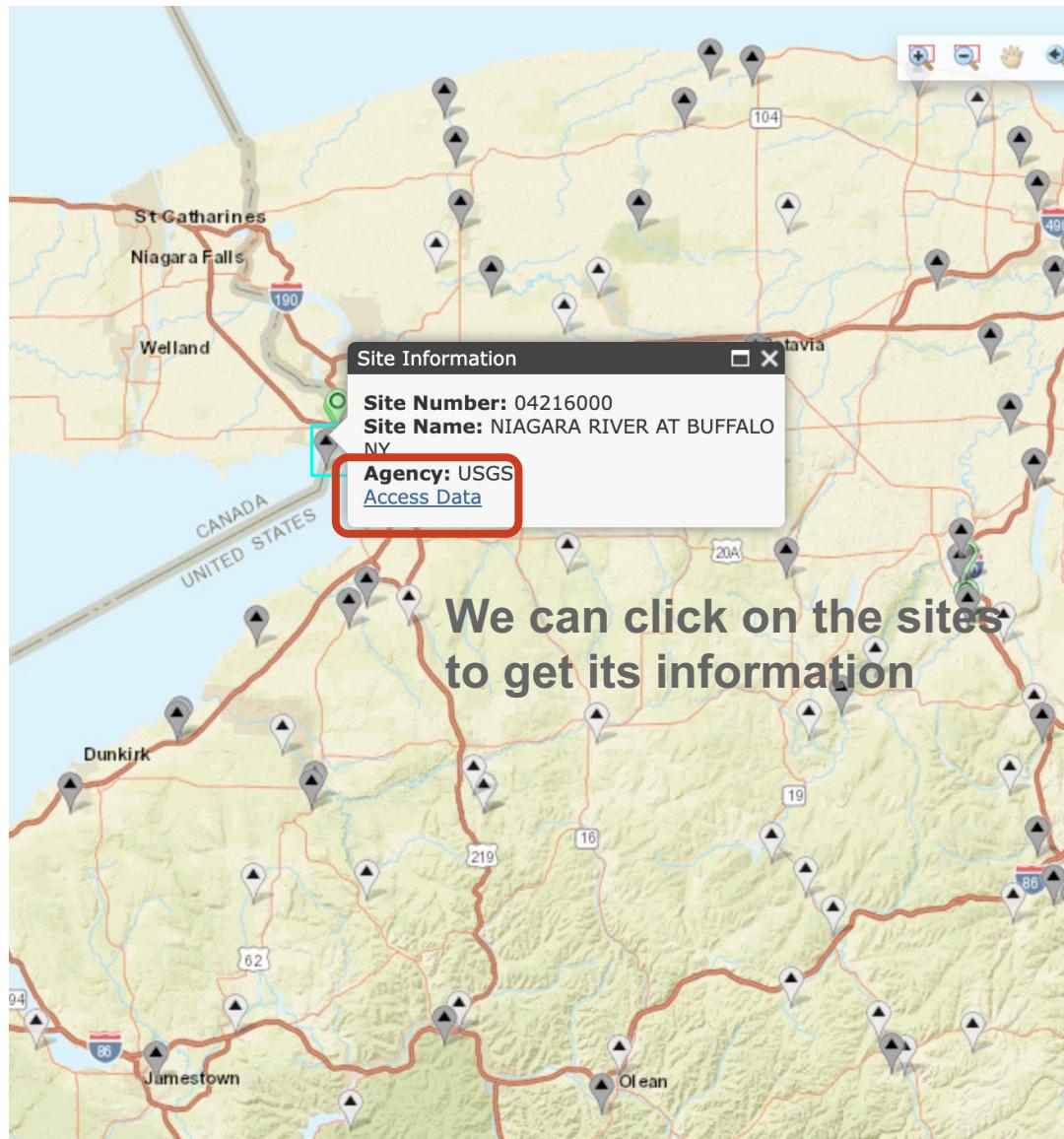
# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State



# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State



# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

Click “Daily data” to access daily streamflow

## USGS 04216000 NIAGARA RIVER AT BUFFALO NY

[Available data for this site](#) [SUMMARY OF ALL AVAILABLE DATA](#) [GO](#)

### Stream Site

#### DESCRIPTION:

Latitude 42°52'40", Longitude 78°54'59" NAD83  
Erie County, New York, Hydrologic Unit 04120200  
Drainage area: 263,700 square miles

