Modeling Instructions for Data Analysis

Why create models?

Data visualization is important because it makes complex data easier to understand, helping to reveal trends, patterns, and outliers at a glance. It supports better decision-making by presenting insights clearly and efficiently, saving time and making communication more effective—especially for non-technical audiences. By turning raw data into visual stories, it encourages exploration, understanding, and action.
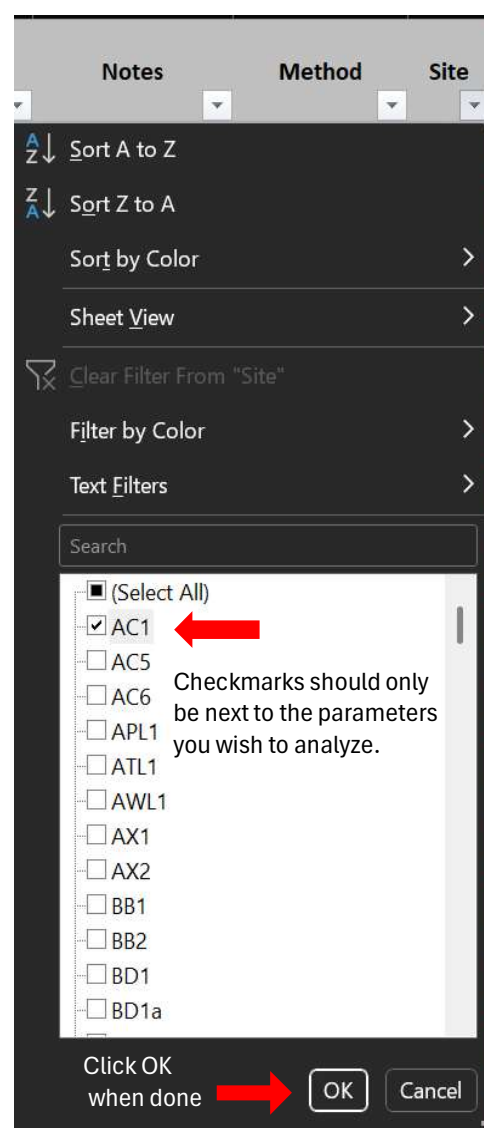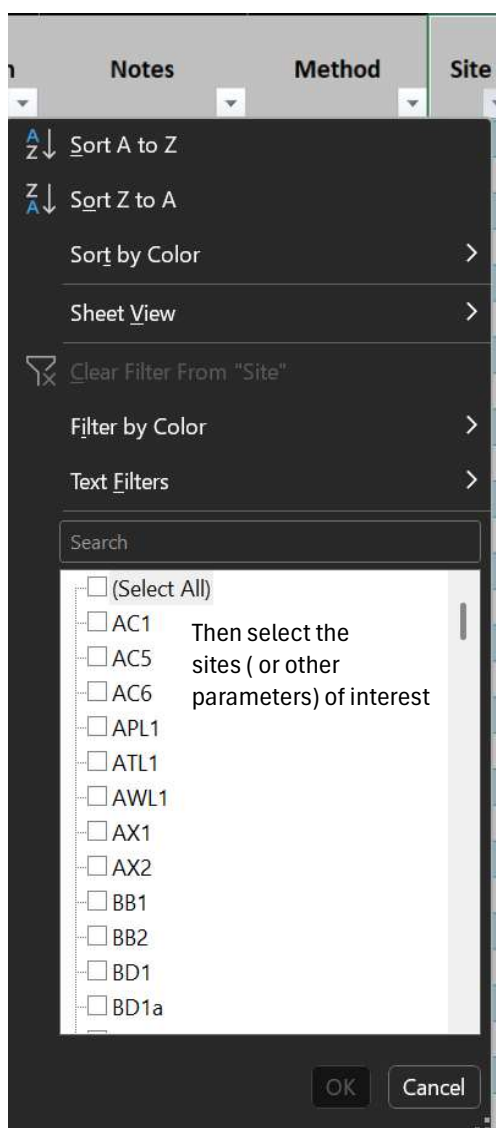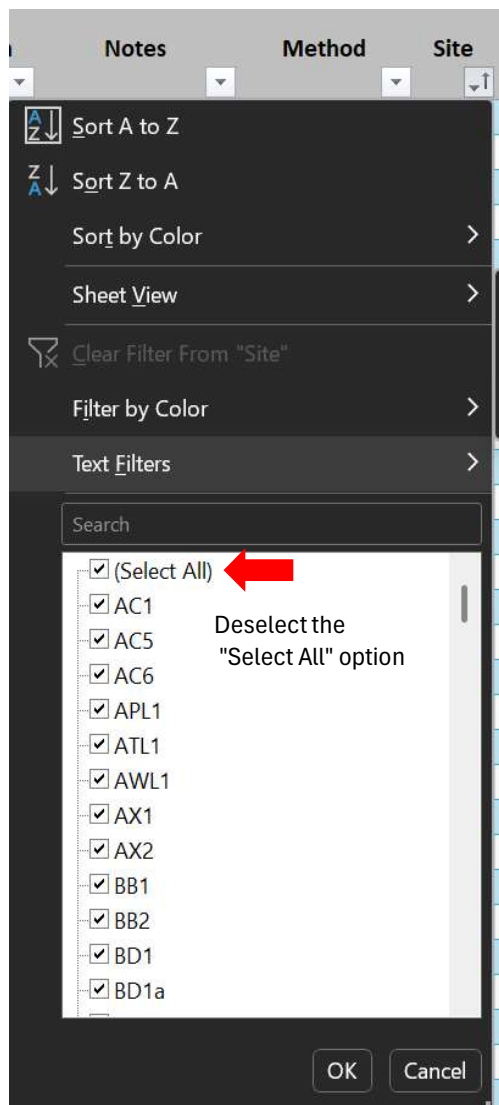
Step 1: Picking Data

