Lesson 3.1 – Data Analysis

In order to study how clean and healthy our water is, scientists gather data (information) about the water in different lakes, rivers, and streams. They also keep track of data for weather patterns like rainfall to see if there is a relationship between climate and water quality.

You are going to analyze water quality and precipitation (rain and snow) data that was collected in and around the area where the Maumee River enters Lake Erie. This data was collected every day for over 10 years (2002-2014)! That is a lot of information. To save you some time, we have selected different pieces of the data that we want you to analyze. When you analyze these data, we want you to look for patterns. In particular, we want you to see if different types of data seem related to each other. In other words, you should try to figure out if changes or patterns in one set of numbers seem to have a clear relationship with changes or patterns in other sets of numbers.

But what does this mean? Here is an example about global temperatures and arctic ice cover to show you what this might look like and give you some practice.

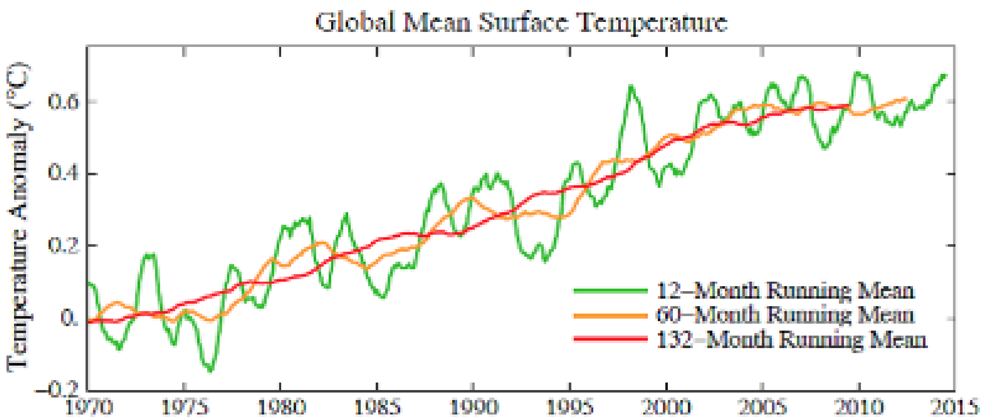


Fig. 6. 12-month, 60-month and 132-month running means of global surface temperature.

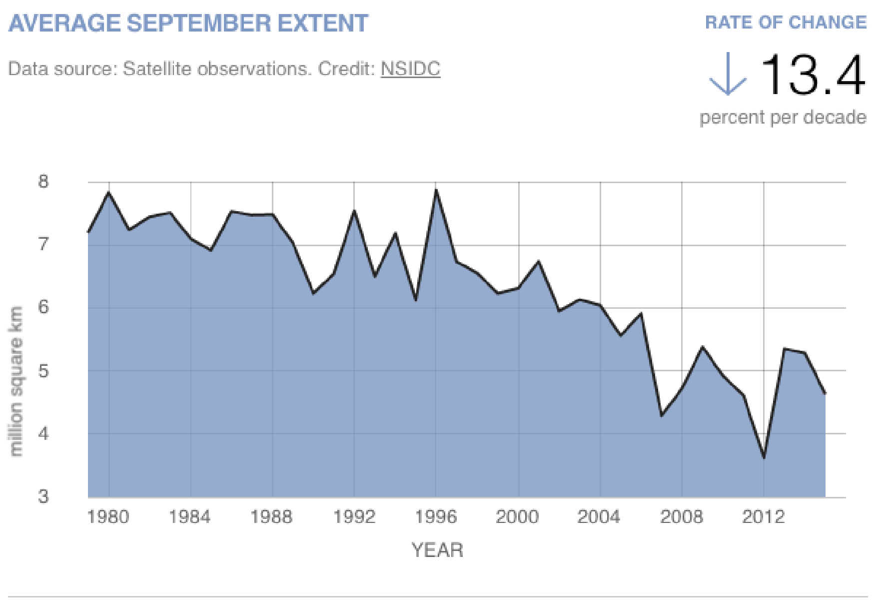
http://csas.ei.columbia.edu/2015/01/16/global-temperature-in-2014-and-2015/

The graph above shows changes in the mean, or average, surface temperature of Earth from 1970 to 2015.

1) What is the overall pattern? Are the numbers overall going up or down?

The green line shows yearly averages, the yellow line shows the 5-year average, and the red line shows the 11-year average. The green line shows more ups and downs because it has more data points on it (yearly) and shows more short-term changes. The longer time periods (5 and 11 years) flatten out a bit because they show the long-term trends. However, all three lines show a general increase in global surface temperatures.

