ENGR 112: Engineering Decisions

Week # 9: Uploading, manipulating, and plotting data

DEPARTMENT OF
ENGINEERING
JAMES MADISON UNIVERSITY

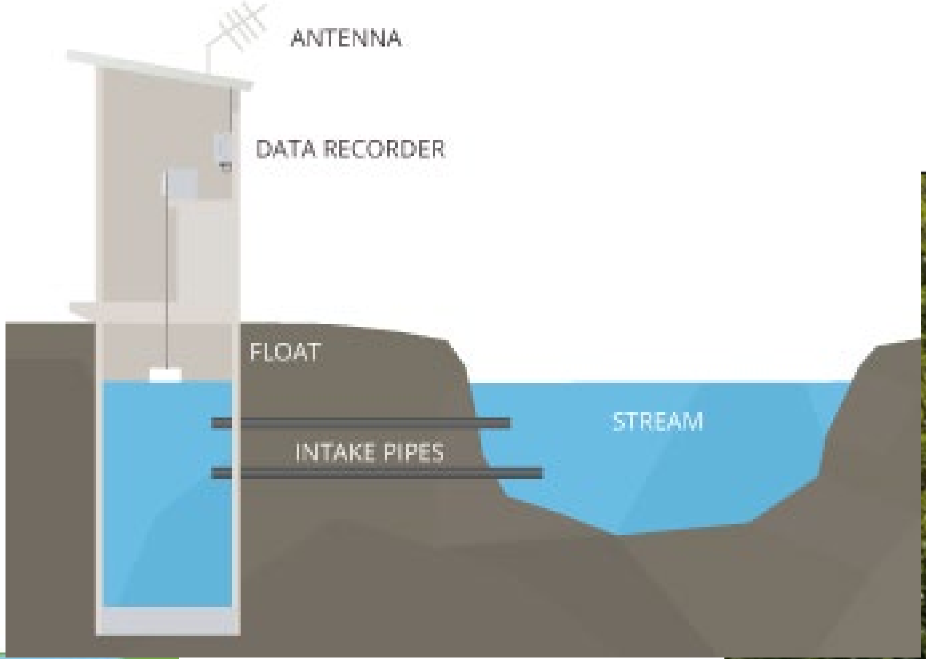
Spring 2022

Jason Forsyth, Ph.D.
Kyle Gipson, Ph.D.
Shraddha Joshi, Ph.D.
Steven Woodruff, ABD

Background

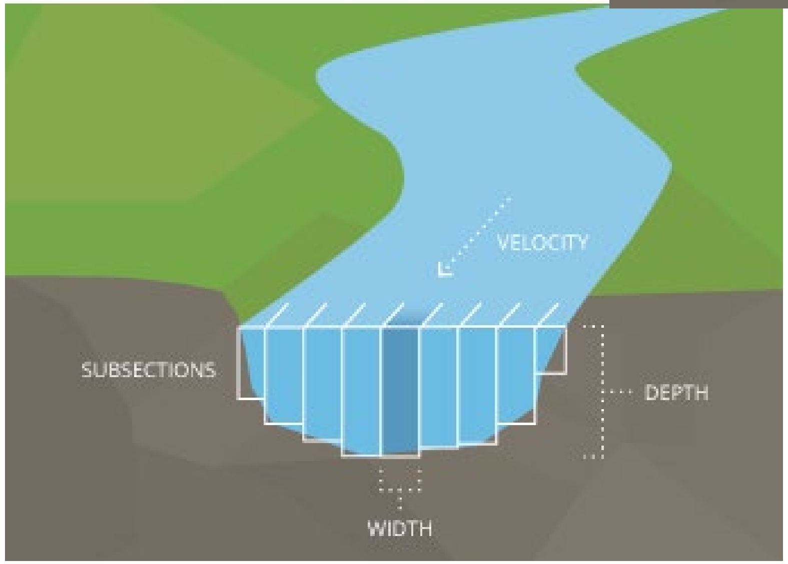
River discharge (flowrate)

- Volume of water flowing through a river per unit time
- Common units: [ft^3/sec], [m^3/sec]
- Determine cross-section geometry
- Measure depth
- Calculate flowrate