First, choose which data is of interest. Analyzing all data will likely provide a larger timeline of data for predicting trends, however large masses of data are often tricky to analyze. Focusing on a smaller subset, such as the the precreated subwatershed tabs, more provide visualizations that are easier to interpret. You can also create your own subsets of data using the filtering tool.

Creating Custom Subsets

To create a custom subset that is not one of the sub watersheds, you can utilize the filtering tool in excel to select exactly what you are interested in.
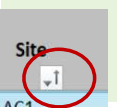
Click on the inverted triangle symbol located on each parameter at the top of the sheet. See the red circled logo above.

Notes	Method	Site

Sort A to Z

Sort Z to A

Sort by Color >

Sheet View >

Clear Filter From "Site"

Filter by Color >

Text Filters >

Search

☑ (Select All) ⬅ Deselect the
☑ AC1   "Select All" option
☑ AC5
☑ AC6
☑ APL1
☑ ATL1
☑ AWL1
☑ AX1
☑ AX2
☑ BB1
☑ BB2
☑ BD1
☑ BD1a

OK   Cancel

Notes	Method	Site

Sort A to Z

Sort Z to A

Sort by Color >

Sheet View >

Clear Filter From "Site"

Filter by Color >

Text Filters >

Search

☐ (Select All) Then select the
☐ AC1   sites (or other
☐ AC5   parameters) of interest
☐ AC6
☐ APL1
☐ ATL1
☐ AWL1
☐ AX1
☐ AX2
☐ BB1
☐ BB2
☐ BD1
☐ BD1a

OK   Cancel

Notes	Method	Site

Sort A to Z

Sort Z to A

Sort by Color >

Sheet View >

Clear Filter From "Site"

Filter by Color >

Text Filters >

Search

◼ (Select All)
☑ AC1   ⬅
☐ AC5   Checkmarks should only
☐ AC6   be next to the parameters
☐ APL1   you wish to analyze.
☐ ATL1
☐ AWL1
☐ AX1
☐ AX2
☐ BB1
☐ BB2
☐ BD1
☐ BD1a

Click OK
when done ⬅ OK   Cancel

Once you have selected the parameters of interest and have clicked okay, the excel sheet should update. Now, only entries with the parameters you selected should be visible on the sheet. See an example below:

Data Condition	Notes	Method	Site	Date	Time	Air
Provisional		LaMotte	AC1	Saturday, November 21, 2020	1005	
Provisional		LaMotte	AC1	Saturday, December 19, 2020	1000	
Provisional		LaMotte	AC1	Saturday, January 16, 2021	1000	
Provisional		LaMotte	AC1	Saturday, March 20, 2021	1000	
Provisional		LaMotte	AC1	Saturday, April 17, 2021	1000	
Provisional		LaMotte	AC1	Sunday, May 16, 2021	1000	
Provisional		LaMotte	AC1	Saturday, June 19, 2021	1000	
Provisional		LaMotte	AC1	Saturday, July 17, 2021	1000	
Unchecked		BACT	AC1	Sunday, August 1, 2021		
Unchecked		BACT	AC1	Sunday, August 8, 2021		
Unchecked		BACT	AC1	Sunday, August 15, 2021		
Unchecked		BACT	AC1	Sunday, August 22, 2021		
Unchecked		BACT	AC1	Sunday, August 29, 2021		
Provisional		Hanna	AC1	Thursday, April 21, 2022	1159	
Unchecked		BACT	AC1	Sunday, May 22, 2022		
Unchecked		BACT	AC1	Sunday, May 29, 2022		
Unchecked		BACT	AC1	Sunday, June 5, 2022		
Unchecked		BACT	AC1	Sunday, June 12, 2022		
Unchecked		BACT	AC1	Sunday, June 19, 2022		
Unchecked		BACT	AC1	Sunday, June 26, 2022		
Unchecked		BACT	AC1	Sunday, July 31, 2022		
Unchecked		BACT	AC1	Sunday, August 7, 2022		

Identify your _____ vs _____

When analyzing a large dataset, choose two characteristics (or variables) that are relevant to your research question or objective. Look for traits that may have a logical or potential relationship—such as cause and effect, correlation, or contrast. It's helpful to begin with exploratory data analysis to identify patterns, trends, or variability across different variables. Consider the type of data (e.g., numerical vs. categorical) and how easy it is to visualize or interpret. Ultimately, selecting meaningful and insightful variables improves the clarity and impact of your analysis.

Step 2: Modeling

Modeling can be done inside excel or on alternate data visualization platforms .

Preparing Data

To prepare data for modeling, first select your parameters of interest. In this example, we will select one site (AC1), the date, and the water temperature. To select just one site, follow the instructions located in the "Creating Custom Subsets" section. Filter to just the desired site(s).

