

US EPA ARCHIVE DOCUMENT

National Lakes Assessment Data Viewer: Finding the Story in the Data

NEAEB Conference
March 19, 2009

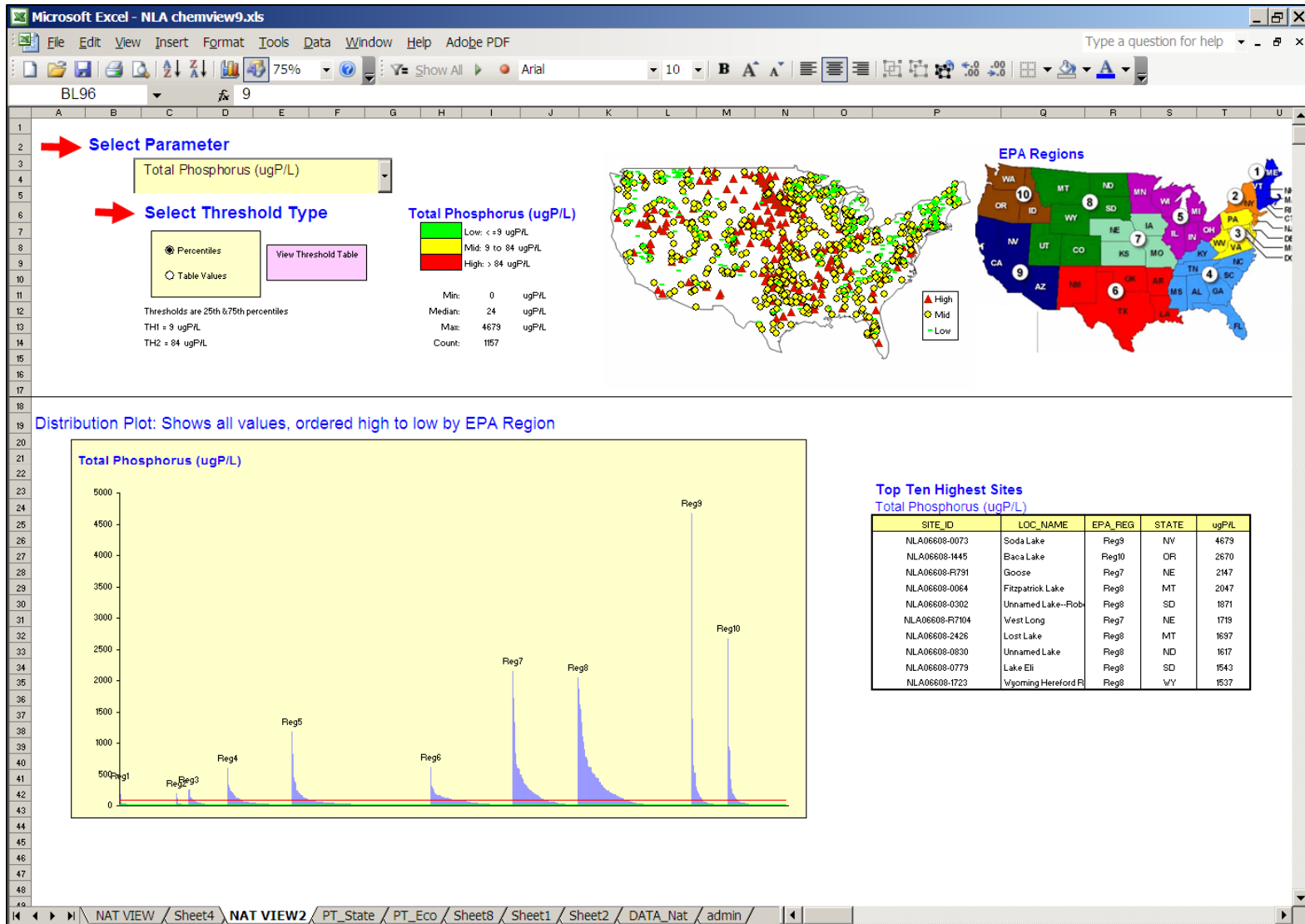
John Kiddon, USEPA Atlantic Ecology Division
Narragansett, RI

Interactive Excel spreadsheet:

