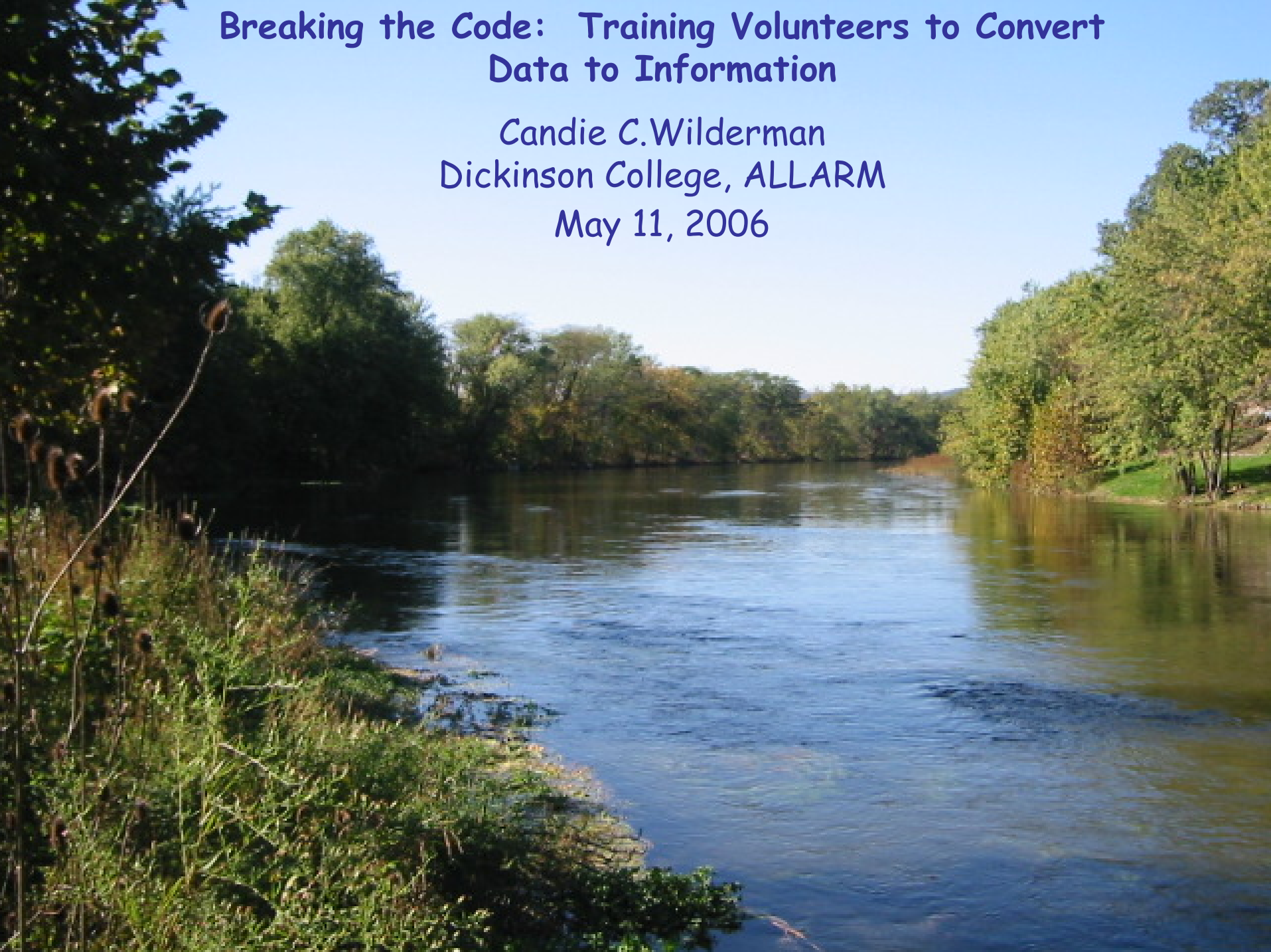
Breaking the Code: Training Volunteers to Convert Data to Information

Candie C.Wilderman
Dickinson College, ALLARM

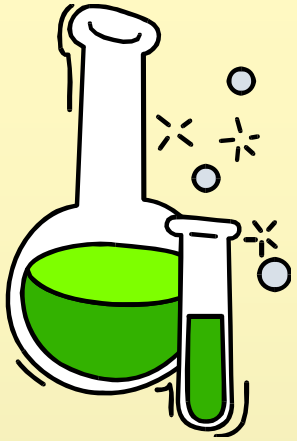
May 11, 2006



Topics to be covered

- ✓ Different operational models for citizen science, and the role of the service provider
- ✓ The importance of volunteers engaging in data analysis and interpretation
- ✓ A two-phase model for training volunteers to interpret data

What are some models for citizen science?



- Citizen science involves a research partnership between community people and professional scientists.
- There are a variety of successful operational models for this partnership.
- These models differ in their goals, the nature and scope of the projects, and the extent of community control over the definition and implementation of the project.

Categorizing the various models for community science can be based on answers to five questions:

- Who defines the problem?
- Who designs the study?
- Who collects the samples?
- Who analyzes the samples?
- Who interprets the data?

Community Workers Model #1

Who defines the problem?	Who designs the study?	Who collects the samples?	Who analyzes the samples?	Who interprets the data?
Professional scientists	Professional scientists	Community	Professional scientists	Professional scientists

Examples of Community Workers Model #1

CORNELL LAB of ORNITHOLOGY

Study of the infestation of blue bird nests by Protocalliphora (blowflies)

Rivers and Streams

MD DNR Stream Waders Volunteer Monitoring Program (Macroinvertebrate Analysis)

Community Consulting Model (Science for the People)

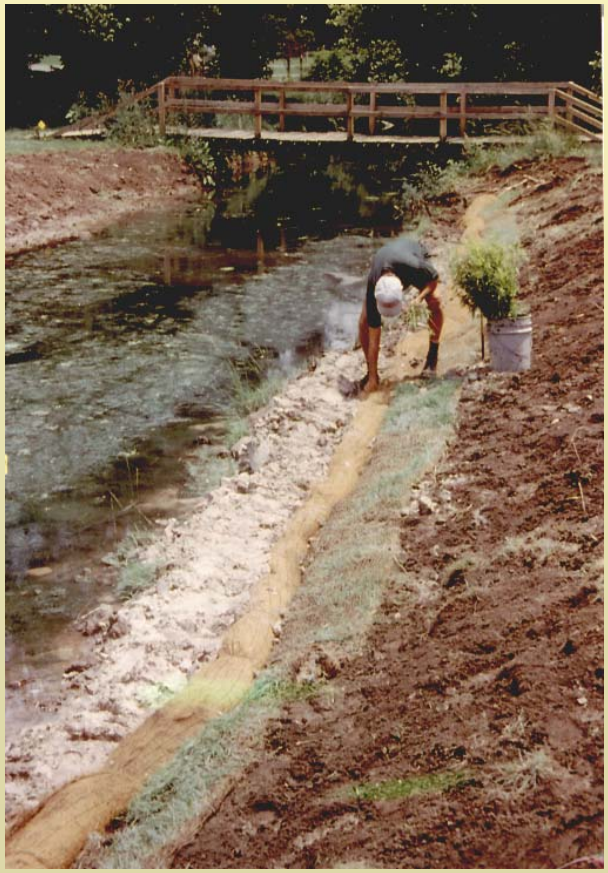
Who defines the problem?	Who designs the study?	Who collects the samples?	Who analyzes the samples?	Who interprets the data?
Community	Professional scientists	Professional scientists	Professional scientists	Professional scientists

Examples of the consulting model

Some PA Growing Greener Grants support programs using this model

European Science Shops

The Mully Grub Restoration Project conducted by ALLARM



Community-based, Participatory Research Model (Science by the People)

Who defines the problem?	Who designs the study?	Who collects the samples?	Who analyzes the samples?	Who interprets the data?
Community	Community	Community	Community	Community

Examples of Community-based, Participatory Research Model

Watershed-based projects

The Alliance for Aquatic Resource Monitoring (ALLARM) is:

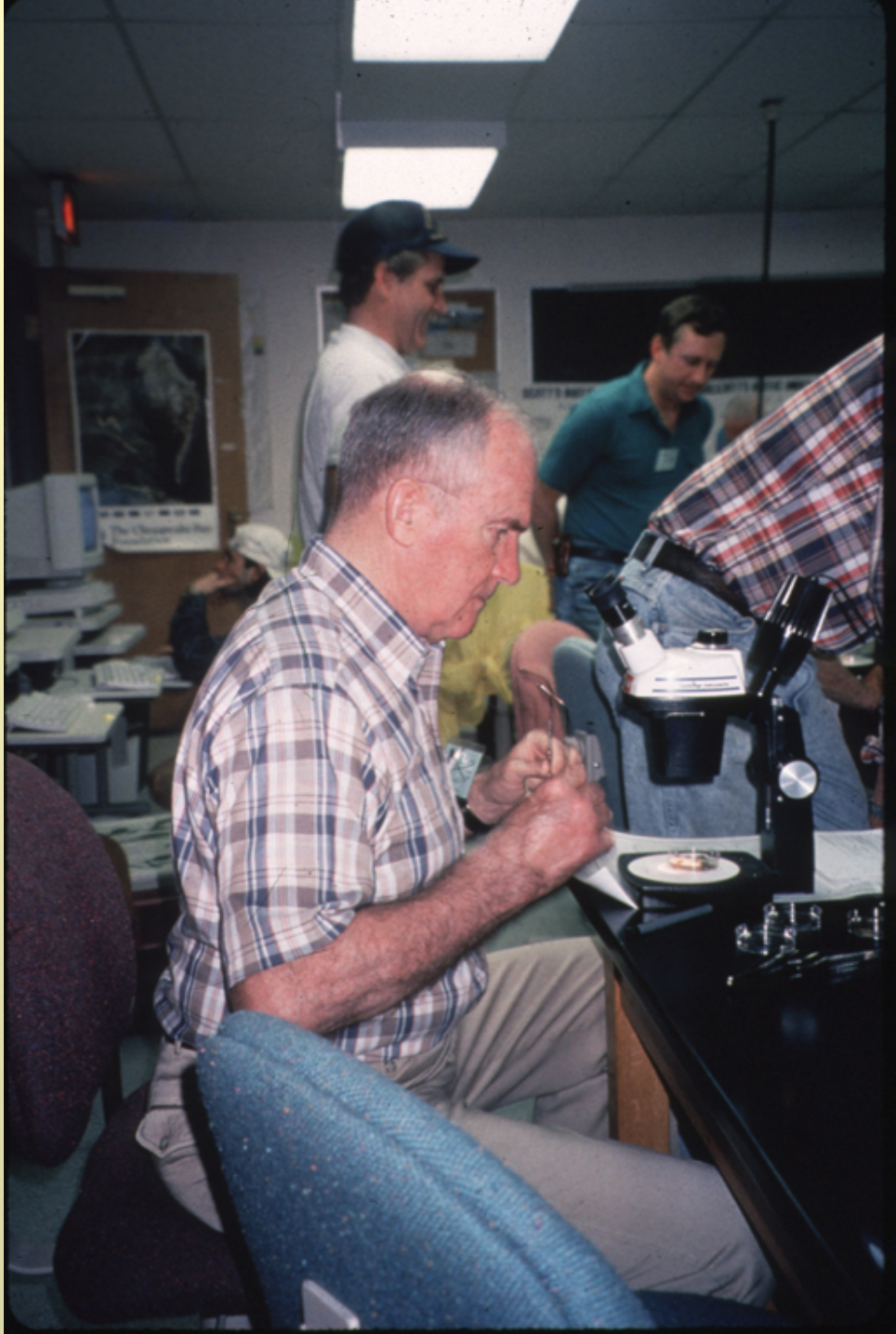
- A project of the Environmental Studies Department at Dickinson College in Carlisle, PA, founded in 1986 by Candie C. Wilderman.
- Staffed by two full-time professionals, a part-time faculty Science Director and 10-12 students.
- A service provider offering capacity-building programmatic and scientific technical assistance to watershed groups throughout the Commonwealth of PA.