Then, create a new sheet for your custom subset. This can be done by clicking the plus sign (+) icon located at the bottom right of the excel window. The icon should look like this:

Once in the new sheet, you will need to copy over the data of interest. Again, make sure only sites you wish to analyze are selected. Select all the data of the parameter of interest by clicking the column header (usually represented by a capital letter) and pressing the control key and the C key if on a windows device, or command key and the C key if on an iOS device.

By clicking the H column header, which represents Water Temperature, I have selected the entirirety of the Water tempertaure data. Now, I can copy this data and paste into the new sheet I created earlier.


Once this data is copied, paste into the first avilable column in the new sheet you created. Remember, you can navigate between sheets using the tabs located at the bottom of the excel window. Paste by selecting the column header and pressing Ctrl C or Command C. Repeat for any parameters of interest.

Cleaning data

To clean the data, you will need to add filers. As described above, use the column headers to select all the parameters you have added to you new sheet. In this example, they are date and water temperature. Once all parameters are selected, navigate to the "Sort & Filter" dropdown, and click "Filter". This should allow you to select and deselect specific data entries. Now, evaluate each parameter indivudually. Click the filtering drop down (For more information on how to do this, refer to "Creating Custom Subsets"), and scroll until you find "(Blanks)". Deselect this.
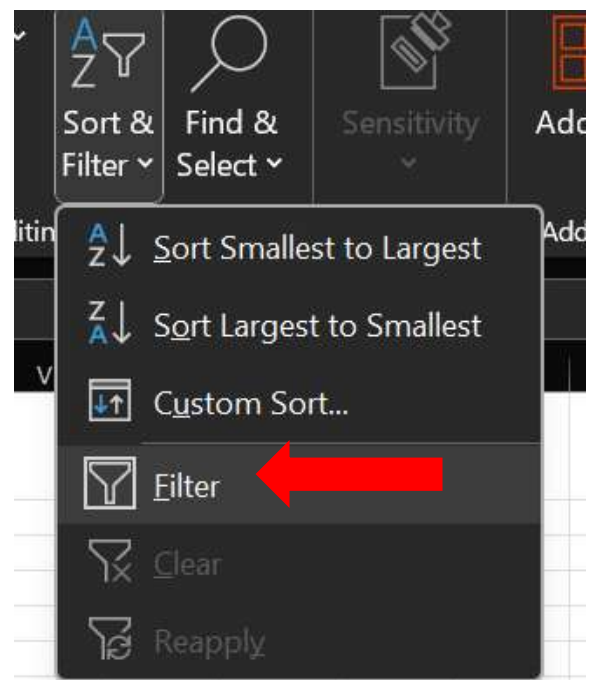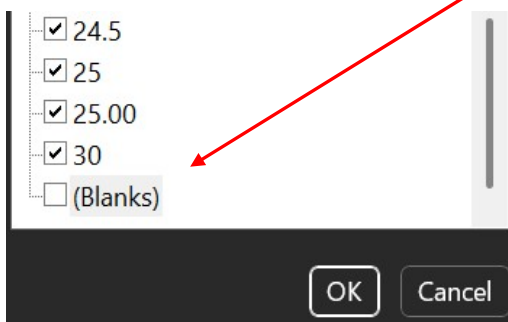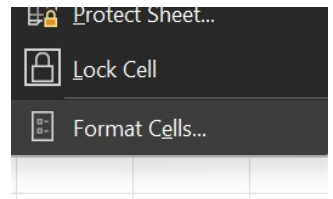
Formatting Dates

Dates in this sheet come automatically formatted in Day of the Week, Month Day, Year format. For easier analysis, we want to change the date to Month / Day / Year format. To do this, select the year column by clicking on the header. Then navigate to the formatting tab

Scroll down to the bottom of the dropdown and click " Format Cells"

Protect Sheet...

Lock Cell

Format Cells...

Format Cells   ?   ✕

| Number | Alignment | Font | Border | Fill | Protection |

Category:

General
Number
Currency
Accounting
Date
Time
Percentage
Fraction
Scientific
Text
Special
Custom

Sample
Date

Type:

*3/14/2012
*Wednesday, March 14, 2012
2012-03-14
3/14
3/14/12
03/14/12
14-Mar

Locale (location):

English (United States)

Under "Category", ensure that "Date" is selected. Then, to the right and listed under "Type", select 3/14/2012. Once these conditions are satisfied, click "OK" at the bottom of the popup. Now, the date should be correctly formatted.
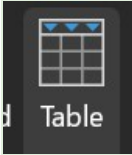
Date formats display date and time serial numbers as date values. Date formats that begin with an

Converting to a Table

Ensure all your data is selected by clicking each column header in this case A and B) while shift.

	Date	
	A	B
	Date	**Water Temperature**
	11/21/2020	9.00
	12/19/2020	2.50
	1/16/2021	6.00
	3/20/2021	5.00

Table

Insert

☑ My table has headers

OK   Cancel

Date	Water Temperature
11/21/2020	9.00
12/19/2020	2.50
1/16/2021	6.00
3/20/2021	5.00
4/17/2021	10.00
5/16/2021	15.50
6/19/2021	21.00
7/17/2021	25.00
8/1/2021	
8/8/2021	19
8/15/2021	21
8/22/2021	20
8/29/2021	21
4/21/2022	9.61
5/22/2022	20

Your data should now appear with this blue and white repeted pattern.