Digital depth sensor

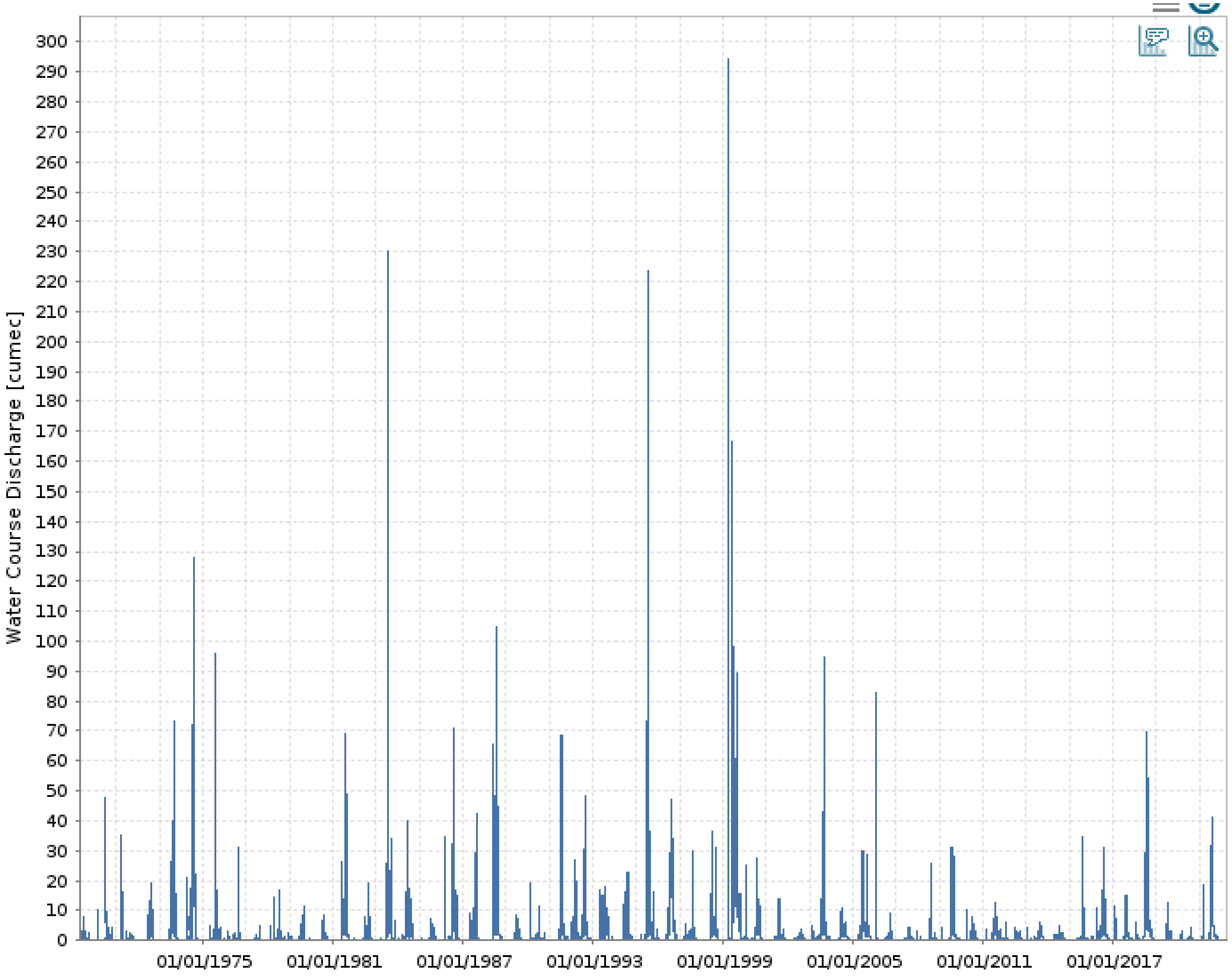
Analog staff gauge



River cross-section geometry

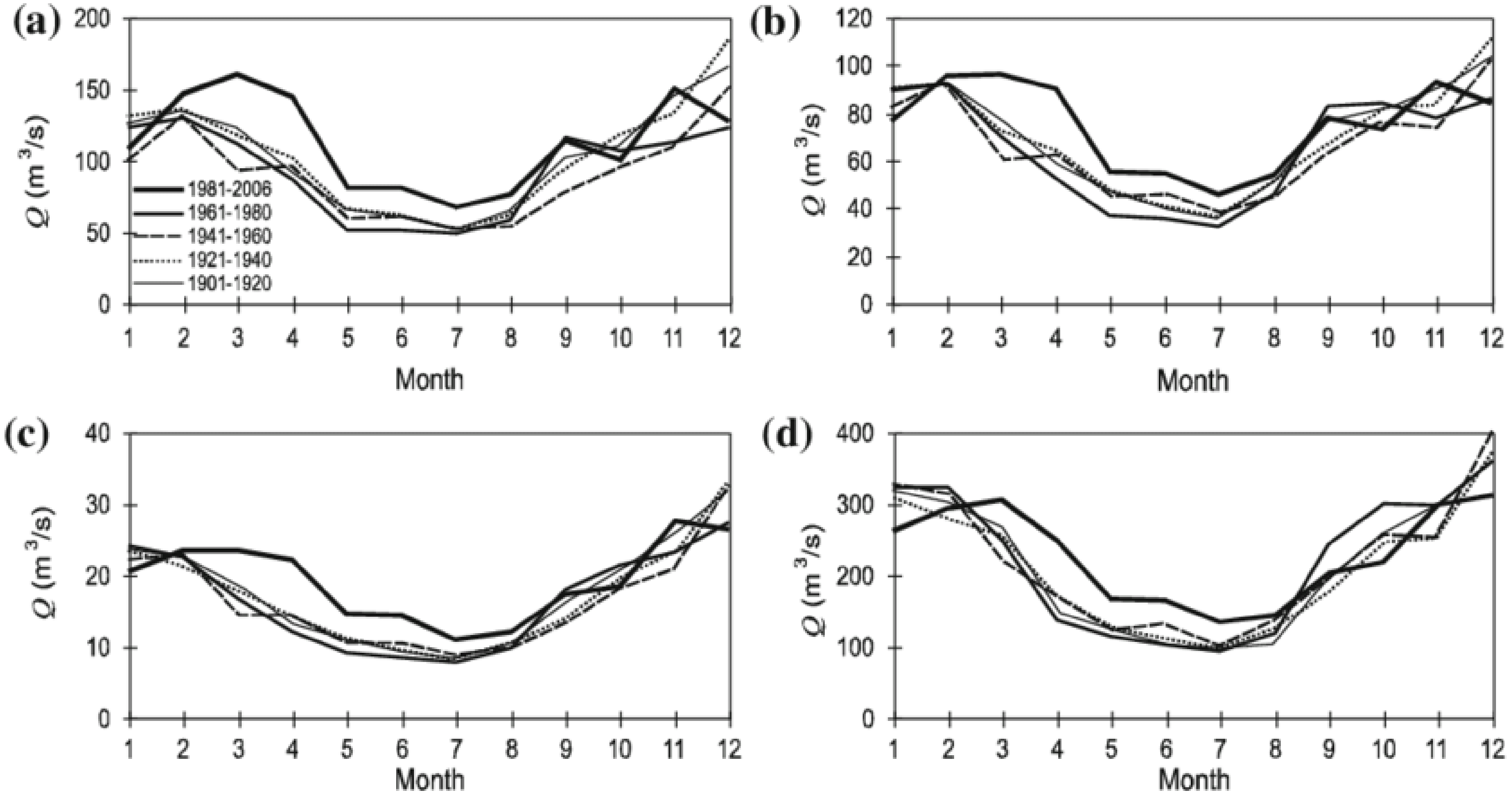
Visualizing discharge

- Discharge can be measured every hour, day, month, year
- Sample rate impacts how we understand the data
  - Daily discharge plot →
  - Not very clear



Visualizing discharge

- We know that water events (e.g., rain or snowmelt) change somewhat predictably with the seasons
- We can better understand the water cycle through a river by looking at the mean flow for each month over many years



The problem

Context

You are a hydrological engineer with experience designing a dam in Michigan. You have been asked to design a similar dam in Western Australia. To ensure that the dam will not fail, you must know the maximum amount of water that might flow through the river.

