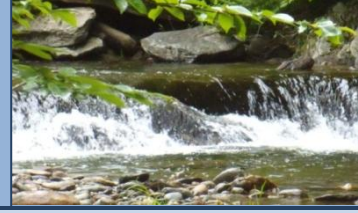
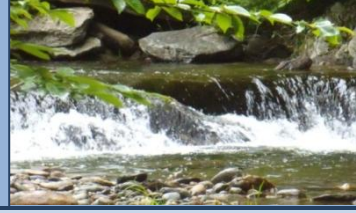
Data Analysis Tutorial



Module 2: Understanding Streams Project Data

Module 2



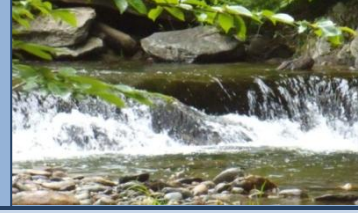
Understanding Streams Project Data

"The whole of science is nothing more than a refinement of everyday thinking" – Albert Einstein, Physics and Reality, 1936

This module helps you understand:

1) **How Streams Project data are collected**
2) **What Streams Project data are available**
3) **How to use the available data to specify your question**
4) **Why you need a hypothesis**
5) **How to finalize your question and hypothesis**

Module 2



How data are collected

You've made valuable contributions to the Streams Project data set throughout the past summer and school year! Check out some of the other people who have contributed data to the Streams Project:

Participants

High Schools
collect water and macroinvertebrate samples in streams near their schools

UVM
Undergraduates carry out independent projects

Baccalaureate College Faculty & Undergraduates
Complementary research projects

Module 2

How data are collected: A diagram of Streams Project data flow

Participants collect water and macroinvertebrate samples and make habitat observations in the field

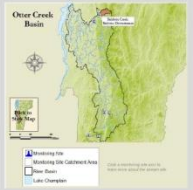
Water samples are sent to the Streams Project analytical lab at UVM

Undergraduates analyze the samples

Bacteria

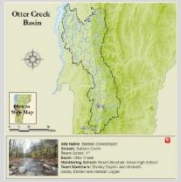
Total Phosphorus

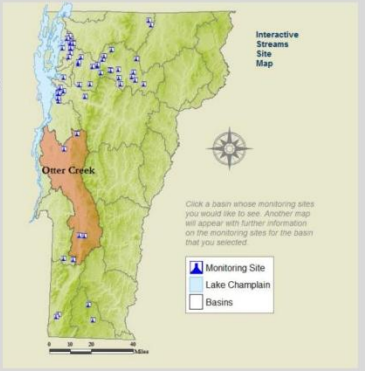
Total Suspended Solids



Undergraduates work with the Project's GIS specialist to analyze land use data and patterns. The GIS specialist also ensures that land use data are user-friendly.

High school groups monitor sites in two watersheds with contrasting land uses (e.g., agricultural, urban, forested).