ALLARM'S goals are:

1. To empower communities with scientific knowledge to address their environmental concerns, and

2. To enhance the quality of undergraduate education at Dickinson College.

The roles in which ALLARM engages citizen-scientists have varied over the past 20 years, but recently we have focused on the community-based participatory research model.

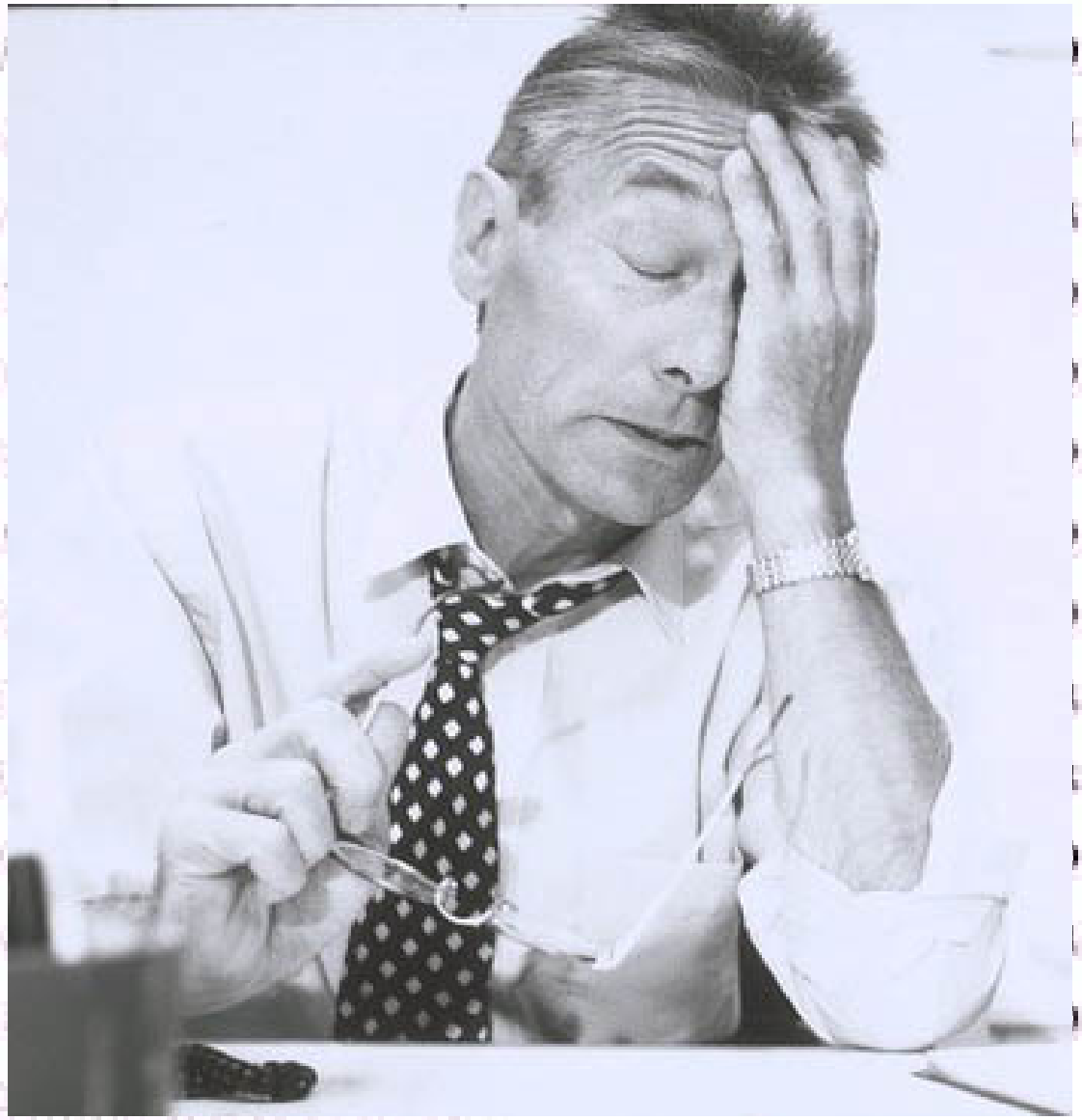


We have found the greatest challenges in this model are the design and the interpretation of the study

Who defines the problem?	Who designs the study?	Who collects the samples?	Who analyzes the samples?	Who interprets the data?
Community	Community	Community	Community	Community

These steps involve intensive mentoring by the professional partner (service provider) and a high level of commitment by the volunteers.

The Data Interpretation Step: Can Volunteers Climb the Learning Curve to Convert their Data to Information?



Presenting the story to the group

Training the volunteers to find the story in the da elves

Steps in the Data to Information Training Process[1]

Workshop #1: Learn the basics using a virtual watershed (Dickinson Creek)

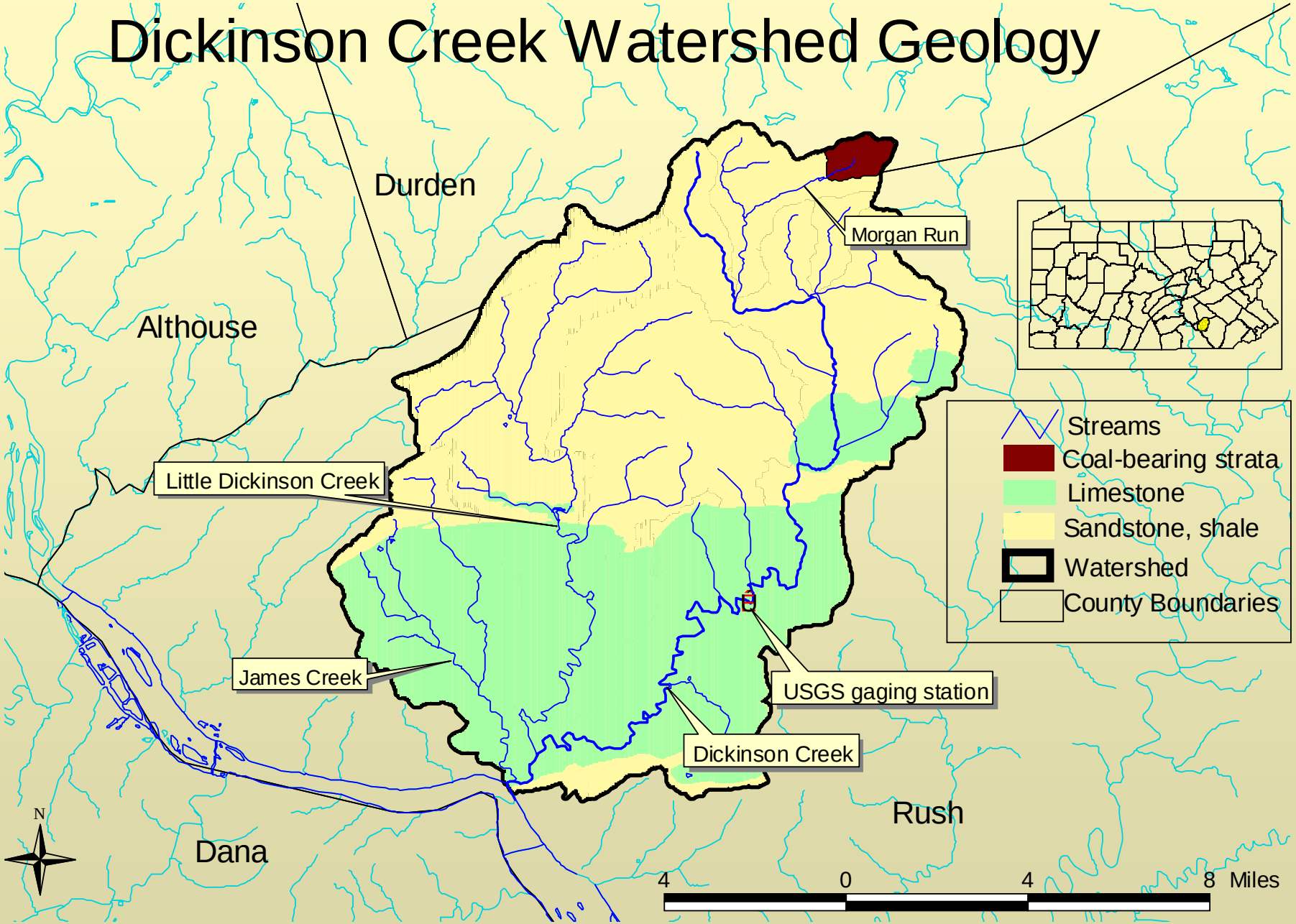
Workshop #2: Apply these skills to the real watershed data collected by volunteers

[1] These workshops were developed in cooperation with River Network, the Stroud Water Research Institute and the Delaware Riverkeeper.