Resources

You have access to governmental data that tells you the average discharge of the river for each month over several decades (for both the USA and AUS rivers).

Question

Taking the average monthly river discharge over the given years, what is the maximum discharge rate for each river?

Understand the data

Huron River

- *Station location:* Ann Arbor, Michigan, USA
- Station number: 04174500
- *Latitude:* 42°17'13"N
- *Longitude:* 83°44'02"W
- *Data owner:* United States Geological Survey
- *Time range:* January 1949 – December 2013
- *Database:* https://dashboard.waterdata.usgs.gov
- *Köppen-Geiger climate class:*

Snow with fully humid hot summer (*Dfa*): A climate where there is at least one month colder than -3°C and precipitation is generally the same throughout the year, and summers can get very hot. This climate is usually found between 35° and 45° latitude.

Moore River

- *Station location:* Red Gully, Western Australia, Australia
- Station number: 617001
- *Latitude:* 30°58'43"S
- *Longitude:* 115°48'27.0"E
- *Data owner:* Western Australia – Department of Water and Environmental Regulation
- *Time range:* April 1969 – January 2022
- *Database:* http://www.bom.gov.au/waterdata/
- *Köppen-Geiger climate class:*

Warm temperate with dry, hot summer (*Csa*): A climate where the coldest month is warmer than -3°C but colder than +18°C and summers are dry and hot. This climate is usually found inland on western sides of continents.

Step 1: Download the data from Canvas

- All data and code are saved in `river_discharge.zip`
- Unzip the folder and place it somewhere on your computer that you will remember (do not move files into or out of the folder)
- Open the MATLAB script `river_discharge_analysis.m`
- Open the `river_discharge_worksheet.docx` file, and answer questions as you complete the lab
- Run the script, change the path new the new folder
  - Your script will throw an error. That’s okay!

Step 2: Explore the Huron River data

- In the Current Folder window in MATLAB, right click on `huron_river_monthly_discharge.dat` and hit `Open as Text`
- DAT files are generic data files that are used by many computer programs to save raw information, but the data is not labeled!
- You should see three columns of number data
- Knowing that the dataset describes the monthly discharge average, what do you think each column represents?
- Looking at the USGS website (https://waterdata.usgs.gov/nwis/uv?04174500), what is the unit of monthly discharge (assume the plot uses the same units as the given data)?

Step 3: Explore the Moore River data

- In the Current Folder window in MATLAB, right click on `moore_river_monthly_discharge.csv` and hit `Open as Text`
- The file name CSV is an initialism that stands for **C**omma **S**eparated **V**alue
- Raw data are often saved in CSV files because they are easy to extract information from and are formatted in a similar manner to the rows and columns of Excel sheets
- You should header text (lines starting with # symbols) that describes the dataset and names the columns
- After the header text, you should see lines of data separated by commas. See how the Quality Code and Interpolation Type data are encoded
- In the Current Folder window in MATLAB, right click on `moore_river_disclaimer.txt` and hit `Open`
- This is the specifications sheet for the dataset. Refer to this document to decode the CSV file information