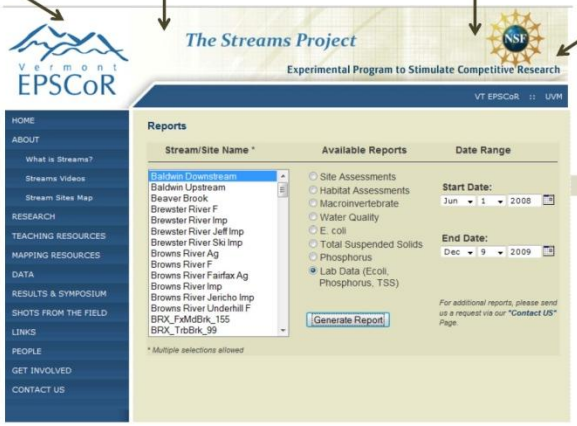




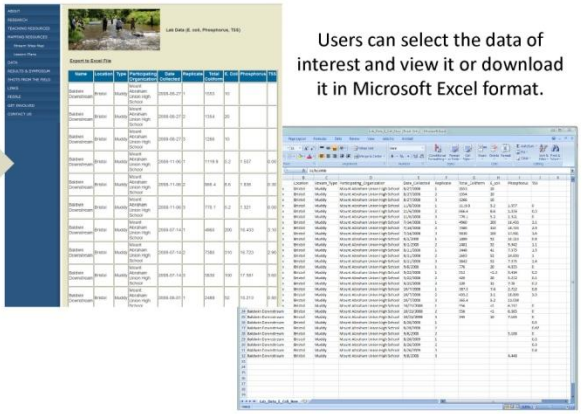
Students enter field data online

High school groups send a sub-sample of macroinvertebrates to St. Michael's College

Stream-specific, web-based macroinvertebrate identification keys are created by undergraduate interns under the direction of a science adviser. High school groups use the keys to guide them as they identify benthic macroinvertebrates.



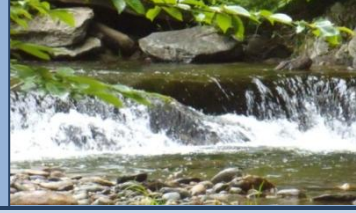
All participants contribute the data they collect to an online publicly-accessible database. Students and other researchers retrieve data from the database and use it to complete their individual research projects.



Users can select the data of interest and view it or download it in Microsoft Excel format.

Contribution to a common database provides students with the option to expand the scope of their analyses and produce more powerful results than any one team could have produced alone.

Module 2

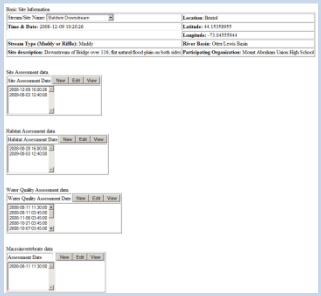


How Streams Project data are collected

The following data are available in the database:

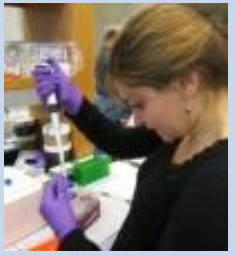
- **Site Assessment Data**
- **Habitat Assessment Data**
- **Water Quality Assessment Data**
- **Macroinvertebrate Data**

Gathered by our high school and undergraduate researchers in the field and entered into the database through this online interface:



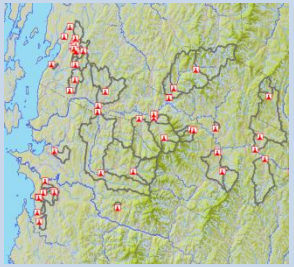
- **E.coli Lab Data**
- **TSS Lab Data**
- **Phosphorus Lab Data**

Analyzed by our lab technician and undergraduate interns at our UVM water quality lab and entered directly into the database by our database specialist.

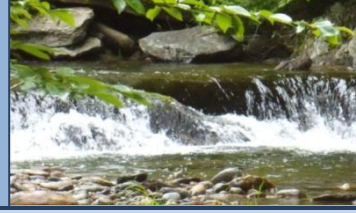


- **GIS Assessment Data**

Analyzed by our GIS specialist and undergraduate interns at UVM and entered directly into the database by our database specialist.