2) The graph below shows the extent of Arctic sea ice, or the area of the Arctic Ocean that is covered by ice, from 1980 to 2012. Although there are a lot of ups and downs, what is the general pattern or trend here? How do the numbers seem to be changing overall? Is Arctic sea ice increasing (going up) or decreasing (going down) overall?



http://climate.nasa.gov/vital-signs/arctic-sea-ice/

Go back to each graph and describe in very basic terms the overall pattern in the data that you see for the years 1980 to 2012:

3) Global mean surface temperature: ____

4) Arctic sea ice coverage: ____

5) Are the trends in the two tables similar or different? If they are different, how are they different? Jot down some ideas below:

____

____

6) So, do you think there is a relationship between changes in global temperature and arctic sea ice coverage between 1980 and 2012? Do you think the patterns are connected? If not, explain why not, and if so, describe the connection below:

You should have noted that the average global temperature increased between 1980 and 2012, while the arctic sea ice coverage decreased. The data have an inverse relationship, which means that they change in opposite directions… as one goes up, the other one goes down.

This relationship does not necessarily mean that one change caused the other, although that might be the case. To figure that out, scientists have to do a lot more research and look at a whole range of other factors.

This is an important point! Correlation (when data sets have changes in patterns that seem similar or almost directly opposite) is not causation! In other words, data patterns that look like they have clear relationships are not always connected. Harvard law student Tyler Vigen uses funny graphs to explore these relationships and to show how data sets with similar trends are not always related. One of his graphs shows that divorce rate in the state of Maine decreased between 2000 and 2009 at almost the same rate as the per capita consumption of margarine in the United States (http://www.tylervigen.com/spurious-correlations). We can reach a common sense conclusion that the amount of margarine people eat probably doesn't cause changes in Maine's divorce rate, or vice versa.

However… we can also use common sense and logic to make conjectures (best guesses) about other data relationships, and these conjectures can become hypotheses we test with ongoing research. For example, we can ask if it makes sense that increasing temperatures might cause decreasing amounts of ice, and of course, this makes a lot of sense. Then we can ask additional questions and do more research to see if this is indeed the case (and it certainly does appear to be what is happening).

When you look at the data below about rainfall and water quality, you will use the same sort of process.

- First, look for patterns and try to describe long-term changes in each piece of data (for each separate variable). Are they going up, going down, staying the same, or are they all over the place?
- Then, see if any of these data have similar patterns, or patterns that are almost opposite.
- Then, think about whether or not the patterns in individual variables might be logically connected.
- Finally, describe the relationships you see and make conjectures that you could investigate with more research.

Table 1.

					Water Quality Variables	
	1	2	3	4	5	6
	Days and precipitation (2002-2014)	**Count of date**	**Average of Precipitation (rainfall+snowfall, mm H2O/day)**	**Average of SS- suspended solid**	**Average of TP - total phosphorous load**	**Average of NO - nitrate + nitrite load**
1	Days with less than 23 mm of precipitation that were not within the 3 days after a day with 23 mm or more of precipitation.	4,658	2	2,042	5	73
2	Days with 23 mm or more of precipitation.	24	29	11,668	23	176
3	Days 1 day after a day with 23 mm or more of precipitation.	22	8	25,431	55	360
4	Days 2 day after a day with 23 mm or more of precipitation.	22	2	24,474	64	392
5	Days 3 day after a day with 23 mm or more of precipitation.	22	2	19,223	54	362

Table 1 allows us to compare days with more than 23 mm of precipitation to days with less than 23 mm of precipitation, and it also provides data for the day after a day with more than 23 mm of precipitation, two days after that day, and three days after that day. With these data, we can see how days that had less than 23 mm of precipitation compare to days with heavy precipitation (23 mm or more), and we can also analyze what happens to water quality during the three days immediately following each heavy precipitation day.

So… there were only 24 days that had at least 23 mm of precipitation. Compare the days with 23 mm or more of precipitation to the days with less than 23 mm that were not within 3 days of a heavy precipitation day (row 2 compared to row 1). Look at each of the water quality variables.

Do days with heavy rainfall have HIGHER or LOWER levels of suspended solids, phosphorous, and nitrogen? Describe what you see in the data in the questions below:

7) What happens to the water quality variables one day after a day with more than 23 mm of rain?

a) 2 days after?

b) 3 days after?

c) What might explain this pattern?