Modeling

Highlight only the data entries of your columns, do not include the headers.
Navigate to the "Insert" section of excel.

Insert

	A	B	C
1	Date	Water Temperat	
2	11/21/2020	9.00	
3	12/19/2020	2.50	
4	1/16/2021	6.00	
5	3/20/2021	5.00	
6	4/17/2021	10.00	
7	5/16/2021	15.50	
8	6/19/2021	21.00	
9	7/17/2021	25.00	
11	8/8/2021	19	
12	8/15/2021	21	
13	8/22/2021	20	
14	8/29/2021	21	

Select what type of chart you wish to visuaize your data with. If you are having trouble deciding, either refer to the guide below or click "Recommended Charts", which will provide some charts that may be a good fit based upon the trends and format of your data.
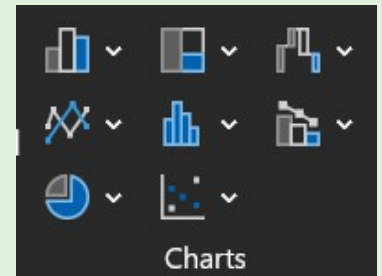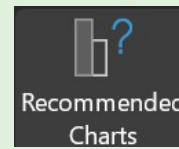
Common Chart Strengths and Weaknesses

Bar Chart
Strengths:
Great for comparing discrete categories.
Easy to read and interpret.
Works well for both small and large datasets.

Weaknesses:
Can become cluttered with too many bars.
Not suitable for showing trends over time.

2-D Column

2-D Bar

3-D Column

3-D Bar

Line Graph
Strengths:
Ideal for showing trends over time.
Good for continuous data.
Easy to compare multiple lines.

Weaknesses:
Hard to read with too many lines.
Not ideal for categorical data.

2-D Line

2-D Area

3-D Line

3-D Area
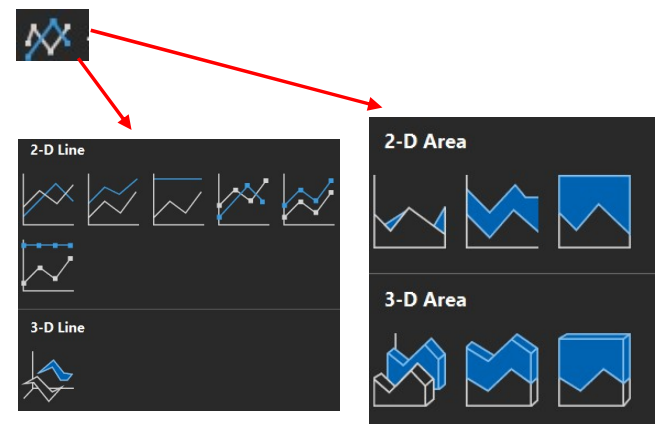
Pie Chart
Strengths:
Visually shows part-to-whole relationships.
Simple and appealing when limited to a few categories.

Weaknesses:
Hard to compare slices.
Ineffective with many or similarly sized categories.
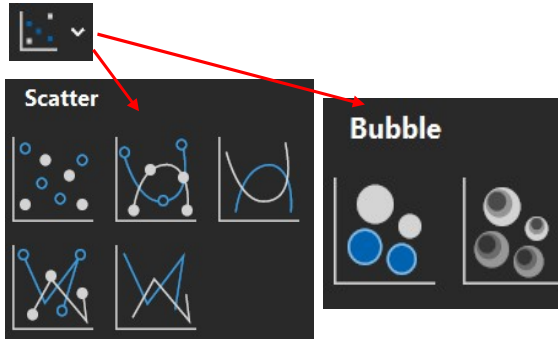Can be misleading without percentages or labels.

2-D Pie

Doughnut

3-D Pie

Scatter Plot
Strengths:
Shows relationships/correlations between two variables.
Highlights outliers and data patterns.

Weaknesses:
Not useful for large datasets without transparency.
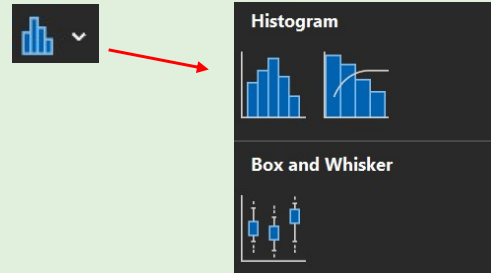Can be hard to interpret if points overlap too much.

Histogram
Strengths:
Shows the distribution of a dataset.
Great for visualizing frequency and spread.

Weaknesses:
Only works with numerical, continuous data.
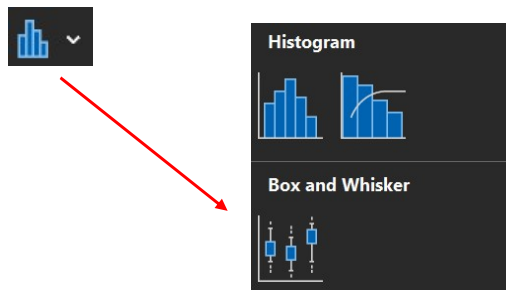Choice of bin size can affect interpretation.

Box Plot (Box-and-Whisker)
Strengths:
Summarizes data with median, quartiles, and outliers.
Great for comparing distributions across groups.