Extract the data

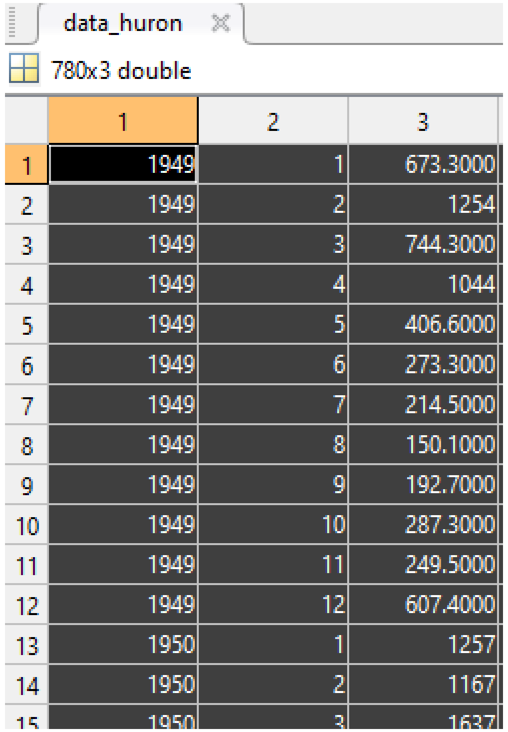
Step 4: Extract the Huron River data

- Read the documentation for the `readmatrix()` function (`doc readmatrix`)
- We want to put the Huron River data into the array called `data_huron`
- What should the input be for the `readmatrix()` function?
  - Hint: Read the documentation for the function and choose the appropriate input
- Put your input into the code

```
%% Place Huron River discharge data into arrays
% Import data from DAT file
data_huron = readmatrix(YOUR_INPUT_HERE);
```

Step 5: Ensure your data is loaded

- Place your cursor on the line that you just changed
- Click Run Section to run the code section you changed
- You should see in the Workspace a new variable called `data_huron`
- Double click this variable name in the Workspace to open it
- The Huron River data should be loaded into an array in a similar shape as the DAT file showed →



	1	2	3
1	1949	1	673.3000
2	1949	2	1254
3	1949	3	744.3000
4	1949	4	1044
5	1949	5	406.6000
6	1949	6	273.3000
7	1949	7	214.5000
8	1949	8	150.1000
9	1949	9	192.7000
10	1949	10	287.3000
11	1949	11	249.5000
12	1949	12	607.4000
13	1950	1	1257
14	1950	2	1167
15	1950	3	1637

Step 6: Separate the data into three arrays

- We want to separate the Huron river data into three arrays, one for each column
- We can do that by slicing the array using indexing
- In the script, find the section titled `Separate Huron River data into three arrays`
- Choose the column number to put into each line of code (replace `YOUR_INPUT_HERE` with the correct column number)
- Click on one of the lines you just changed and hit Run Section. You should now see three more variables in the Workspace

Workspace

Name	Value
data_huron	780x3 double
discharge_huron	780x1 double
month_huron	780x1 double
year_huron	780x1 double

```
%% Separate Huron River data into three arrays
year_huron = data_huron(:, YOUR_INPUT_HERE);
month_huron = data_huron(:, YOUR_INPUT_HERE);
discharge_huron = data_huron(:, YOUR_INPUT_HERE);
```

Step 7: Extract the Moore River data

- Since the Moore River data uses text and numbers, we cannot use `readmatrix()`
- Instead, we can use the `readtable()` function
- Read the documentation for `readtable()` (`doc readtable`)
- Choose the correct input for the `readtable()` function, put your answer in the code, and run the section
- You should see the `data_moore` variable in your Workspace
- Double-click on the `data_moore` variable to see the table you created

	1 x_Timestamp	2 Value	3 QualityCode	4 InterpolationType
1	'1969-05-01T00:00:00.000+08:00'	NaN	-1	603
2	'1969-06-01T00:00:00.000+08:00'	1.0190	10	603
3	'1969-07-01T00:00:00.000+08:00'	1.0310	10	603
4	'1969-08-01T00:00:00.000+08:00'	0.6200	10	603
5	'1969-09-01T00:00:00.000+08:00'	0.3200	10	603
6	'1969-10-01T00:00:00.000+08:00'	0.0490	10	603
7	'1969-11-01T00:00:00.000+08:00'	0.0400	10	603
8	'1969-12-01T00:00:00.000+08:00'	0.0070	10	603

```
%% Import Moore River data from CSV file
data_moore = readtable(YOUR_INPUT_HERE);
```

Step 8: Understand tables vs. arrays

- Notice that the `data_moore` table looks different from your `data_huron` array
- A table is a much more readable way to store data in MATLAB
- An array is a much more manipulatable way to store data in MATLAB
- Since the year and month data is combined in the `x_Timestamp` column, we need to convert the table to an array to extract these values
- Note the `NaN` values; these are values that are blank in the CSV file and represent not-a-number inputs, since blank table entries are not allowed