8) Looking at these data as a whole, describe the relationship between heavy rainfall and levels of suspended solids, phosphorous, and nitrogen:

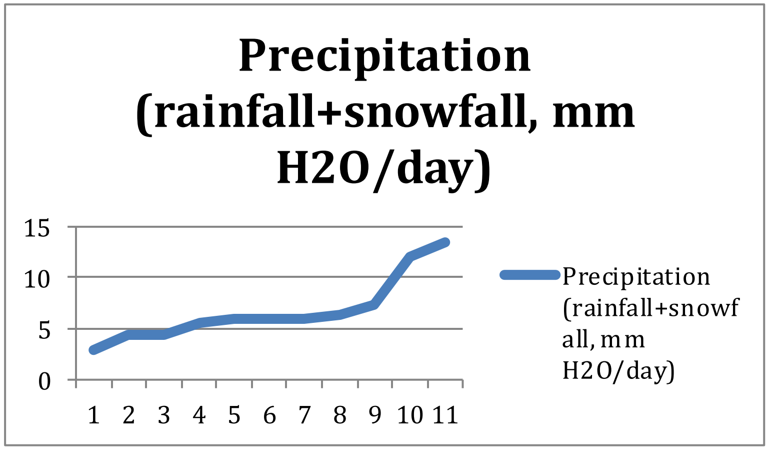
Table 2.

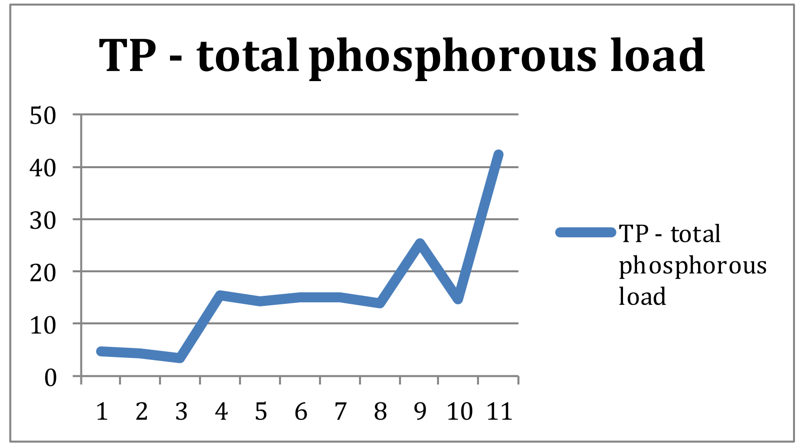
Ranking of year for average rainfall (1 is lowest, 10 is highest)	Year	Precipitation (rainfall+snowfall, mm H2O/day)	TP - total phosphorous load	NO - nitrate + nitrite load
1	2004	2.910686348	4.882224109	76.9753037
2	2012	4.313091977	4.371615994	32.11761803
3	2005	4.411742637	3.41975787	74.11848015
4	2008	5.480396129	15.35629592	157.8883257
5	2010	6.006011318	14.21334539	229.3805708
6	2009	6.025194744	14.95468244	157.6766332
7	2007	6.029941782	14.93903059	149.7008737
8	2013	6.409984104	13.7150662	235.7055895
9	2011	7.44053657	25.19256338	265.9436263
10	2006	12.0789119	14.76551731	324.8754242
11	2003	13.47617575	42.33962402	754.1997691

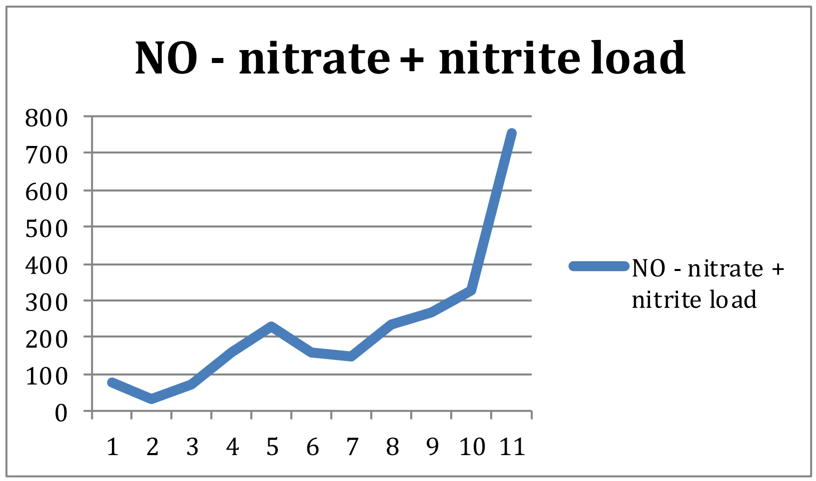
Table 2. provides a different way to look at some of the same patterns, and it gets more complicated. Table 2 shows data for 11 years, 2003-2013, but it has been organized by sorting the data to show you the years ordered lowest to highest for average daily precipitation. So, between 2003 and 2013, 2004 had the lowest daily average of precipitation and 2003 had the highest daily average.