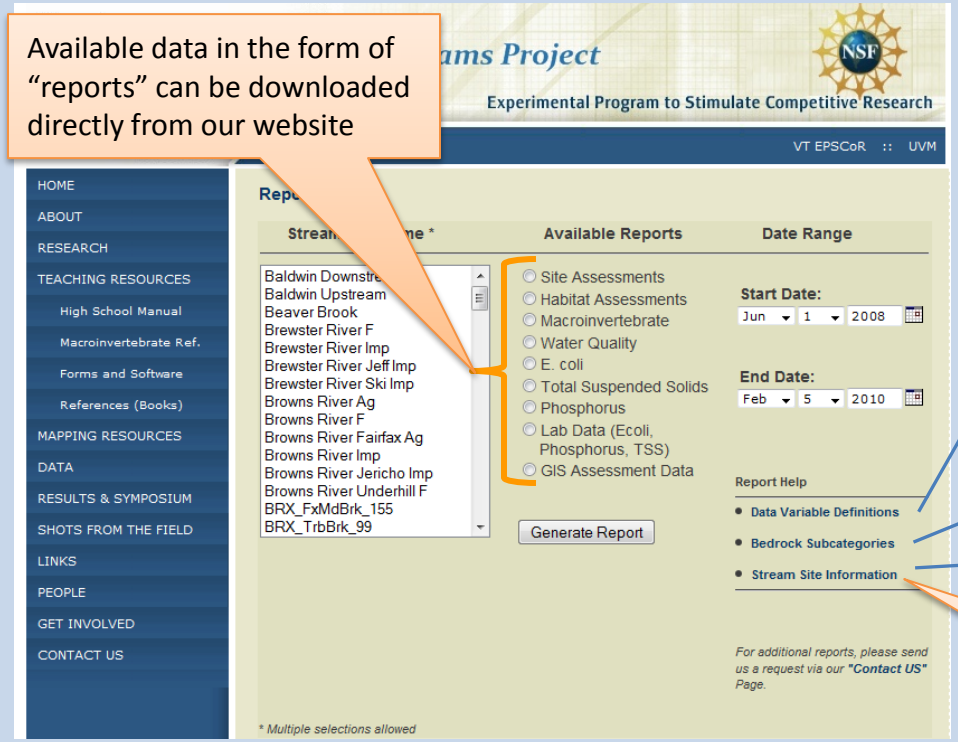
Module 2

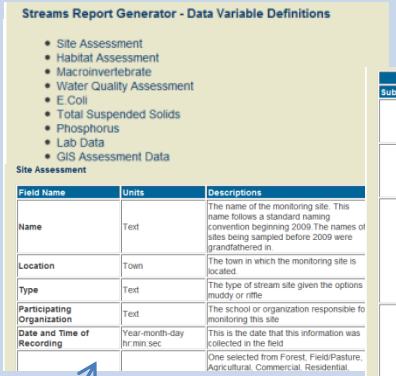


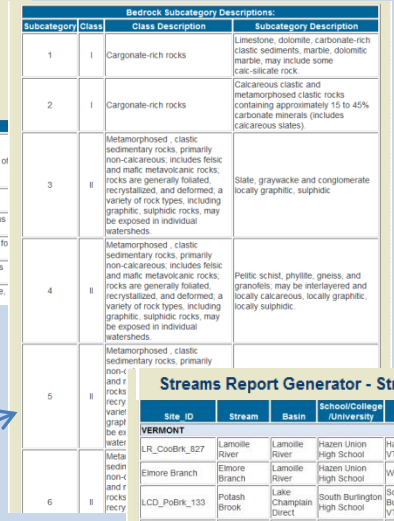
The Streams Project Database

Below is the web interface through which you will download your data. All the data sources mentioned on the previous page are listed as “Available Reports”:

Available data in the form of “reports” can be downloaded directly from our website



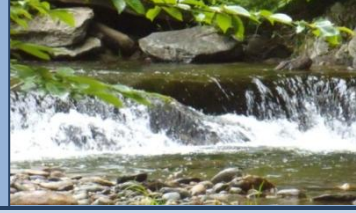




These three “Help” pages give detailed definitions of each variable. They can help you understand what each variable means and are a valuable initial resource.

We will show you how to retrieve data in the next module.