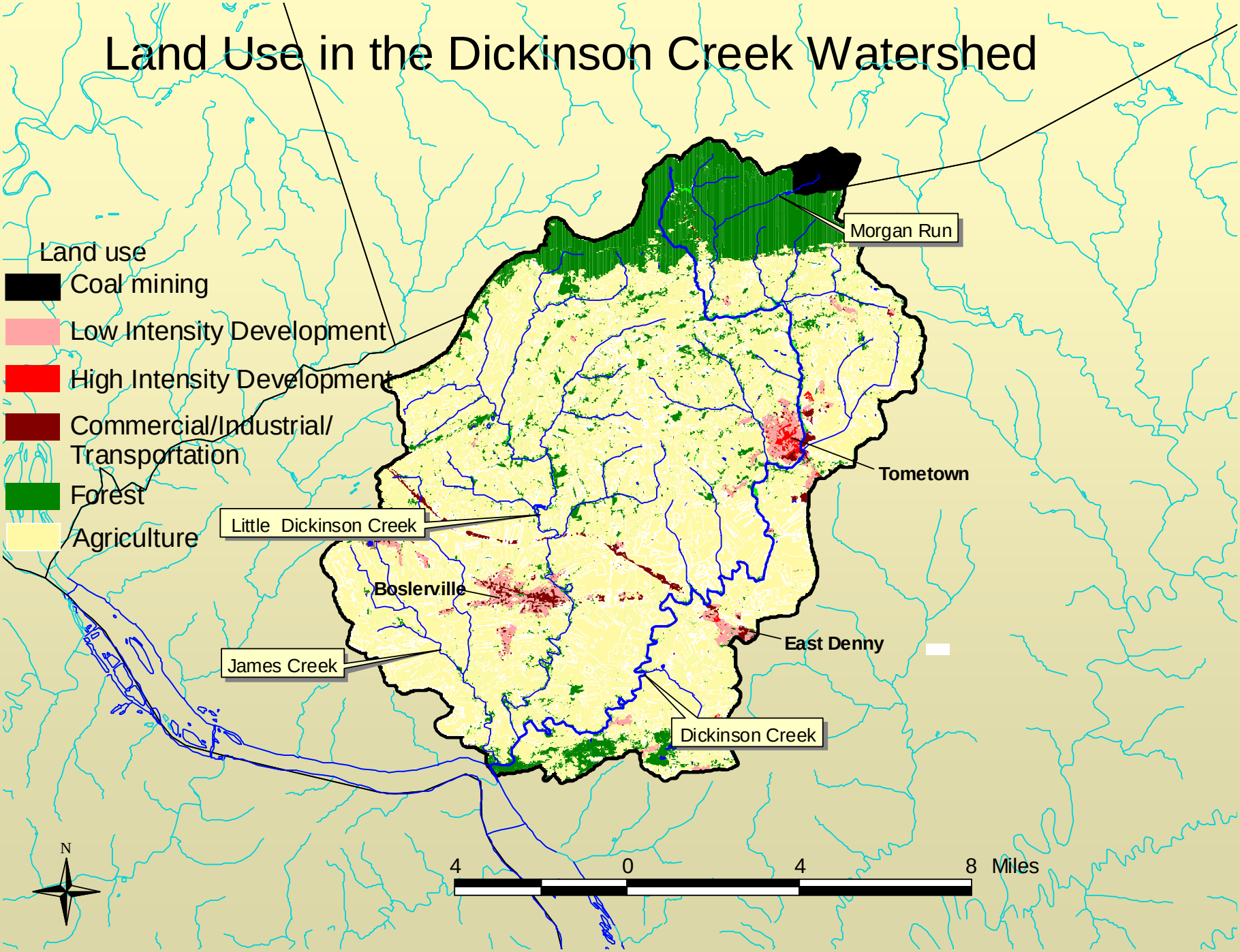
Why a virtual watershed?

- Allows us to keep the data simple, the sites limited to a few, and the patterns clear.
- Allows us to use just those indicators with which the group is familiar.
- Allows us to demonstrate expected relationships between indicators, land use, geology, and seasons.
- Allows us to add outliers and impossible values for volunteers to discover, discuss, and figure out how to handle.

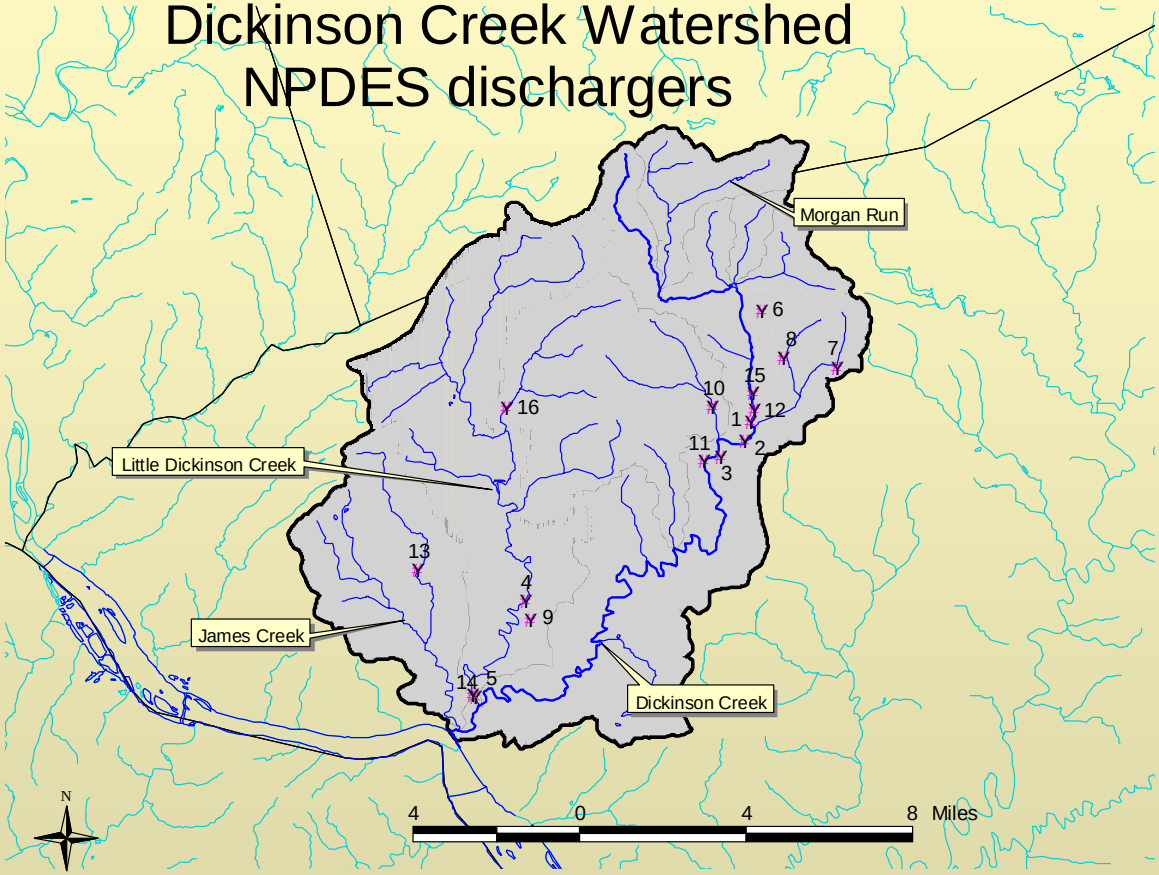
Dickinson Creek Watershed Geology



Land Use in the Dickinson Creek Watershed

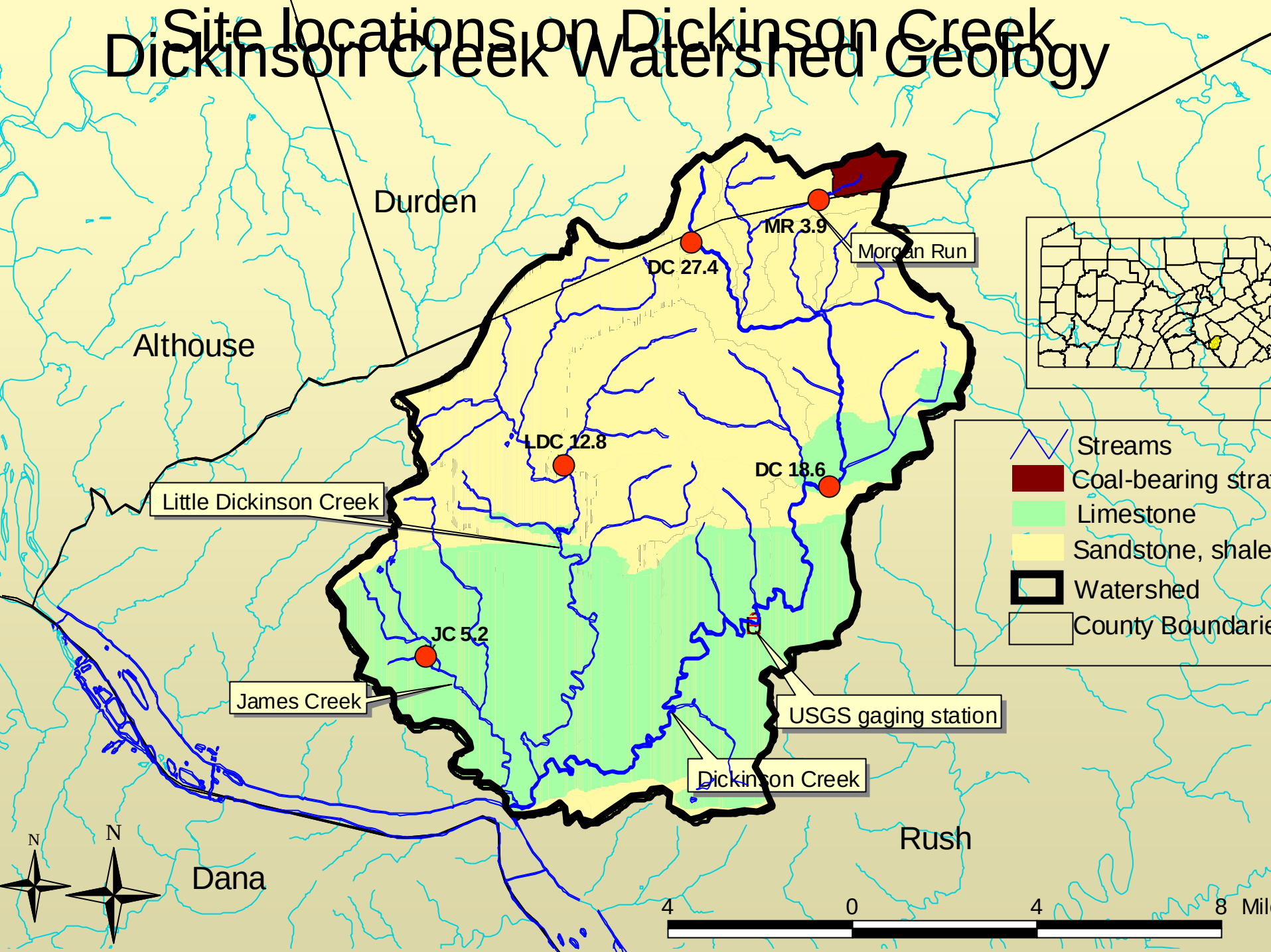


Dickinson Creek Watershed NPDES dischargers



Facility Name	Number on map
Raymark Inc.	1
Fuller Co. -- Tometown Plant	2
Tometown STP	3
Boslerville Borough Authority	4
Marietta-Donegal Jnt Sew Auth	5
Pine View Acres MHP	6
Chadaga, PCS. MD.	7
Pleasant View Retirement Comm	8
Boslerville Wire Corporation	9
Tometown Borough	10
Hilltop Acres MHP	11
Northwestern Rush County	12
Imgrund, Lauren	13
Boslerville Borough Authority	14
Worley & Obetz, Inc.	15