#### AVAILABLE DATA:

Data Type	Begin Date	End Date	Count	
<a href="#">Daily Data</a>	Discharge, cubic feet per second	1926-01-01	2024-02-29	35852
<a href="#">Daily Statistics</a>	Discharge, cubic feet per second	1926-01-01	2024-02-28	35851
<a href="#">Monthly Statistics</a>	Discharge, cubic feet per second	1926-01	2024-02	
<a href="#">Annual Statistics</a>	Discharge, cubic feet per second	1926	2024	
<a href="#">Peak streamflow</a>	1960-11-30	2022-12-24	63	
<a href="#">Field/Lab water-quality samples</a>	1957-08-27	1957-08-27	1	
<a href="#">Water-Year Summary</a>	2005	2023	19	

#### OPERATION:

Record for this site is maintained by the USGS New York Water Science Center

Email questions about this site to [New York Water Science Center Water-Data Inquiries](#)

# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

## USGS 04216000 NIAGARA RIVER AT BUFFALO NY PROVISIONAL DATA SUBJECT TO REVISION

Available data for this site Time-series: Daily data

Available Parameters      Period of Record      Output format

All 1 Available Parameters for this site

00060 Discharge(Mean) [From USACOE Buffalo] 1926-01-01 2024-02-29

Graph

Graph w/ stats

Graph w/ (up to 3) parms

Table

Tab-separated

1

Days (365)

or

Begin date  3

End date

2. set "2022-09-01" as begin date,  
set "2023-09-01" as end date

[Summary of all available data for this site](#)  
[Instantaneous-data availability statement](#)

# USGS GUI

- <https://waterdata.usgs.gov/nwis/sw>
- Explore available data
  - Available streamflow observations in New York State

```
# -----  
# WARNING --  
# Some of the data that you have obtained from this U.S. Geological Survey database  
# may not have received Director's approval. Any such data values are qualified  
# as provisional and are subject to revision. Provisional data are released on the  
# condition that neither the USGS nor the United States Government may be held liable  
# for any damages resulting from its use.  
#  
# Additional info: https://waterdata.usgs.gov/provisional-data-statement/  
#  
# Contact: gs-w_waterdata_support@usgs.gov  
# retrieved: 2024-09-16 12:12:23 EDT (caww02)  
#  
# Data for the following 1 site(s) are contained in this file  
# USGS 04216000 NIAGARA RIVER AT BUFFALO NY  
# -----  
#  
# Data provided for site 04216000  
#      TS   parameter   statistic   Description  
#      217733     00060     00003   Discharge, cubic feet per second (Mean), [From USACOE Buffalo]  
#  
# Data-value qualification codes included in this output:  
#  
#      A Approved for publication -- Processing and review completed.  
#  
agency_cd    site_no  datetime      217733_00060_00003      217733_00060_00003_cd  
5s      15s    20d    14n   10s  
USGS 04216000 2022-09-01 227000 A  
USGS 04216000 2022-09-02 222000 A  
USGS 04216000 2022-09-03 223000 A  
USGS 04216000 2022-09-04 219000 A  
USGS 04216000 2022-09-05 213000 A  
USGS 04216000 2022-09-06 214000 A  
USGS 04216000 2022-09-07 220000 A  
USGS 04216000 2022-09-08 221000 A  
USGS 04216000 2022-09-09 222000 A  
USGS 04216000 2022-09-10 222000 A  
USGS 04216000 2022-09-11 220000 A  
USGS 04216000 2022-09-12 228000 A  
USGS 04216000 2022-09-13 225000 A  
USGS 04216000 2022-09-14 228000 A  
USGS 04216000 2022-09-15 216000 A  
USGS 04216000 2022-09-16 222000 A  
USGS 04216000 2022-09-17 219000 A  
USGS 04216000 2022-09-18 226000 A  
USGS 04216000 2022-09-19 227000 A  
USGS 04216000 2022-09-20 224000 A  
USGS 04216000 2022-09-21 223000 A  
USGS 04216000 2022-09-22 225000 A
```

# Is there simpler ways for us to easily access it?

- Yes!
- More than one way too!!
  - We will introduce the first way today!
  - Access the USGS data using urllib



## urllib

What's the url?