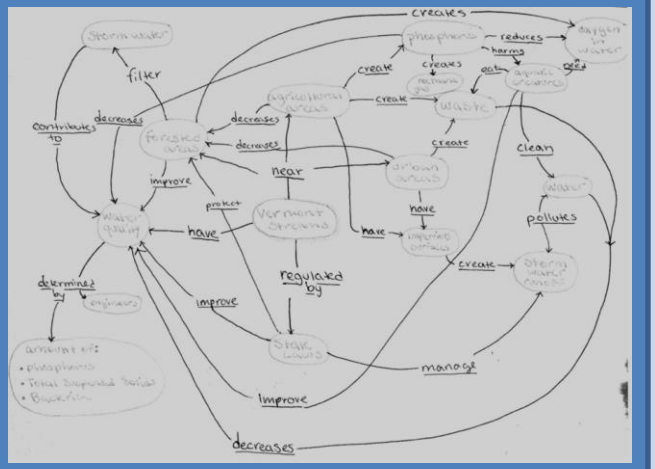
Module 2



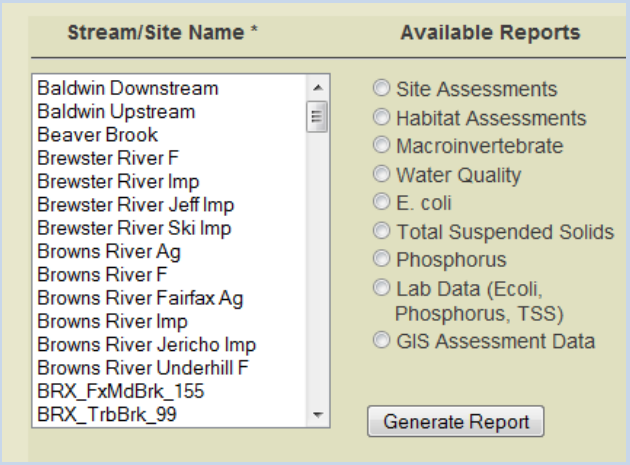
How to use the available data to specify your question

Let's learn how to turn those observations into ***answerable*** scientific questions!

Recall your concept map or web of ideas that you made in Module 1. It should look something like this:

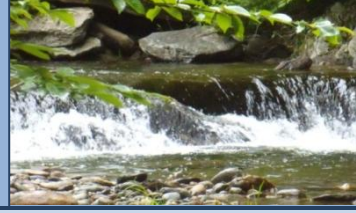


1. Look at each of the related ideas on your concept map/web.
2. Evaluate which ideas on the map have data in the database.
3. Cross out the ideas for which no data are available.



For example, if you're interested in how fish abundance is related to temperature, cross that idea out since the Streams Project doesn't collect data on fishes. However, before you write that idea off completely, ask yourself if there is some other type of data that we DO have that could substitute for fishes, like macroinvertebrates. Re-write the relationship using macroinvertebrates in place of fish!

Module 2

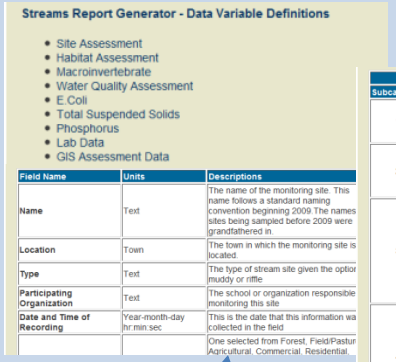


How to use available data to specify your question

At this point, you probably have at least 2 or 3 ideas about what differences or relationships you want to investigate.

Specify the following for each idea from your concept map:

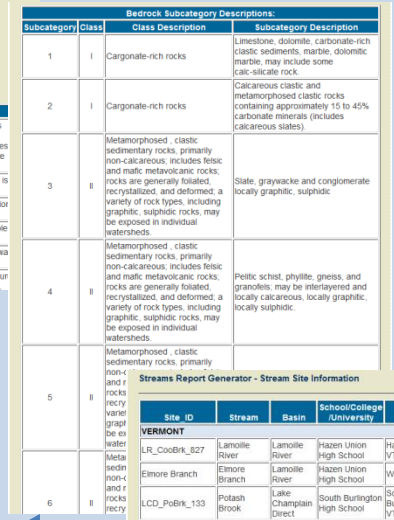
1. **The time range.** Do you want to investigate:
   - across multiple years?
   - different months in the same year?
   - the same month in different years?
2. **The spatial extent.** Do you want to investigate:
   - different watersheds?
   - different catchments?
   - streams with different surrounding land uses?
3. **The parameters of stream health.** Do you want to investigate:
   - nutrients, like phosphorus?
   - total suspended solids?
   - macroinvertebrates?
   - discharge?
   - temperature?
   - other parameters we measure?