Weaknesses:
Not intuitive for general audiences.
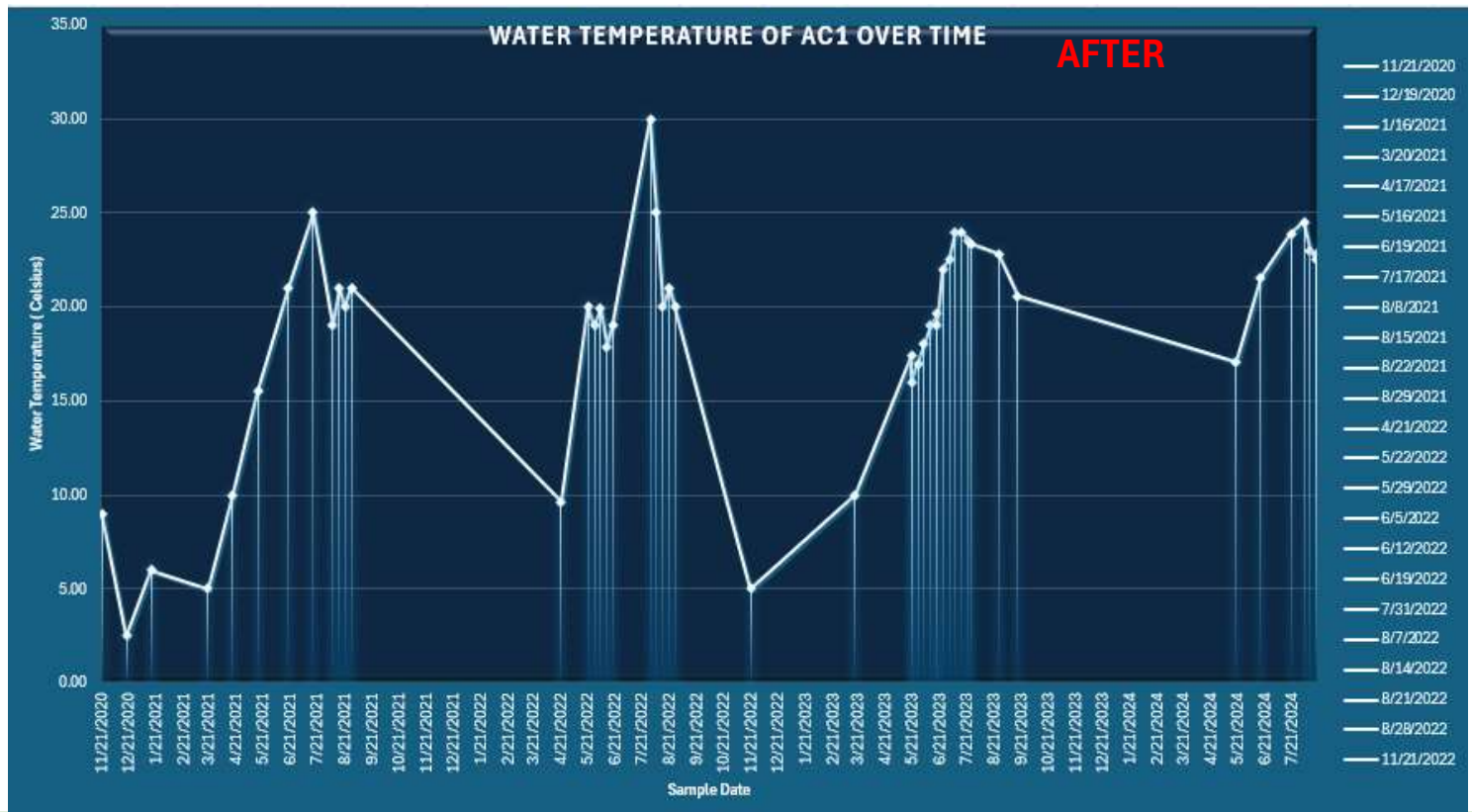Doesn't show the actual shape of the distribution.

Customize

Excel will generate a graph for you that is plain, and possibly unintuitive for analysis. Utilize excel customization features to make the graph easier to understand and better visually.
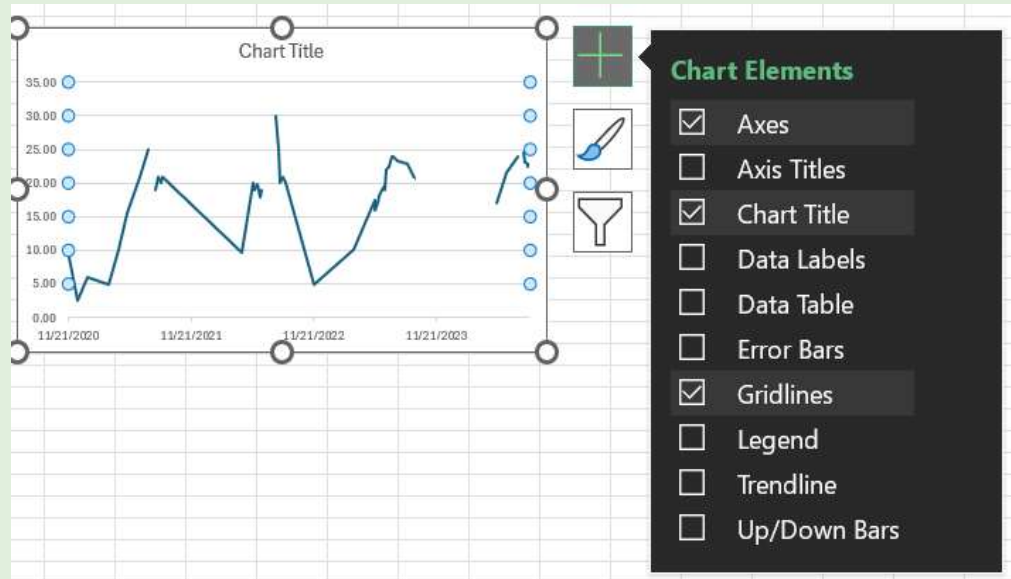