Site locations on Dickinson Creek
Dickinson Creek Watershed Geology



AN AGENDA

6:00-7:00 PM Statistics 101: Introduction to data analysis

7:00-8:00 PM **The pieces:** Finding the story of Dickinson Creek, one indicator at a time (small group activity)

Small group presentations

8:00-9:00 PM **The big picture:** large group compilation and discussion

Statistics 101

THE DATA STRUCTURE OF THE DICKINSON CREEK VIRTUAL WATERSHED

Sample Raw Data for Statistics

Total Suspended Solids (TSS) (ppm)													
LAND USE	**SITE**	**JAN**	**FEB**	**MAR**	**APR**	**MAY**	**JUN**	**JUL**	**AUG**	**SEP**	**OCT**	**NOV**	**DEC**
Forest	DC 27.4	9.3	9.9	10.4	10.0	13.3	7.8	0.0	0.0	0.0	0.0	6.7	9.6
Ag	JC 5.2	35.4	45.3	52.2	48.5	136.7	26.5	15.3	151.6	13.1	13.6	25.1	32.6
Ag; Pt. Dis.	LDC 12.8	52.2	152.3	189.3	175.2	456.3	125.6	100.3	50.2	56.3	62.3	103.2	135.2
Forest; AMD	MR 3.9	4.8	5.9	6.9	6.3	12.0	4.6	0.0	0.0	0.0	0.1	5.6	5.9
Urban	DC 18.6	133.8	158.2	185.6	178.3	365.2	85.6	25.6	20.8	19.8	25.4	75.8	125.6

Sample Data Summary Tables for Box and Whisker Plot

Total Suspended Solids (TSS) (ppm) - Annual Summary for Each Site

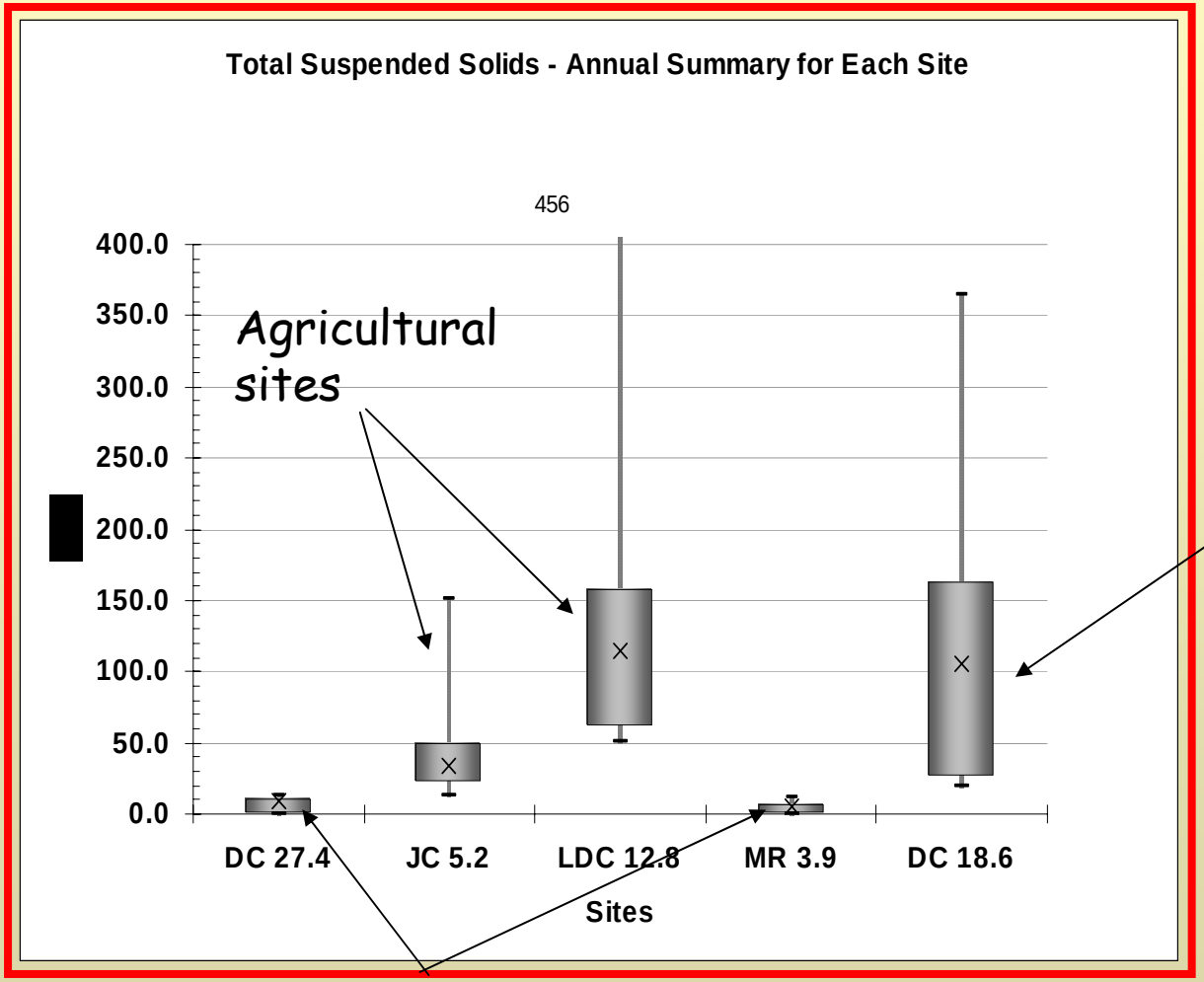
LAND USE	SITE	Average	Min	25th	Median	75th	Max	Range	IQ Range
Forest	DC 27.4	6.4	0.0	0.0	8.6	9.9	13.3	13.3	9.9
Ag	JC 5.2	49.7	13.1	22.7	34.0	49.4	151.6	138.5	26.8
Ag; Pt. Dis.	LDC 12.8	138.2	50.2	60.8	114.4	158.0	456.3	406.1	97.2
Forest; AMD	MR 3.9	4.3	0.0	0.1	5.2	6.0	12.0	12.0	5.9
Urban	DC 18.6	116.6	19.8	25.6	105.6	163.2	365.2	345.4	137.7

Total Suspended Solids (TSS) (ppm) - Monthly Summary for *All Sites*

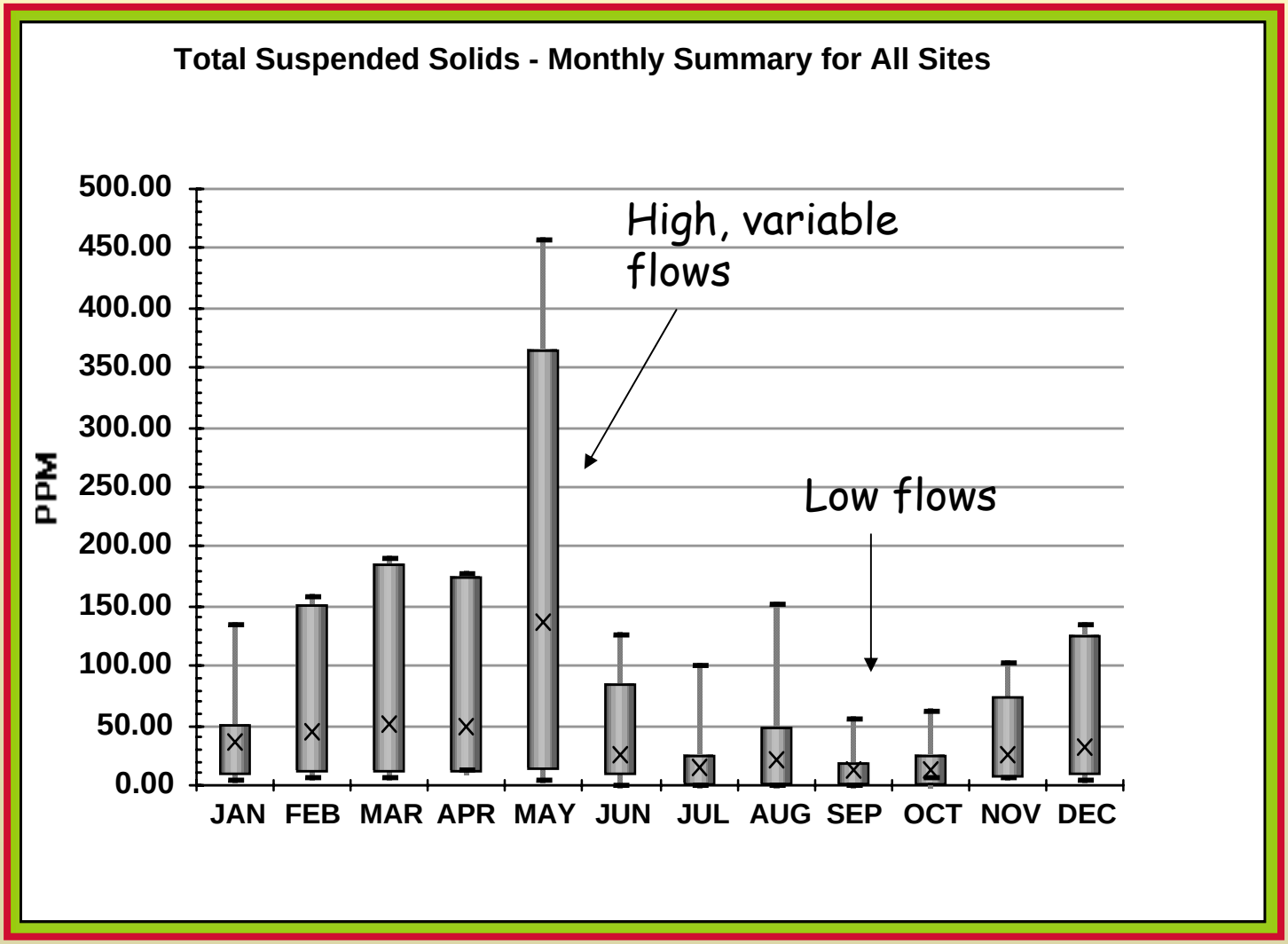
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Average	47.10	74.32	88.88	83.66	196.70	50.02	28.24	44.52	17.84	20.28	43.28	61.78
Min	4.80	6.90	6.30	12.00	4.60	0.00	0.00	0.00	0.00	5.60	5.90	4.34
25th	9.30	9.90	10.40	10.00	13.30	7.80	0.00	0.00	0.00	0.10	6.70	9.60
Median	35.40	45.30	52.20	48.50	136.70	26.50	15.30	20.80	13.10	13.60	25.10	32.60
75th	52.20	152.30	185.60	175.20	365.20	85.60	25.60	50.20	19.80	25.40	75.80	125.60
Max	133.80	158.20	189.30	178.30	456.30	125.60	100.30	151.60	56.30	62.30	103.20	135.20

THE BOX AND WHISKER PLOTS OF THE DICKINSON CREEK VIRTUAL WATERSHED

Sample Box and Whisker Plot for Sites



Sample Box and Whisker Plot for Months



The pieces: finding the story of Dickinson Creek

IA. Using Raw Data and Maps

1. e.g. What is the highest nitrate concentration recorded at any site?