Streams Report Generator - Data Variable Definitions

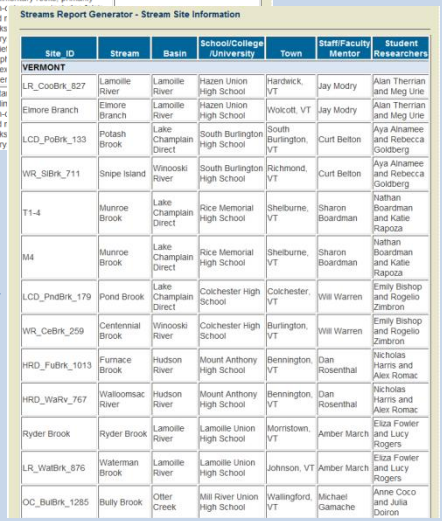
- Site Assessment
- Habitat Assessment
- Macroinvertebrate
- Water Quality Assessment
- E.Coli
- Total Suspended Solids
- Phosphorus
- Lab Data
- GIS Assessment Data

Field Name	Units	Descriptions
Name	Text	The name of the monitoring site. This name follows a standard naming convention beginning 2009.The names sites being sampled before 2009 were grandfathered in
Location	Town	The town in which the monitoring site is located.
Type	Text	The type of stream site given the optior muddy or riffle
Participating Organization	Text	The school or organization responsible monitoring this site
Date and Time of Recording	Year-month-day hr:min:sec	This is the date that this information wa collected in the field
		One selected from Forest, Field/Pastur Agricultural, Commercial, Residential,



Bedrock Subcategory Descriptions:

Subcategory	Class	Class Description	Subcategory Description
1	I	Cargonate-rich rocks	Limestone, dolomite, carbonate-rich clastic sediments, marble, dolomitic marble, may include some calc-silicate rock.
2	I	Cargonate-rich rocks	Calcareous clastic and metamorphosed clastic rocks containing approximately 15 to 45% carbonate minerals (includes calcareous slates).
3	II	Metamorphosed , clastic sedimentary rocks, primarily non-calcareous; includes felsic and mafic metavolcanic rocks; rocks are generally foliated, recrystallized, and deformed; a variety of rock types, including graphitic, sulphidic rocks, may be exposed in individual watersheds.	Slate, graywacke and conglomerate locally graphitic, sulphidic
4	II	Metamorphosed , clastic sedimentary rocks, primarily non-calcareous; includes felsic and mafic metavolcanic rocks; rocks are generally foliated, recrystallized, and deformed; a variety of rock types, including graphitic, sulphidic rocks, may be exposed in individual watersheds.	Pelitic schist, phyllite, gneiss, and granofels; may be interlayered and locally calcareous, locally graphitic, locally sulphidic.
5	II	Metamorphosed , clastic sedimentary rocks, primarily non-c and r rocks recry varie graph be ex wate	
6	II	Meta sedin non-c and r rocks recry	