[https://waterdata.usgs.gov/nwis/dv?  
cb\\_00060=on&format=rdb&site\\_no=  
04216000&legacy=&referred\\_modul  
e=sw&period=&begin\\_date=2022-  
09-01&end\\_date=2023-09-01](https://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=rdb&site_no=04216000&legacy=&referred_module=sw&period=&begin_date=2022-09-01&end_date=2023-09-01)

**https://waterdata.usgs.gov/nwis/dv?**  
**cb\_00060=on&**  
**format=rdb&**  
**site\_no=04216000&**  
**referred\_module=sw&**  
**begin\_date=2022-09-01&**  
**end\_date=2023-09-01**

## urllib

What's the url?

The base url

**https://waterdata.usgs.gov/nwis/dv?**

**cb\_00060=on&**

**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**

## urllib

What's the url?

“?” indicates the beginning of a query

**https://waterdata.usgs.gov/nwis/dv?cb\_00060=on&format=rdb&site\_no=04216000&referred\_module=sw&begin\_date=2022-09-01&end\_date=2023-09-01**

# urllib

What's the url?

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

**[https://waterdata.usgs.gov/nwis/dv?](https://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=rdb&site_no=04216000&referred_module=sw&begin_date=2022-09-01&end_date=2023-09-01)**

**cb\_00060=on&**

**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**

# urllib

What's the url?

Query parameters are separated by the “&” sign

**https://waterdata.usgs.gov/nwis/dv?**

**cb\_00060=on&**

**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**

## urllib

What's the url?

cb\_00060 is USGS jargon  
code for discharge in CFS

USGS parameter code for other  
variables

[https://help.waterdata.usgs.gov/code/parameter\\_cd\\_query?fmt=rdb&group\\_cd=PHY&inline=true](https://help.waterdata.usgs.gov/code/parameter_cd_query?fmt=rdb&group_cd=PHY&inline=true)

**https://waterdata.usgs.gov/nwis/dv?**

**cb\_00060=on&**

**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**

## urllib

What's the url?

“rdb” format is the USGS  
tab-separated table format

**https://waterdata.usgs.gov/nwis/dv?**

**cb\_00060=on&**

**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**

## urllib

What's the url?

The USGS site ID for  
Niagara River at Buffalo NY

**https://waterdata.usgs.gov/nwis/dv?**

**cb\_00060=on&**

**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**



## urllib

What's the url?

"sw" refers to the surface water module

**https://waterdata.usgs.gov/nwis/dv?**

**cb\_00060=on&**

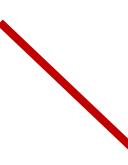
**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**



## urllib

What's the url?

Start and end dates to  
retrieve data for

**[https://waterdata.usgs.gov/nwis/dv?](https://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=rdb&site_no=04216000&referred_module=sw&begin_date=2022-09-01&end_date=2023-09-01)**

**cb\_00060=on&**

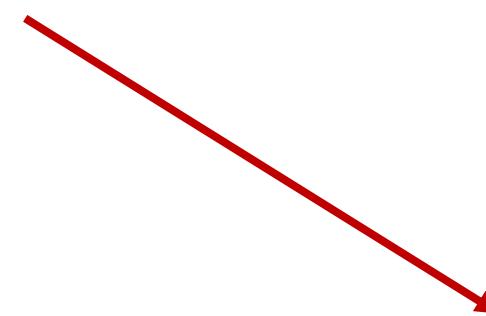
**format=rdb&**

**site\_no=04216000&**

**referred\_module=sw&**

**begin\_date=2022-09-01&**

**end\_date=2023-09-01**



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

- Download the exercise on your laptop
  - [https://github.com/act-hydro/GLY606\\_2024/blob/main/in\\_class\\_practice/python\\_practice/python\\_inclass\\_4\\_data\\_access\\_urllib.ipynb](https://github.com/act-hydro/GLY606_2024/blob/main/in_class_practice/python_practice/python_inclass_4_data_access_urllib.ipynb)
- Practice it on CUAHSI Jupyterhub

**https://waterdata.usgs.gov/nwis/dv?**  
**cb\_00060=on&**  
**format=rdb&**  
**site\_no=<SITE\_ID>&**  
**referred\_module=sw&**  
**begin\_date=<START\_DATE>&**  
**end\_date=<END\_DATE>**