	1 x_Timestamp	2 Value	3 QualityCode	4 InterpolationType
1	'1969-05-01T00:00:00.000+08:00'	NaN	-1	603
2	'1969-06-01T00:00:00.000+08:00'	1.0190	10	603
3	'1969-07-01T00:00:00.000+08:00'	1.0310	10	603
4	'1969-08-01T00:00:00.000+08:00'	0.6200	10	603
5	'1969-09-01T00:00:00.000+08:00'	0.3200	10	603
6	'1969-10-01T00:00:00.000+08:00'	0.0490	10	603
7	'1969-11-01T00:00:00.000+08:00'	0.0400	10	603
8	'1969-12-01T00:00:00.000+08:00'	0.0070	10	603

data_huron

780x3 double

	1	2	3
1	1949	1	673.3000
2	1949	2	1254
3	1949	3	744.3000
4	1949	4	1044
5	1949	5	406.6000
6	1949	6	273.3000
7	1949	7	214.5000
8	1949	8	150.1000

Step 9: Extract the Moore River year and month data

- In the code section `Extract the data from the CSV table`, we are looping through each row of data to extract the important values
- Note that we start by converting the table `x_Timestamp` value into a cell
- We then convert the cell into a string datatype (a list of letters)
- Essentially, we have isolated the cell value for the time stamp (e.g., '1969-05-01T00:00:00.000+08:00')
- This time stamp shows the :

  Year-Month-DayTHour:Minute:Second.FractionsOfASecond+TimeZone
- We only want the year and month
- Notice that the year and month values are separated by a hyphen, we can use the `strsplit()` function to isolate the values between the hyphens

Step 9 (continued)

- Read the documentation for the `strsplit()` function (`doc strsplit`)
- Determine what inputs we need to put in the line of code that says

```
time_stamp = time_stamp_cell{1};
split_time_stamp = strsplit(YOUR_INPUT_HERE, YOUR_INPUT_HERE);
```

- Note that the `time_stamp` variable contains the string we are separating
- Type your inputs into the `strsplit()` function and run the code section
- The code has taken the year and month data you separated and placed them into arrays called `year_moore` and `month_moore`
- The river discharge rate has been extracted directly from the table and placed in the `discharge_moore` variable
- Your Workspace should look like

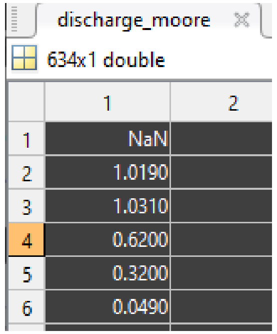
Workspace

Name	Value
data_huron	780x3 double
data_moore	634x4 table
discharge_huron	780x1 double
discharge_moore	634x1 double
month_huron	780x1 double
month_moore	634x1 double
year_huron	780x1 double
year_moore	634x1 double

Clean and convert the data

Step 10: Clean the Moore River data

- Double-click on the `discharge_moore` variable in your Workspace to see the values



- You will see that the first value is NaN (not-a-number)
- We will need to remove the bad datapoints from our array
- Place your cursor in the `Clean Moore River data` code section
- In this code, we slice out the NaN values using the `isnan()` function
  - The `~isnan()` means we are finding the indices for values that are not NaN
- Click Run Section to remove the NaN values from the array
- Double-click on the `discharge_moore` variable in your Workspace again to confirm that the NaN values are now gone

Step 11: Calculate the mean monthly discharge values

- We now have arrays that contain the discharge, year, and month values for each river
- The next step we need to do is calculate the mean monthly discharge value
- For instance, we have the discharge values for several years for the month of January; we can take a simple mean average for these values to find one mean discharge rate for the month of January, representing the amount of river discharge we should expect for that month
- We repeat this process for all twelve months (as done by the loop in the `Calculate monthly mean discharge` code section)
- Place your cursor in the `Calculate monthly mean discharge` code section and hit Run Section
- Your Workspace should now contain new arrays `monthly_mean_huron` and `monthly_mean_moore`