- Group is divided into small groups; each small group has **1-2 chemical indicators** to study

Ib: Using Summary Data

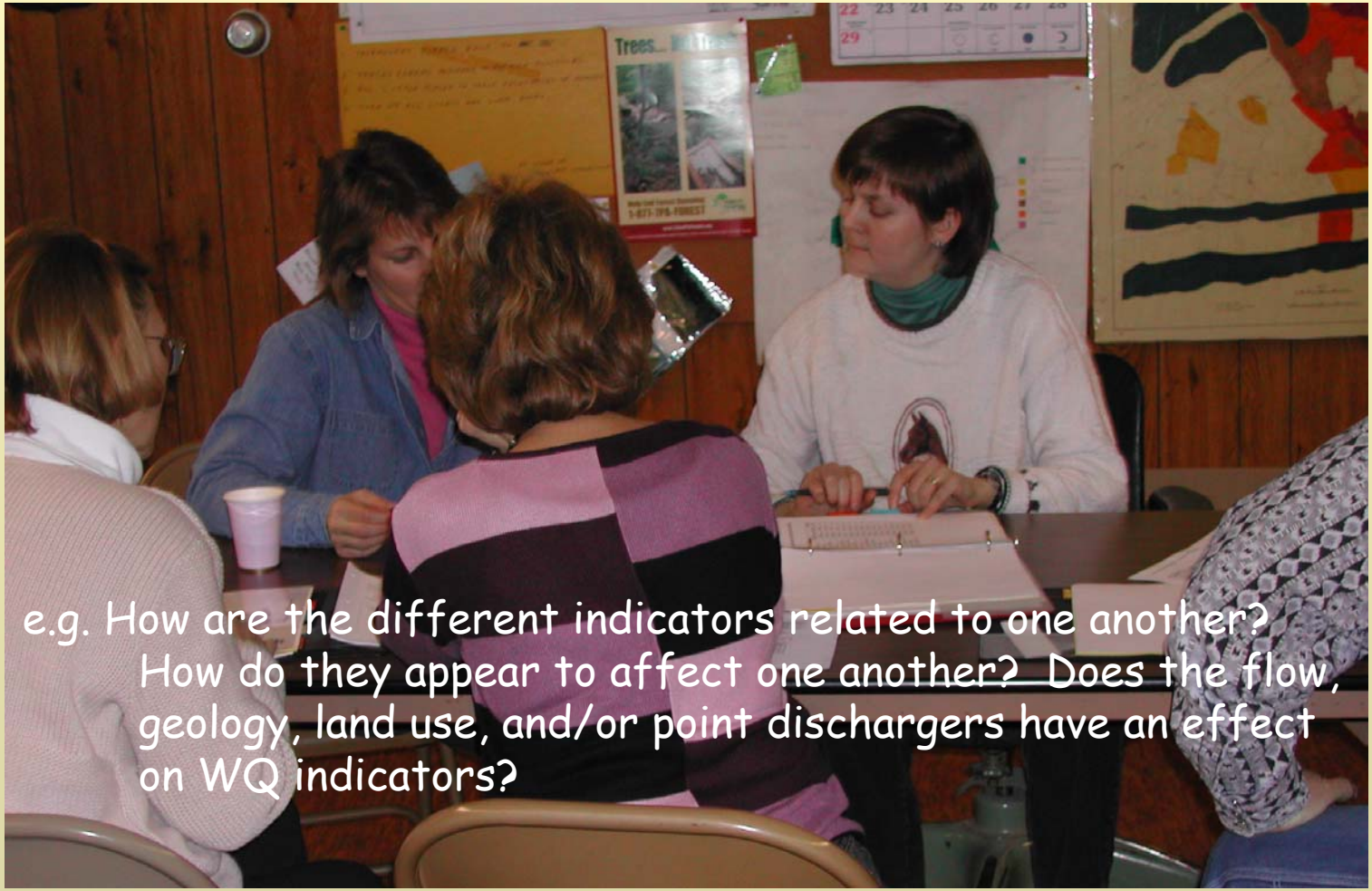
1. e.g. Based on the Box-and-Whisker graphs for your indicator, which sites are most different from each other?

II: Interpretation

1. e.g. If levels of your indicator are different between sites, what might explain this?

The big picture: large group compilation and discussion

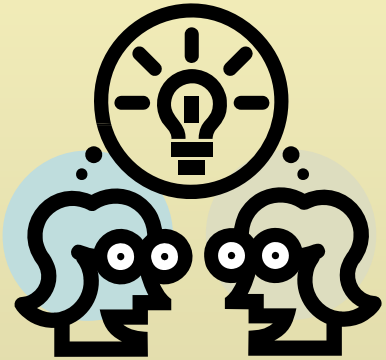
Section III: Looking At All the Data Together



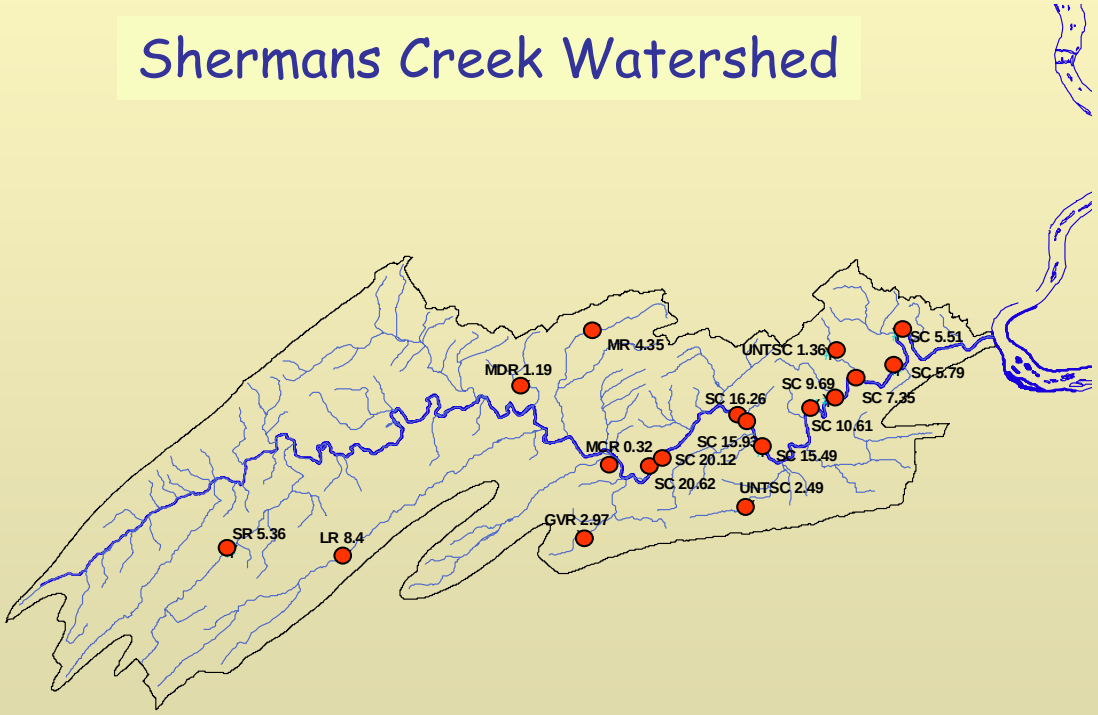
e.g. How are the different indicators related to one another?
How do they appear to affect one another? Does the flow, geology, land use, and/or point dischargers have an effect on WQ indicators?

What they learn beyond basic statistical and graphical concepts

- A feel for typical concentrations of indicators and the range of variation
- A sense of how indicators vary with seasons, land use, and geology
- An understanding of the need for criteria with which to compare data
- A sense of empowerment and an eagerness to move on to analyze the real data



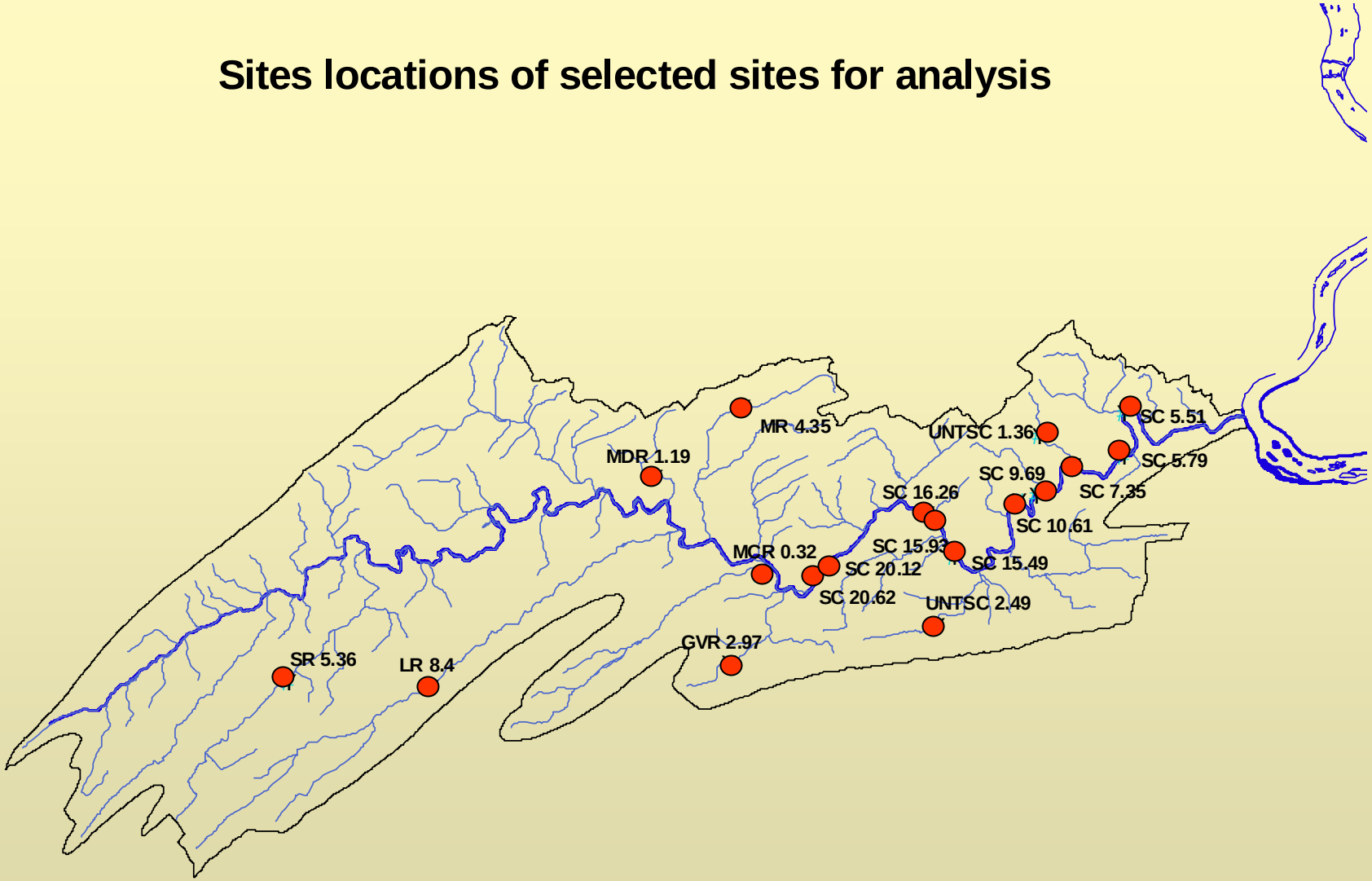
Workshop #2. Applying new skills to find the story in the real watershed data