BEFORE

WATER TEMPERATURE OF AC1 OVER TIME

AFTER
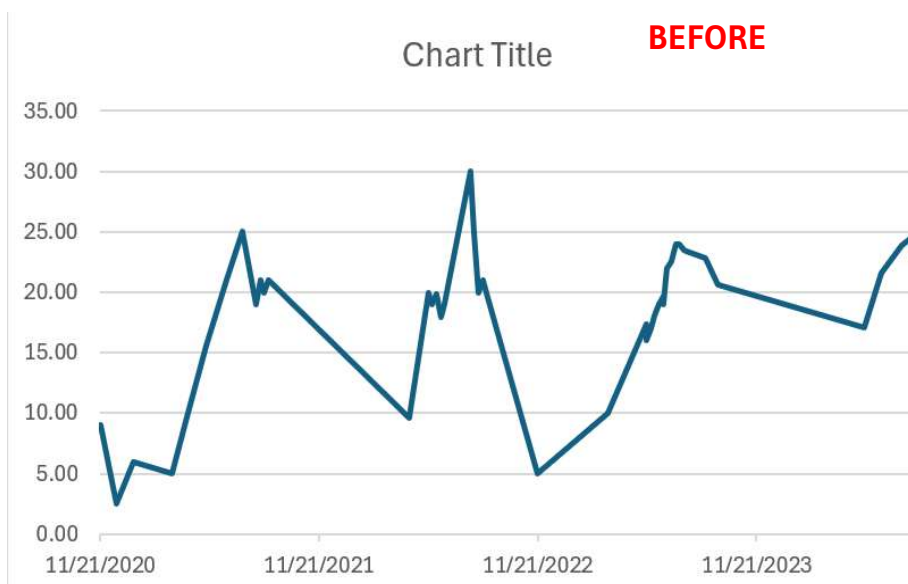
All of this can be done through the three icons that appear next to your chart. Select what elemtents and design features you wish for your graph to display. Additional features can be modified through the "Chart Design" tab

Chart Design

Chart Title

Chart Elements

- ☑ Axes
- ☐ Axis Titles
- ☑ Chart Title
- ☐ Data Labels
- ☐ Data Table
- ☐ Error Bars
- ☑ Gridlines
- ☐ Legend
- ☐ Trendline
- ☐ Up/Down Bars

Look at the Overall Shape or Pattern

Is there a trend (e.g., upward or downward)?

Are there peaks, valleys, or clusters?

For pie or bar charts, which categories are largest or smallest?

Focus on Key Data Points

Look for highs and lows, outliers, or sharp changes.

In a line graph, note when trends start, shift, or flatten.

In a scatter plot, check for correlation or groupings.

Make Comparisons

Compare values between categories or time points.

Ask: "Which is bigger?" "How much did it change?" or "Which group performs better?"

Consider the Scale and Units

A misleading scale can exaggerate or minimize differences.

Always check the y-axis and x-axis intervals and starting points.