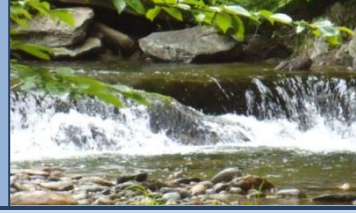


Streams Report Generator - Stream Site Information

Site_ID	Stream	Basin	School/College /University	Town	Staff/Faculty Mentor	Student Researchers
VERMONT						
LR_CooBrk_827	Lamoille River	Lamoille River	Hazen Union High School	Hardwick, VT	Jay Modry	Alan Therrian and Meg Urie
Elmore Branch	Elmore Branch	Lamoille River	Hazen Union High School	Wolcott, VT	Jay Modry	Alan Therrian and Meg Urie
LCD_PoBrk_133	Potash Brook	Lake Champlain Direct	South Burlington High School	South Burlington, VT	Curt Belton	Aya Ainamee and Rebecca Goldberg
WR_SlBrk_711	Snipe Island	Winooski River	South Burlington High School	Richmond, VT	Curt Belton	Aya Ainamee and Rebecca Goldberg
T1-4	Munroe Brook	Lake Champlain Direct	Rice Memorial High School	Shelburne, VT	Sharon Boardman	Nathan Boardman and Katie Rapoza
M4	Munroe Brook	Lake Champlain Direct	Rice Memorial High School	Shelburne, VT	Sharon Boardman	Nathan Boardman and Katie Rapoza
LCD_PndBrk_179	Pond Brook	Lake Champlain Direct	Colchester High School	Colchester, VT	Will Warren	Emily Bishop and Rogelio Zimbron
WR_CeBrk_259	Centennial Brook	Winooski River	Colchester High School	Burlington, VT	Will Warren	Emily Bishop and Rogelio Zimbron
HRD_FuBrk_1013	Furnace Brook	Hudson River	Mount Anthony High School	Bennington, VT	Dan Rosenthal	Nicholas Harris and Alex Romac
HRD_WaRv_767	Walloomsac River	Hudson River	Mount Anthony High School	Bennington, VT	Dan Rosenthal	Nicholas Harris and Alex Romac
Ryder Brook	Ryder Brook	Lamoille River	Lamoille Union High School	Morristown, VT	Amber March	Eliza Fowler and Lucy Rogers
LR_WatBrk_876	Waterman Brook	Lamoille River	Lamoille Union High School	Johnson, VT	Amber March	Eliza Fowler and Lucy Rogers
OC_BuiBrk_1285	Bully Brook	Otter Creek	Mill River Union High School	Wallingford, VT	Michael Gamache	Anne Coco and Julia Doiron

Tip: You may want to refer to the "Help" pages that define each variable. They can help you determine the scale and intervals on which data are collected.

Module 2



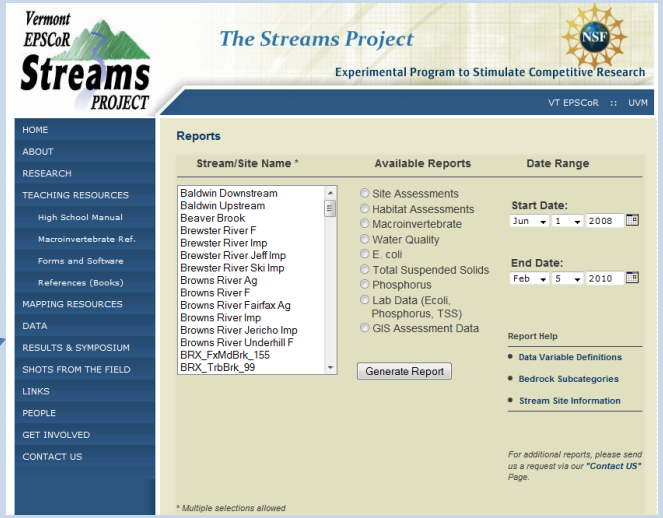
How to use available data to specify your question

At its essence, research addresses **a question** about the phenomena of interest. In the case of the Streams Project, the phenomena are several aspects of stream health. The scientific community refers to this question as a **hypothesis** which you, the scientist, **test** by collecting data and analyzing it with statistics.