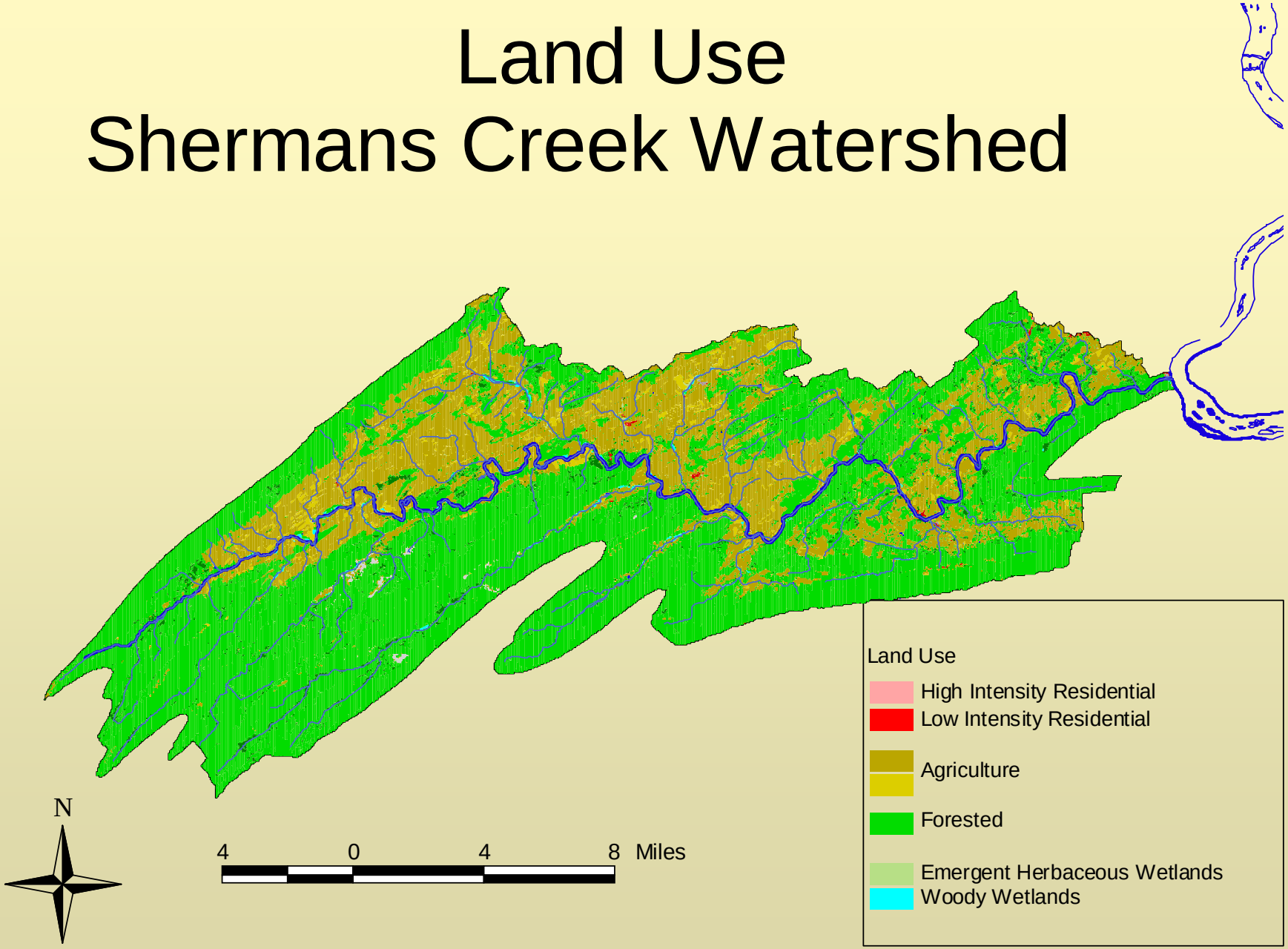
Indicators for which data have been collected by SCCA

- DO (dissolved oxygen)
- Temperature
- Using DO and Temperature, we can calculate Percent Saturation of DO
- Nitrate
- Alkalinity
- pH
- Macroinvertebrates (7 sites)
- Bacteria (4 sites)
- Visual Assessments

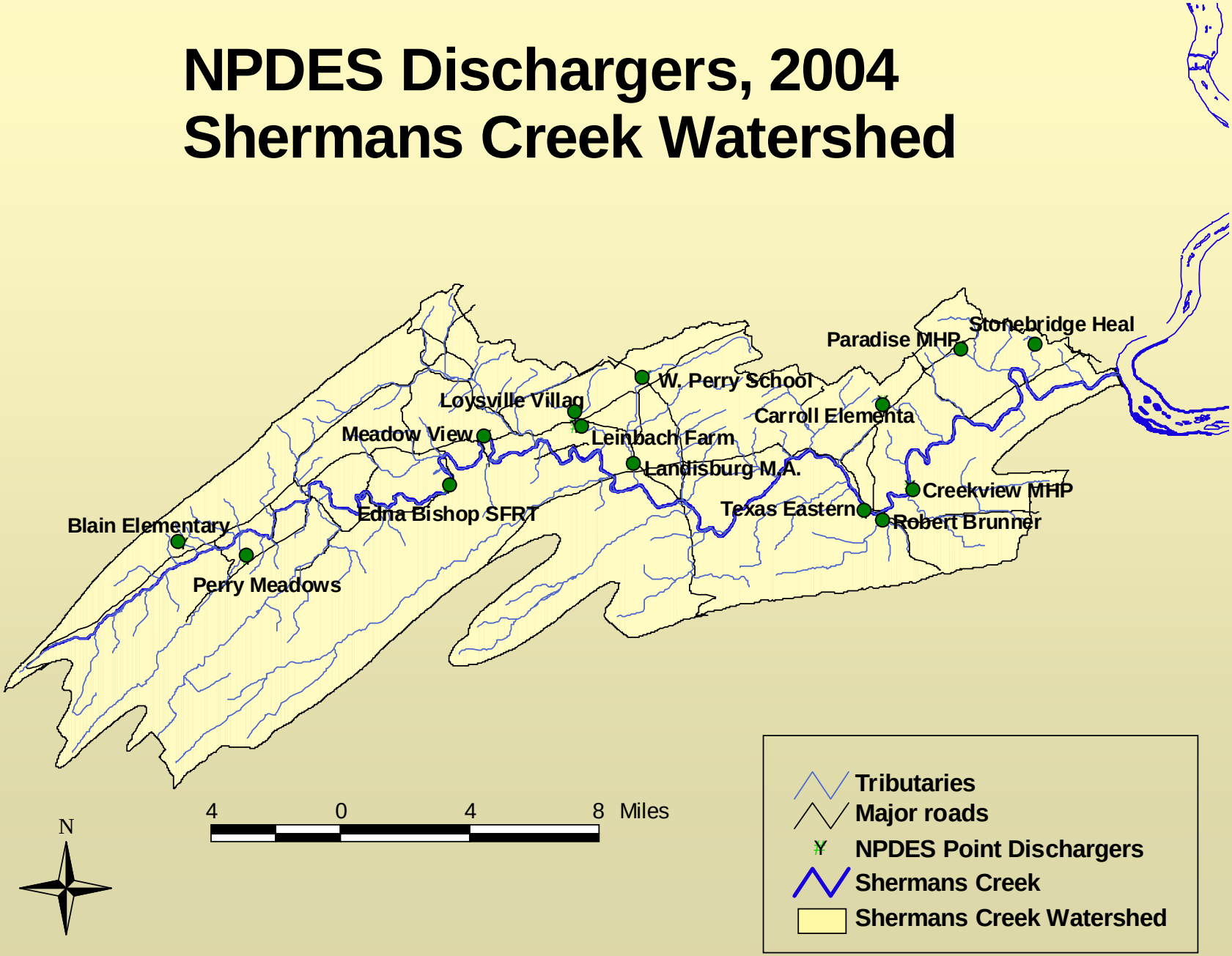
Sites locations of selected sites for analysis



Land Use
Shermans Creek Watershed



NPDES Dischargers, 2004
Shermans Creek Watershed



NITRATE CONCENTRATION DATA TABLES

NITRATES (mg/L)

SITE	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
SC 10.61		1.00	0.70	0.70	0.50	0.70	0.62	0.52			0.20	
SC 15.93	0.7	1	1.00	0.90	0.96	0.58		0.42	0.39	0.40	0.27	0.46
SC 16.26	0.64	0.89	0.91	0.88	0.86	0.80	0.76	0.68	0.38		0.36	0.37
SC 20.62		0.9	0.93	0.88	0.94		0.86	0.57	0.43			0.36
SC 5.51							0.63	0.37	0.08	0.10	0.00	0.31
SC 9.69					0.58		0.64	0.36	0.12	0.00	0.03	0.06
SR 5.36					0.17	0.20	0.22	0.14	0.18	0.18	0.07	0.12

NITRATES (mg/L) - Annual Summary For Each Site

SITE	Average	Min	25th	Median	75th	Max	Range	IQ Range
SC 10.61	0.62	0.20	0.52	0.66	0.70	1.00	0.80	0.19
SC 15.93	0.64	0.27	0.41	0.58	0.93	1.00	0.73	0.52
SC 16.26	0.68	0.36	0.51	0.76	0.87	0.91	0.55	0.36
SC 20.62	0.73	0.36	0.54	0.87	0.91	0.94	0.58	0.37
SC 5.51	0.25	0.00	0.09	0.21	0.36	0.63	0.63	0.27
SC 9.69	0.26	0.00	0.05	0.12	0.47	0.64	0.64	0.43
SR 5.36	0.16	0.07	0.14	0.18	0.19	0.22	0.15	0.05