Draw Conclusions

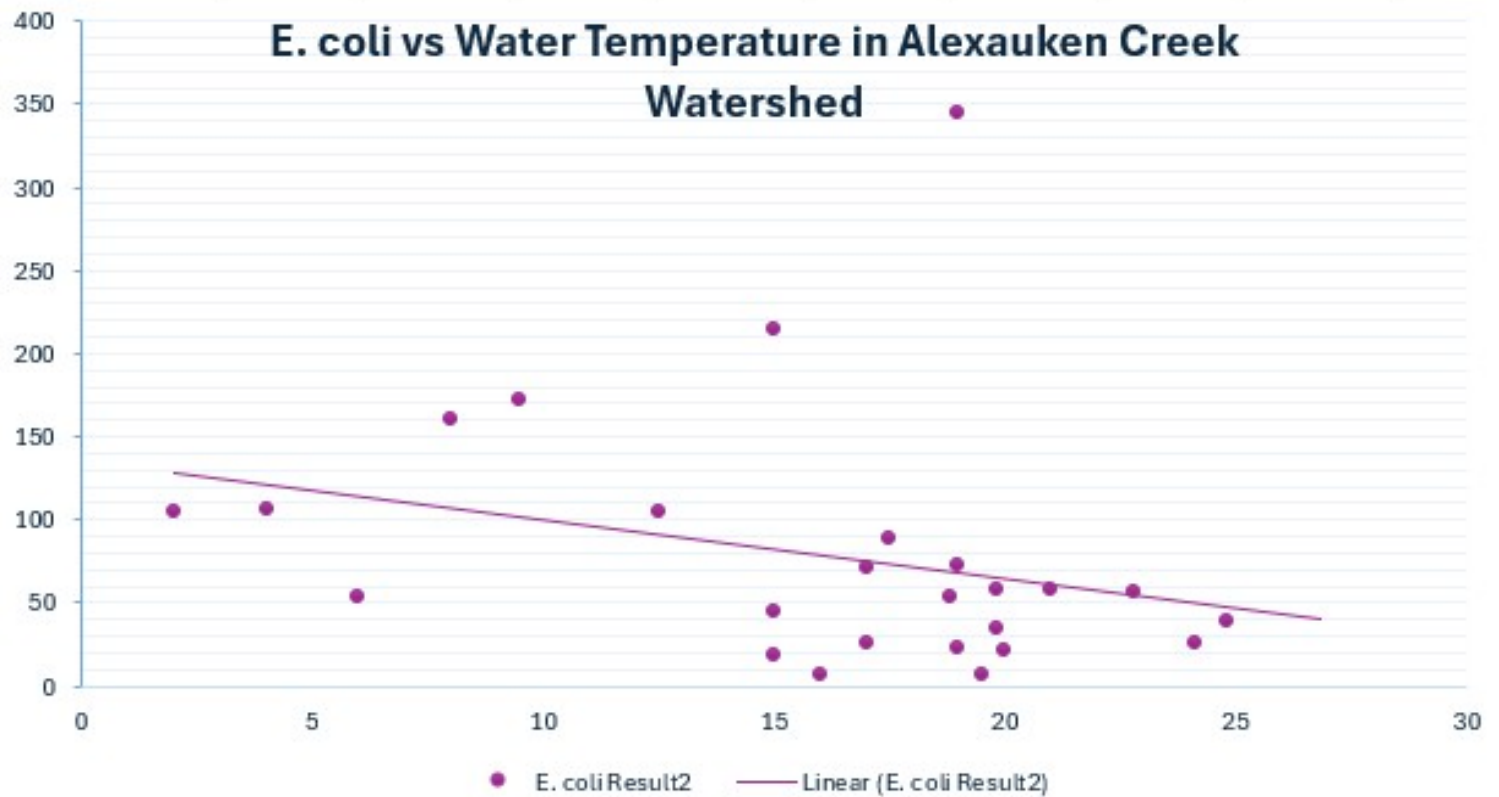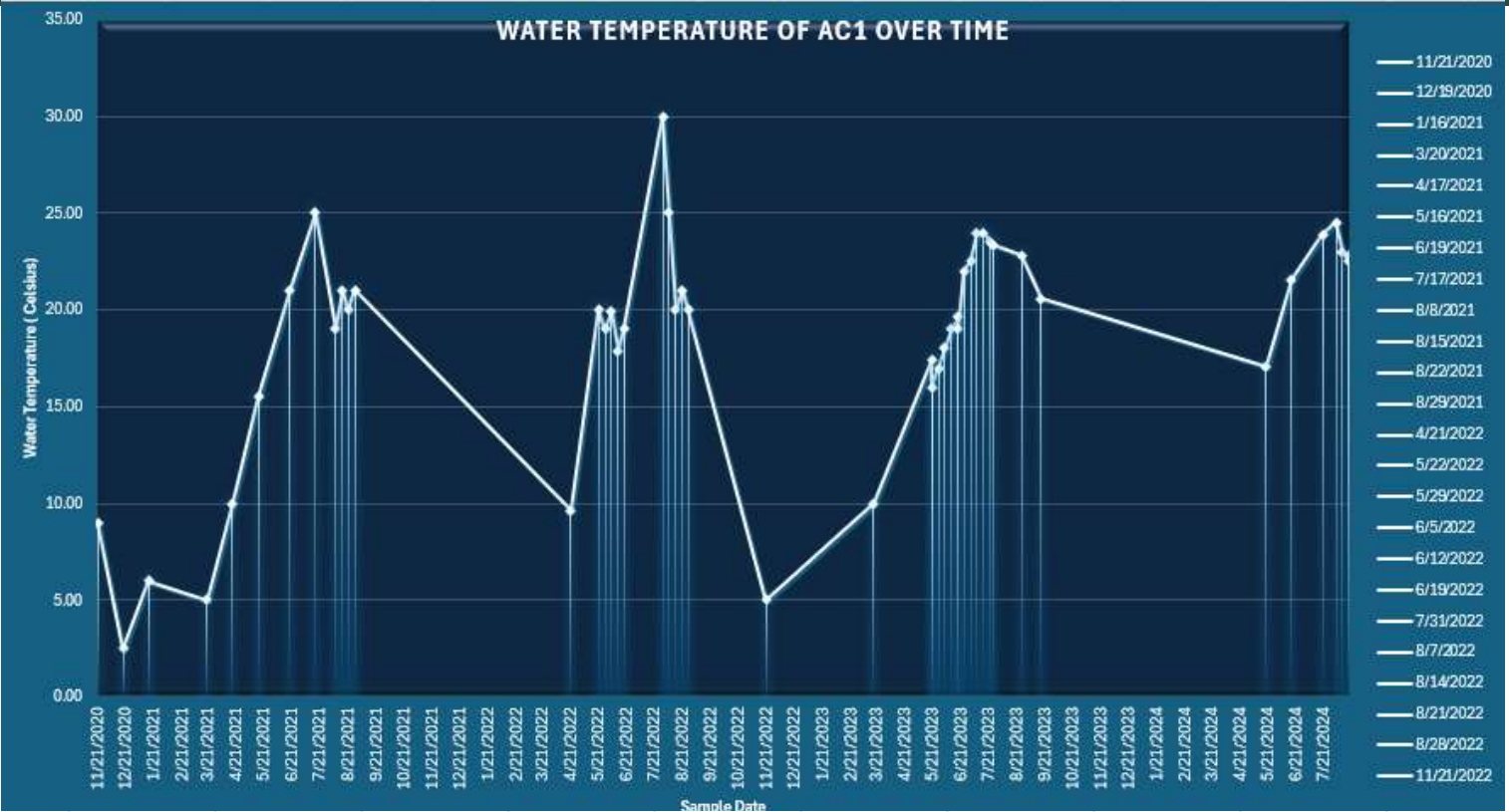Based on what you see, summarize the main insight.