Step 12: Convert units

- Since people in the USA does not always use SI units, we will need to convert the discharge rates for the Huron River to SI units ([m^3/sec])
- We can use dimensional analysis to determine a *conversion factor*
- If we multiply the discharge rates for the Huron river by the conversion factor, we will get the discharge rates in cubic meters per second
- For example, converting x miles per hour to kilometers per second is done by:

$$\text{value in desired unit} = \text{conversion factor} \times \text{value in given unit}$$

$$x\left[\frac{\text{mi}}{\text{hr}}\right] \rightarrow \left[\frac{\text{km}}{\text{hr}}\right] \qquad \left(\frac{x\ \cancel{[\text{mi}]}}{1\ [\text{hr}]}\right)\left(\frac{1.609\ [\text{km}]}{1\ \cancel{[\text{mi}]}}\right) = 1.609x\left[\frac{\text{km}}{\text{hr}}\right]$$

Given value

Conversion factor

Step 12 (continued)

- Determine the appropriate conversion factor(s) for this dataset using dimensional analysis (as shown in the last slide)
- We want both discharge values to be in the same units (cubic meters per second)
- Input your conversion factor to the code section `Convert units`

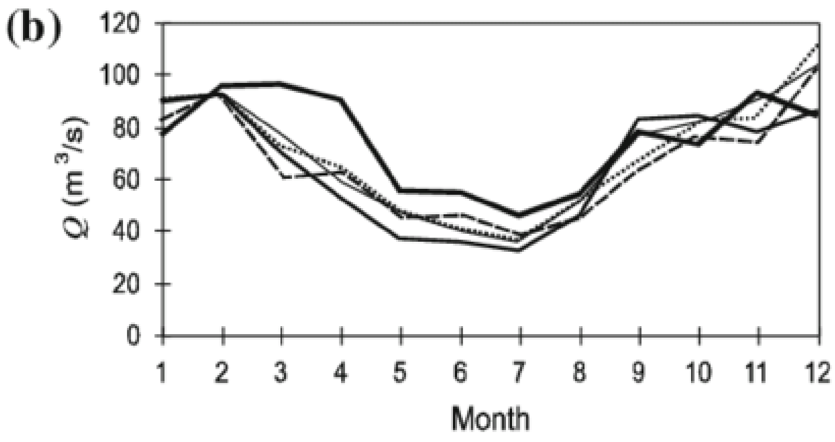
```
%% Convert units
conversion_factor = YOUR_INPUT_HERE;
monthly_mean_huron = conversion_factor * monthly_mean_huron;
```

- Here, we are recalculating the value of the mean monthly discharge values for the Huron River by multiplying the original values by the conversion factor, making the units for discharge of this river the same as that of the Moore River (cubic meters per second)
- Run this section of code

Plot your results

Step 13: Plot the mean monthly discharge rates

- We now have arrays that contain the mean monthly discharge rates for both rivers
- We want to plot the discharge for each month
  - Similar to this plot, but only one line
- We create an array of numbers from 1 to 12, representing each of the months (January = 1, February = 2, ..., December = 12)
- Fill in the code in the last code section to create your plot
- Make sure your labels tell the reader what you are plotting and what the units are (where necessary)
- Include a legend to differentiate your two lines




```
%% Plot monthly mean discharge rates for each river
month_indices = 1:1:12;
figure()
    hold on
    plot(YOUR_INPUT_HERE, YOUR_INPUT_HERE, 'LineWidth', 2)
    plot(YOUR_INPUT_HERE, YOUR_INPUT_HERE, 'LineWidth', 2)
    hold off
    xlabel(YOUR_INPUT_HERE)
    ylabel(YOUR_INPUT_HERE)
    grid on
    xlim([1,12])
    legend(YOUR_INPUT_HERE, YOUR_INPUT_HERE)
```

Step 14: Assess your results

1. What is the peak discharge for the Huron River and during which month?
2. What is the peak discharge for the Moore River and during which month?
3. Why is the peak discharge different for the two rivers
    a. Think about how the values are different and how the months in which the peaks occur are different
    b. Consider what factors go into the discharge rate and how these factors are different in Michigan vs. Western Australia (see the slides showing the aerial views of the sites for ideas)

Conclusion

Before you leave

- Complete the given worksheet (type your answers below the given questions)
- Upload your completed worksheet to Canvas (.docx or .pdf)
- Upload your MATLAB script file (`river_discharge_analysis.m`) with your name, date, and section indicated on the top of the code