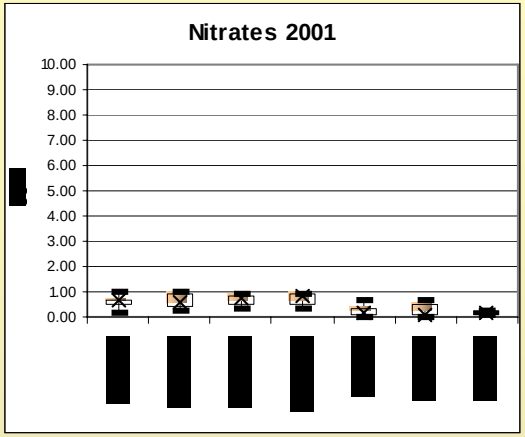
Raw Data, 2001

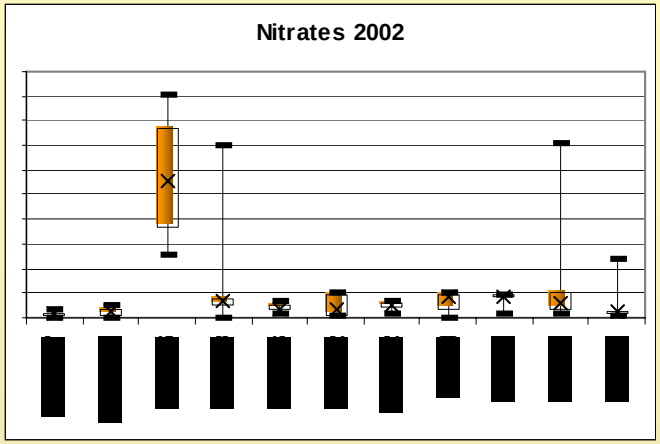
Statistical summaries, 2001

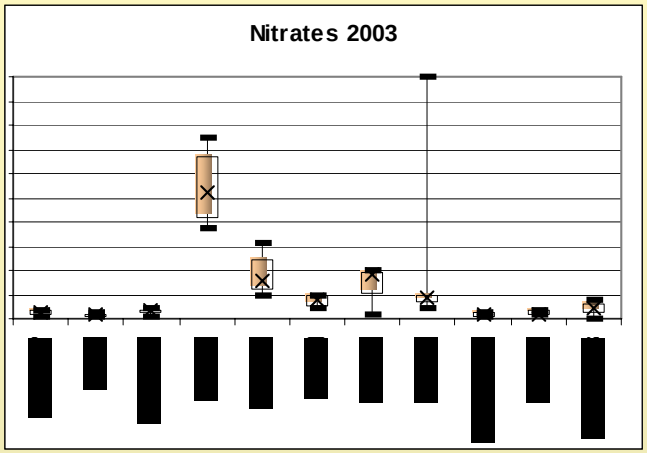
NITRATES (mg/L) - Monthly Summary for All Sites

MONTH	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
Average	0.67	0.95	0.89	0.84	0.67	0.57	0.62	0.44	0.26	0.17	0.16	0.28
Min	0.64	0.89	0.70	0.70	0.17	0.20	0.22	0.14	0.08	0.00	0.00	0.06
25th	0.66	0.90	0.86	0.84	0.52	0.49	0.62	0.37	0.14	0.08	0.04	0.17
Median	0.67	0.95	0.92	0.88	0.72	0.64	0.64	0.42	0.28	0.14	0.14	0.34
75th	0.69	1.00	0.95	0.89	0.92	0.73	0.73	0.55	0.39	0.24	0.25	0.37
Max	0.70	1.00	1.00	0.90	0.96	0.80	0.86	0.68	0.43	0.40	0.36	0.46