Try to answer: "What is the chart telling me?" or "What decision could this inform?"
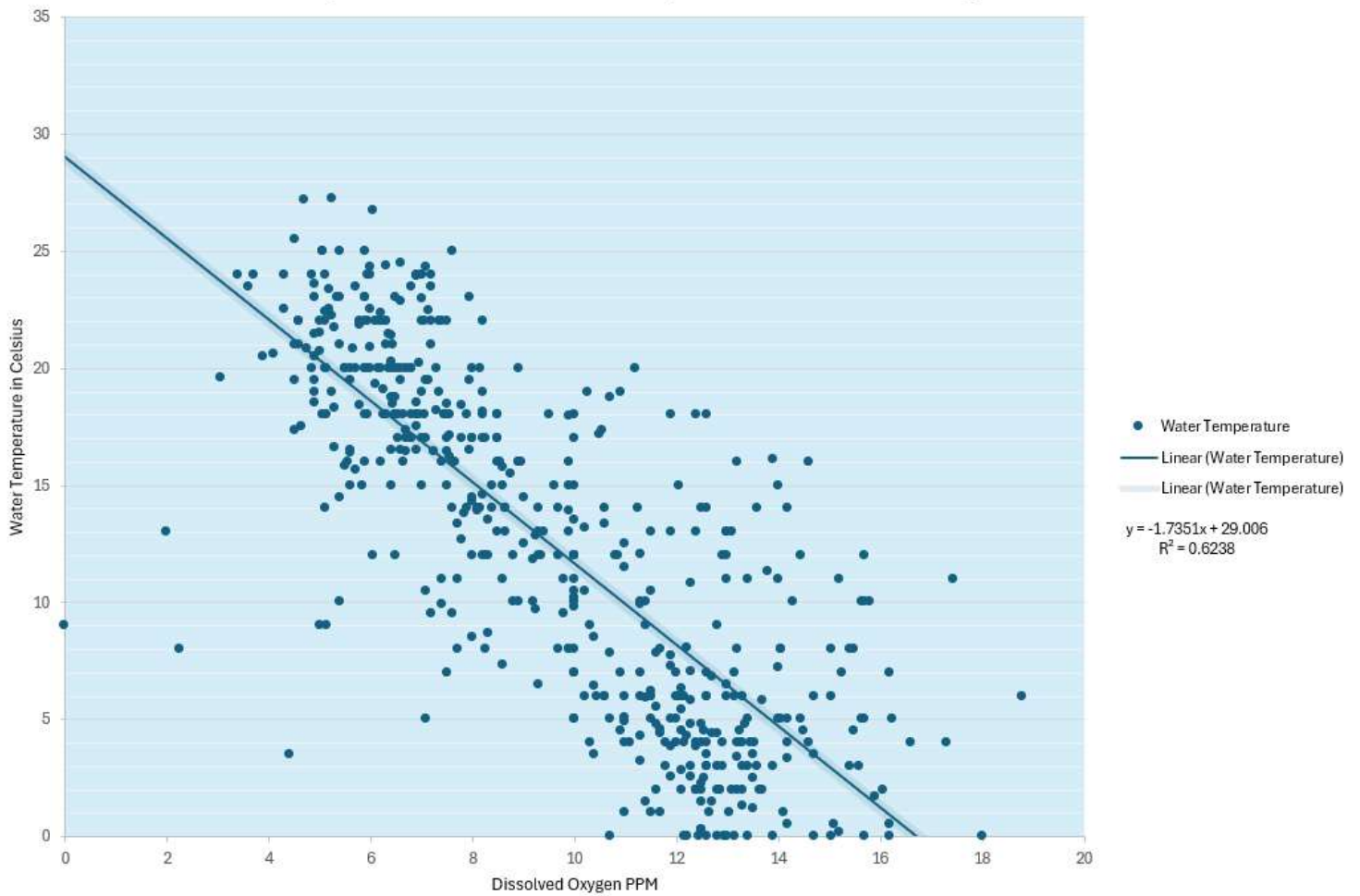
Think Critically

Is there missing context or possible bias in how the data is displayed?

Would a different type of chart tell the story better?

WATER TEMPERATURE OF AC1 OVER TIME

E. coli vs Water Temperature in Alexauken Creek Watershed

StonyBrook Watershed Water Temperature vs Dissolved Oxygen PPM

$y = -1.7351x + 29.006$
$R^2 = 0.6238$

TEMPERATURE VS NITRATE IN ASSUNPINK WATERSHED