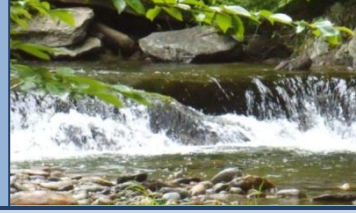
Use these specifications you made about time, space, and stream health on the previous slide to turn your idea into a question.

Take the previous example about our interest in the relationship between macroinvertebrate abundance and temperature. You can turn it into a question by saying, "Are macroinvertebrates more abundant in streams with higher temperatures?"

Check back with the database to make sure that the data you need to answer your question are available. Again, you may want to use the "Help" pages that define each variable.



Module 2



Why you need a hypothesis

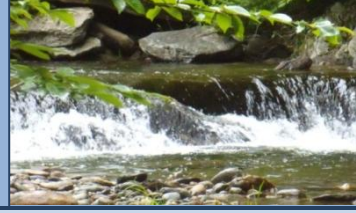
Throughout this module we will use the term **"hypothesis"** to refer to your question. In statistics, a hypothesis is really composed of two hypotheses: a "**null hypothesis (H_o)**" and an "**alternative hypothesis (Ha).**" The hypothesis is your question phrased in the format:

H_o = null hypothesis = *The hypothesis of **no** difference or **no** real relationship between or among parameters of interest.*

Ha = alternative hypothesis = *The hypothesis that there **is** a difference or real relationship between or among parameters of interest.*

Statistical tests allow you to determine if there are or are not differences or relationships in the parameters of interest. Such statistical tests are structured in a way that requires both a null and alternative hypothesis. Module 5 demonstrates this.

Module 2



Finalizing your hypothesis

From here, you should be able to take your questions and turn them into hypotheses!

Take the following examples:

Question 1: Are macroinvertebrates more abundant in streams with higher temperatures?

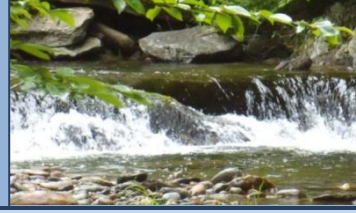
For this question we would write our hypothesis as the following:

H_o = *There is no difference between the number of macroinvertebrates in warm sites and cooler sites.*

Ha = *There is a difference between the number of macroinvertebrates in warm sites and cooler sites.*

Continued…..

Module 2



Finalizing your hypothesis

Take the following example:

Question 2: Are the levels of phosphorus and TSS related?
For this question we would write our hypothesis as the following:

H_o = *There is no relationship between the levels of phosphorus and TSS.*
Ha = *There is a relationship between the levels of phosphorus and TSS.*

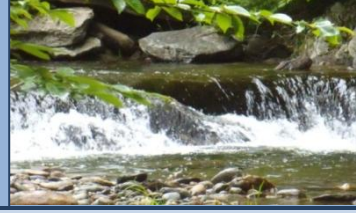
The above examples are very simple. You may choose to ask a similarly simple question, but don't be afraid to add layers of complexity:

Question 2: Are the levels of phosphorus and TSS related at my urban site, but not at my forested site?

For this question we would write our hypothesis as the following for each site:

H_{o_1} = *There is no relationship between the levels of phosphorus and TSS at my urban site.*
Ha_1 = *There is a relationship between the levels of phosphorus and TSS at my urban site.*
H_{o_2} = *There is no relationship between the levels of phosphorus and TSS at my forested site.*
Ha_2 = *There is a relationship between the levels of phosphorus and TSS at my forested site.*

Module 2



Finalizing your hypothesis

SUMMARY

- High school teams, undergraduate students, and faculty researchers all contribute data to the Streams Project database.
- All quality controlled data are available online.
- The “Help” pages define what each variable means and can offer valuable insight into how you can ask your research question.
- Reconcile your ideas with the data available in the streams database, using the “Help” pages.
- Define the time, spatial extent, and parameters of the ideas you have on your concept map.
- Turn your idea into a question and a null and alternative hypothesis.
- Hypotheses are necessary for conducting statistical tests.