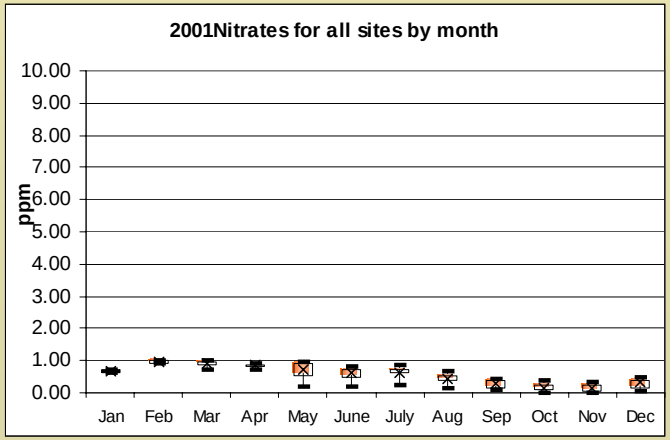
ANNUAL SUMMARY OF EACH SITE

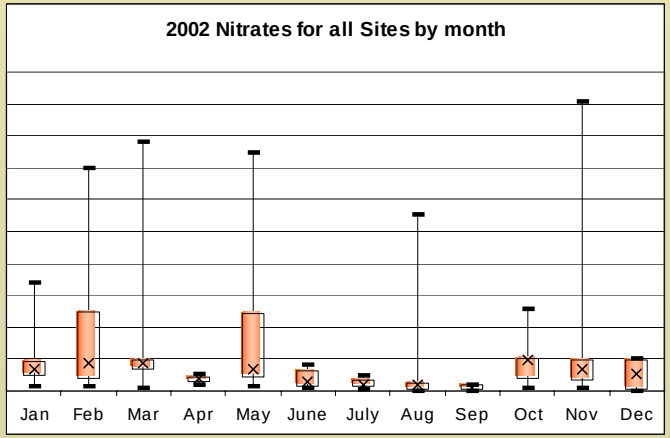


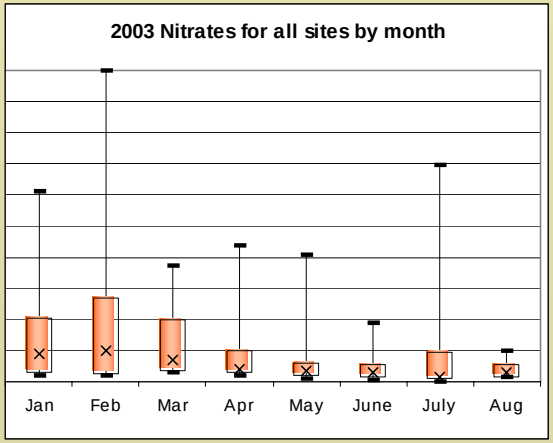




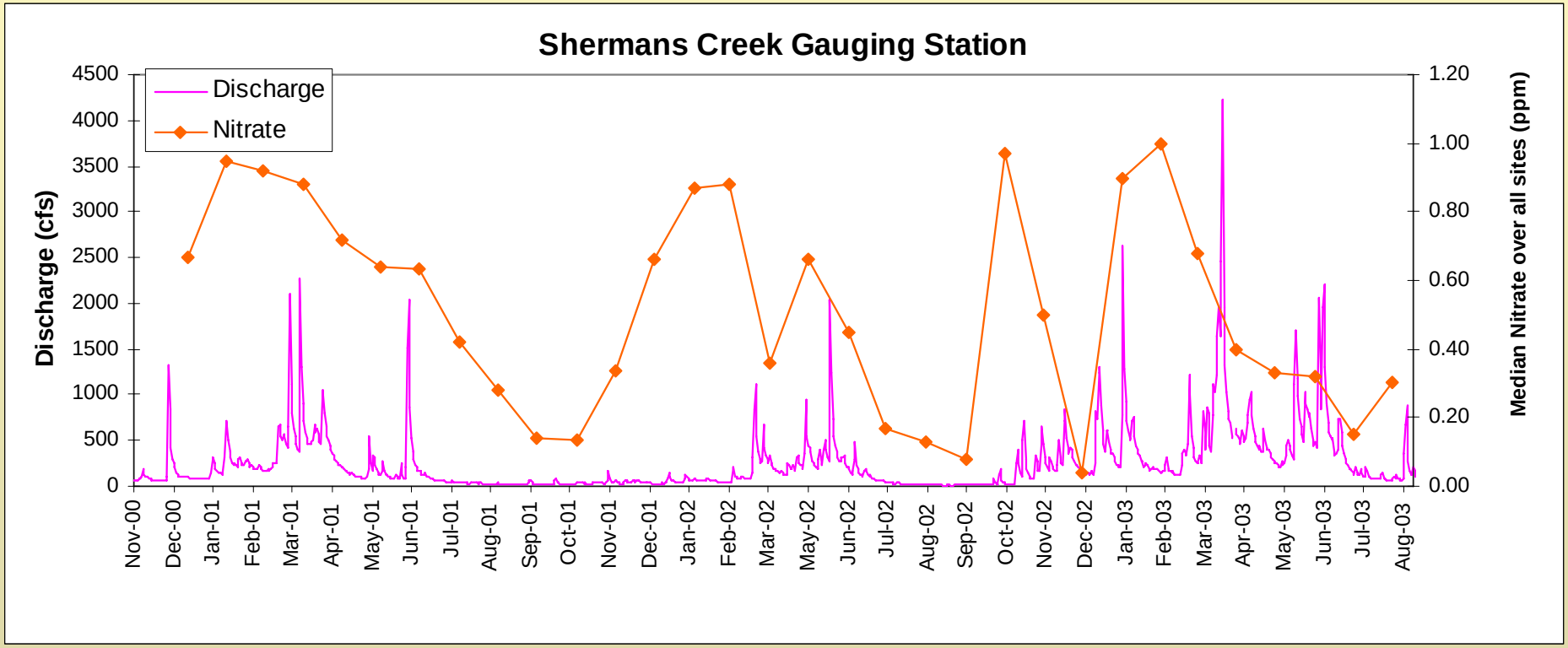
MONTHLY SUMMARIES OF ALL SITES



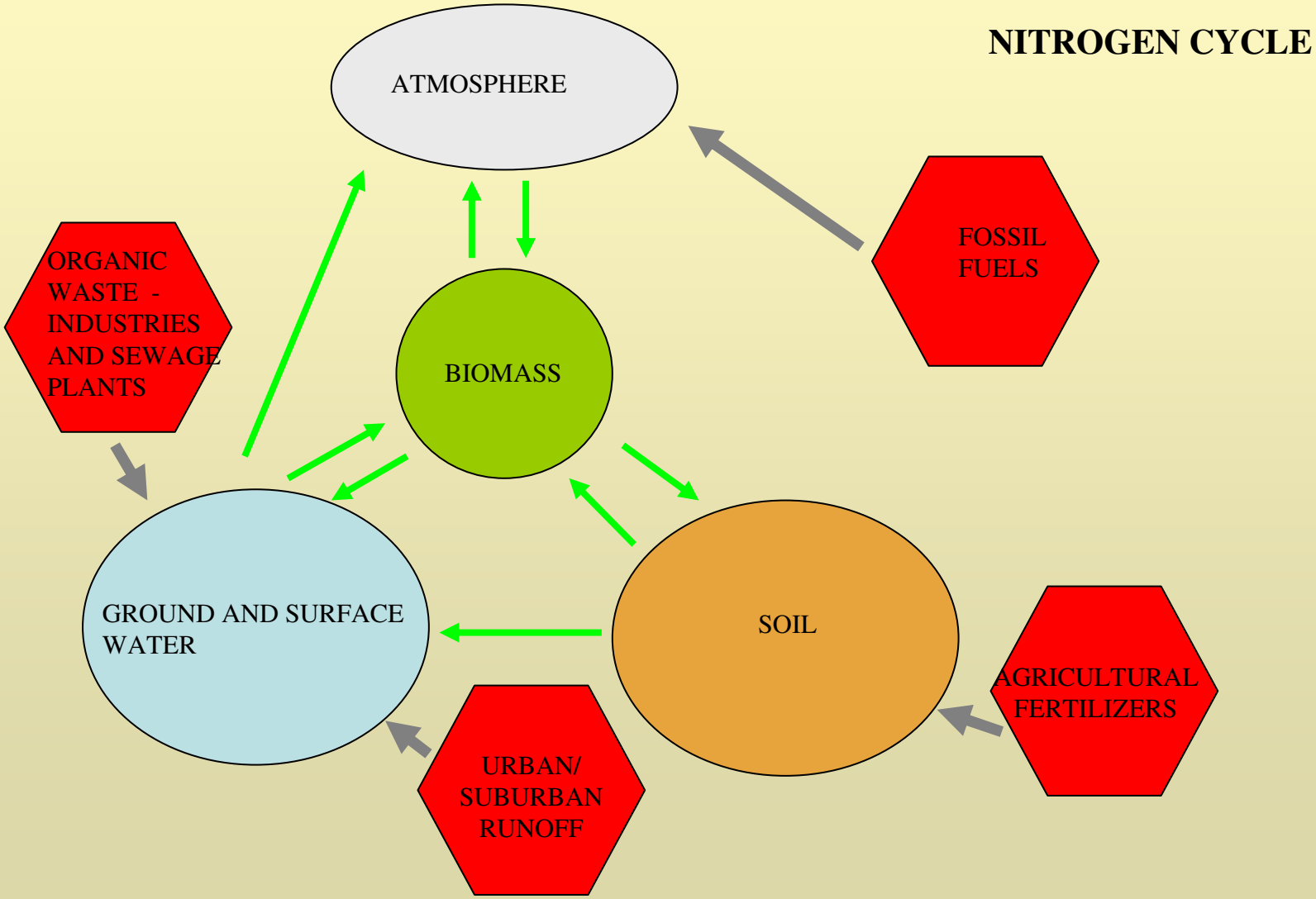




Relationship between flow and median monthly nitrate concentrations



Background Information on Chemical Indicators



Specific Water Quality Criteria (all waters should meet these criteria)

Parameter	Criteria
Alkalinity	Minimum of 20 mg/l as $CaCO_3$, except where natural conditions are less. Where discharges are to waters with 20 mg/l or less alkalinity, the discharge should not further reduce the alkalinity of the receiving waters.
pH	From 6.0 to 9.0 inclusive
Aluminum	Maximum of 0.1 of the 96-hour LC_{50} for representative important species as determined through substantial literature data or bioassay tests tailored to the ambient quality of the receiving waters.
Total Suspended Solids	No state specific criteria
Nitrate	Maximum 10 mg/L as nitrogen
Flow	No state specific criteria
Total Phosphorus	No state specific criteria
Dissolved Oxygen (Warm Water Fishery)	Min. Daily average 5.0 mg/l; min. 4.0 mg/l
Dissolved Oxygen (Cold Water Fishery)	Min. Daily Average 6.0 mg/l, minimum 5.0 mg/l

Note these criteria are based on the critical use of the watershed in our case study and your watershed may have different criteria. Refer to Title 25, Chapter 93 of the Pennsylvania Code for specific criteria for your watershed.

Excerpt from chart showing typical values of chemical parameters found in healthy streams

Percent saturation oxygen	Values of percent saturation of oxygen in healthy streams generally fall between 80% and 120%. Anything above or below that range is cause for concern.
pH	Although some streams are naturally acidic, anything below 6.0 (or above 8.5) is not considered typical.
Alkalinity	Low alkalinity values are of concern. Any stream with less than 20 mg/L (ppm) of alkalinity has little buffering capacity and is at risk for impact from acidic deposition.
Nitrates	Typical values in non-impacted streams are below 1 mg/L.

ETC.

The pieces: finding the story of OUR watershed

I. Groups use skills learned in first workshop to find patterns in their parameter and to identify problem sites.

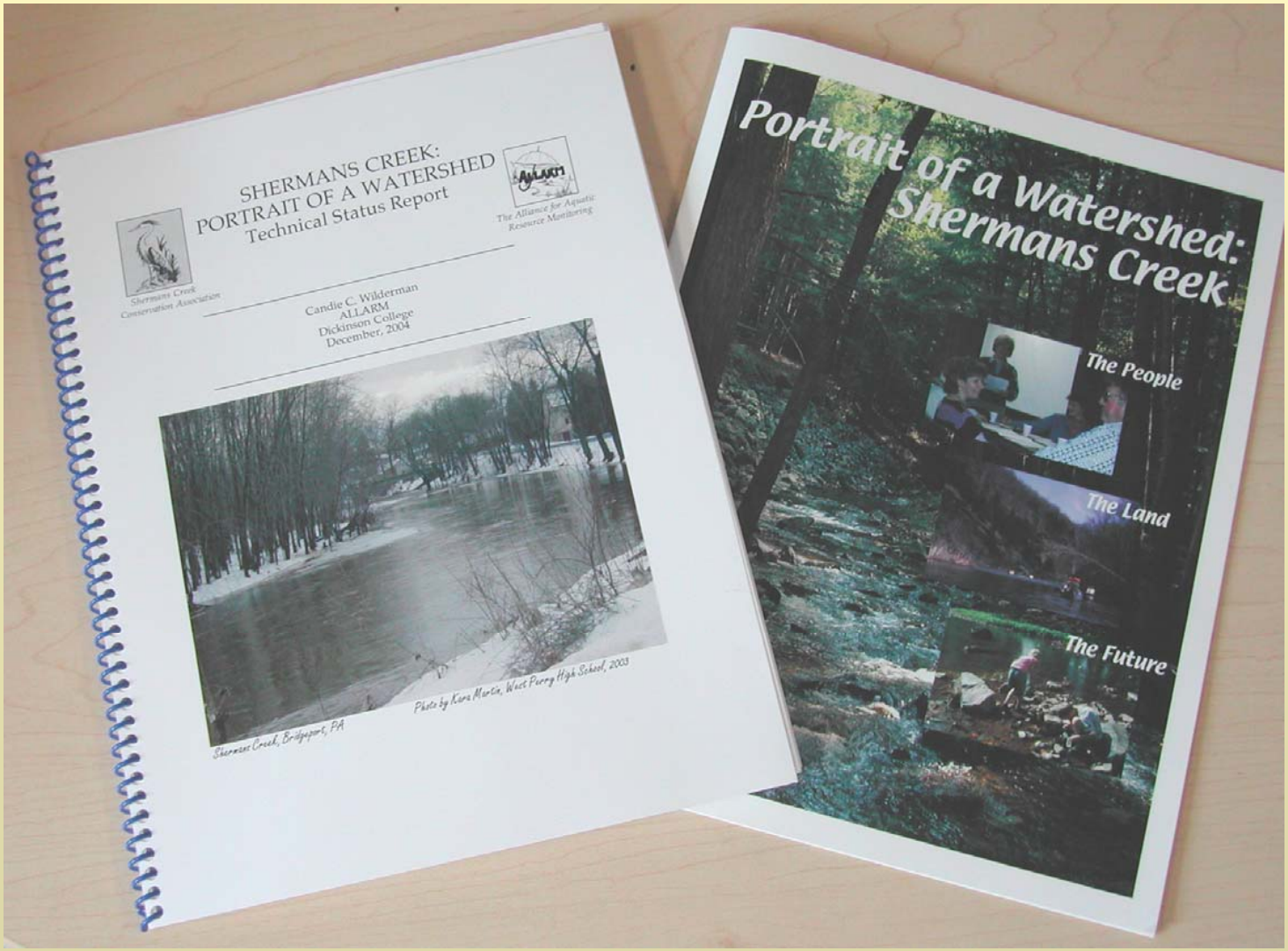
- Group is divided into small groups; each small group has **1-2 chemical indicators** to study

III. Together, we compile and record the big picture on a large summary table.

MASTER WORKSHEET TO COMPILE ALL INFORMATION GATHERED FROM ALL GROUPS (ALL INDICATORS)

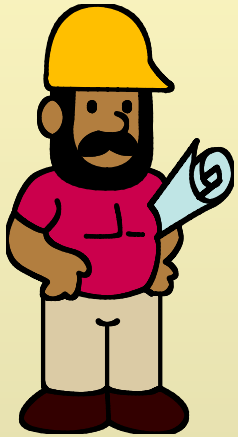
PROBLEM SITES	YEAR OF PROBLEM	INDICATORS OF PROBLEM	POSSIBLE CAUSE(S) OF PROBLEM
1.			
2.			
3.			
4.			
5.			

Examples of findings and final report



Typical watershed action plans based on volunteer data and carried out by volunteers

- Developing grant proposals for restoration projects
- Developing watershed fact sheets for public education
- Working with landowners to implement "best management practices"
- Developing conservation easement programs
- Upgrading stream protection status
- Removing dams
- Implementing stream and riparian zone restoration projects
- Using data to advocate for sound land use decisions by local municipalities



In the gathering of scientific knowledge, there is a trade-off between efficiency on the one hand and democracy and sustainability on the other hand.

Operational Model	Efficiency	Democracy "knowledge is power"	Sustainability
Consulting or Community Workers	Immediate, measurable scientific results	Only experts can use the data; volunteers are dependent on them	Money runs out, consultants leave, activities end
Community-based Participatory Research	Requires time, patience, and commitment for complex training process	Volunteers can shape the interpretations based on their own knowledge and can use the data; levels the playing field in decision-making	Builds community capacity to continue even after experts and monies are gone

Remember: the story in the data belongs to those who understand it, and knowledge is power!

For more information: http.//www.dickinson.edu/allarm

- Wilderman, C.C., and J. Vastine, 2005. "Breaking the Code: Data Analysis Workshops," *The Volunteer Monitor,* 17(1), pp. 11-14.

- Wilderman, C.C., A. Barron and L. Imgrund, 2004, "From the Field: A service provider's experience with two operational models for community science," Community-Based Collaboratives Research Consortium Journal, http://www.cbcrc.org/journal.html.

- Wilderman, C.C., A. Barron and L. Imgrund, 2003, "The ALLARM program: growth, change, and lessons learned," *The Volunteer Monitor*, 15(1), pp.1-4.