9) Do you see a pattern in the other variables? Do years with higher average rainfall also have higher average levels of phosphorous and nitrogen?

To help you see the patterns more clearly, look at basic trends in the graph for each variable on the next page.







10) Do the patterns look similar overall? Do you think there is a relationship between precipitation and water quality as seen in phosphorous and nitrogen levels? How would you describe that relationship?

11) What conjectures, or best guesses, can you make about possible causal relationships (changes in one variable cause changes in others) between precipitation and the other two variables?

Table 3.

20 days from 11 years with no rainfall and lowest discharge rates

date	Precipitation (rainfall+snowfall, mm H2O/day)	TP - total phosphorous load	NO - nitrate + nitrite load
10/10/10	0	0.021267414	0.01519101
9/13/02	0	0.024327632	0
9/12/02	0	0.026859467	0
10/9/10	0	0.031871767	0.00247068
7/12/12	0	0.042622892	0.026639308
7/11/12	0	0.044168901	0.026859467
9/6/02	0	0.041556341	0.006702636
10/8/10	0	0.044643468	0.007142955
8/26/08	0	0.079345508	0.011228138
9/9/02	0	0.038542601	0
11/8/10	0	0.041075252	0.002902845
9/8/02	0	0.049756062	0
10/17/10	0	0.065717631	0.004867973
9/12/05	0	0.068031752	0.007448732
10/11/08	0	0.060260119	0.199024248
10/16/05	0	0.075623642	2.309877475
9/11/05	0	0.08200635	0.021812921
10/11/04	0	0.082047355	0
8/25/12	0	0.069482359	0.012867104
9/6/13	0	0.094002753	0.013258498
AVG	*0*	*0.054160463*	*0.1334147*

[redacted]

20 days from 11 years with highest precipitation

date	Precipitation (rainfall+snowfall, mm H2O/day)	TP - total phosphorous load	NO - nitrate + nitrite load
3/5/11	24.846	49.69148181	522.4207116
7/3/08	25.533	8.546631052	116.6247601
5/11/06	25.761	1.623407717	77.11186654
7/28/11	25.881	0.837412355	2.099078767
6/13/13	25.946	3.683476655	182.6438868
9/1/03	26.054	0.868162556	0
12/22/13	26.979	35.74696629	434.5264895
2/6/08	27.229	137.007255	290.9090659
2/28/11	27.241	28.93324787	334.6193015
3/8/09	27.527	17.38379934	124.5222506
11/29/11	32.207	40.21157398	337.5747398
5/26/11	32.328	107.3617909	520.5697404
10/20/11	32.453	12.32893571	118.0495594
9/13/08	32.67	0.382148083	0.052349052
10/17/06	32.676	3.460247901	36.37696512
12/1/06	33.326	20.62650513	115.6571588
8/20/07	34.164	4.188690673	38.73789107
9/14/08	34.862	1.014642054	3.354864856
4/11/13	35.19	5.799188885	125.1440791
8/21/07	42.191	31.53027458	410.0349609
AVG	*29.98785714*	*24.43672231*	*180.7765212*

The left side of Table 3 shows the 20 days in 2003-2013 that had the lowest river discharge rates, all of which had no precipitation. The right side of Table 3 shows the 20 days with the most rainfall in these 11 years.

12) How do the other variables compare? Does lower rainfall seem to predict lower phosphorous and nitrogen levels? Summarize the pattern you see in Table 3. What do you notice when looking at phosphorous and nitrogen levels when there is no precipitation as compared to when there is a lot of precipitation (more than 23 mm in 24 hours)?

- Now, fill out the chart below based on your analysis and review of all the data tables:

	Description of what the table shows	Summary statement: what relationships do you see between precipitation and the water quality indicators of phosphorous and nitrogen levels?
13) Table 1		
14) Table 2		
15) Table 3		
16) Looking across all three data samples, how can you summarize the relationship between precipitation and phosphorous and nitrogen levels?		

17) How might you explain the relationship between precipitation and phosphorous and nitrate levels? What do you think is happening? Jot down an explanation, or develop an illustration that might show the connection between these variables.	
18) Scientists who study climate change have noticed that the frequency of severe weather events (such as very heavy rainfall) is increasing. If we expect more extreme rain events in the spring and summer in the future, what effect might we expect on water quality? Remember that phosphorous and nitrate levels provide one way to measure water quality, and that high levels of these nutrients are signs of poor water quality (you will learn more about why in the last lesson in this unit).	