Allows user to view and explore National Lakes Assessment (NLA) data

This version focuses on EPA Regions

others built for 9 WSA Ecoregions and for 50 States



All NLA data are loaded on spreadsheet

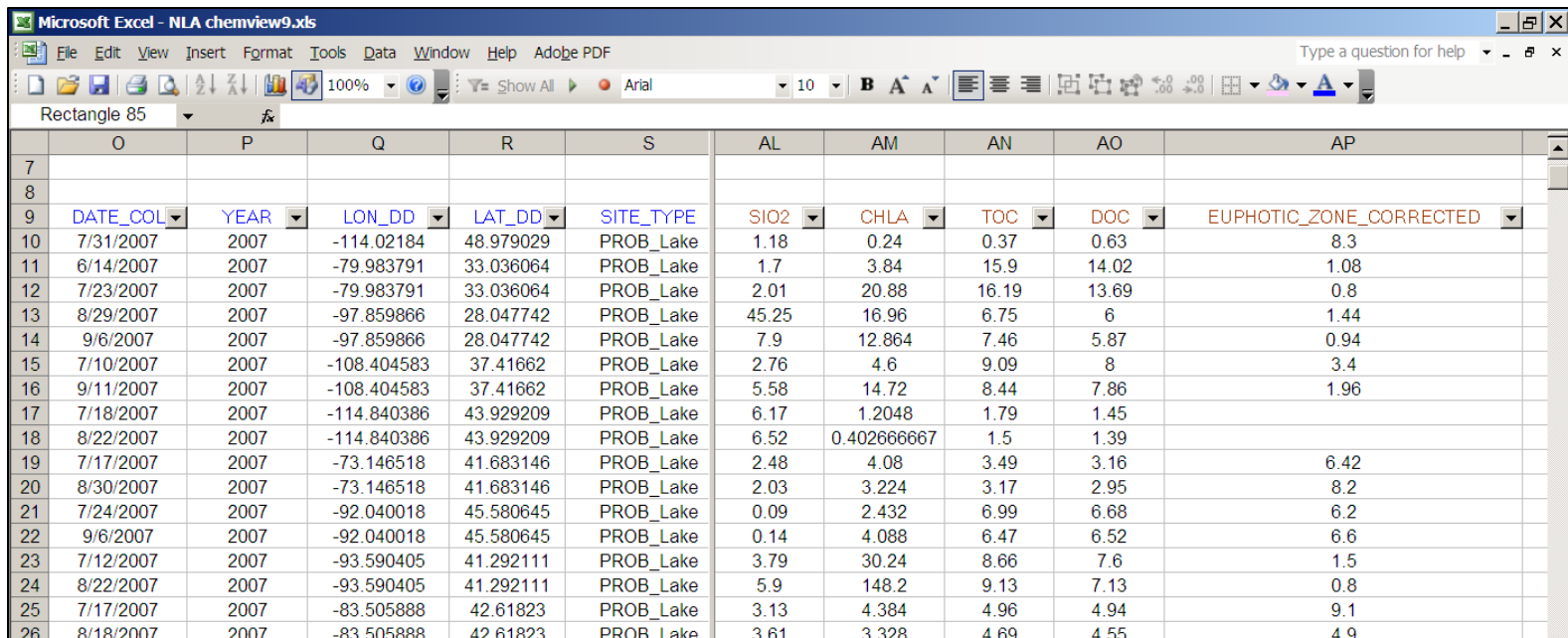
Pros:

Data are in “spreadsheet” format

Easy to view and extract data

New or derived NLA parameters can be added to data sheet (and viewed in graphics)

Other independent datasets (e.g., State data) may be added and viewed



	O	P	Q	R	S	AL	AM	AN	AO	AP
	DATE_COL	YEAR	LON_DD	LAT_DD	SITE_TYPE	SIO2	CHLA	TOC	DOC	EUPHOTIC_ZONE_CORRECTED
7	7/31/2007	2007	-114.02184	48.979029	PROB_Lake	1.18	0.24	0.37	0.63	8.3
8	6/14/2007	2007	-79.983791	33.036064	PROB_Lake	1.7	3.84	15.9	14.02	1.08
9	7/23/2007	2007	-79.983791	33.036064	PROB_Lake	2.01	20.88	16.19	13.69	0.8
10	8/29/2007	2007	-97.859866	28.047742	PROB_Lake	45.25	16.96	6.75	6	1.44
11	9/6/2007	2007	-97.859866	28.047742	PROB_Lake	7.9	12.864	7.46	5.87	0.94
12	7/10/2007	2007	-108.404583	37.41662	PROB_Lake	2.76	4.6	9.09	8	3.4
13	9/11/2007	2007	-108.404583	37.41662	PROB_Lake	5.58	14.72	8.44	7.86	1.96
14	7/18/2007	2007	-114.840386	43.929209	PROB_Lake	6.17	1.2048	1.79	1.45	
15	8/22/2007	2007	-114.840386	43.929209	PROB_Lake	6.52	0.402666667	1.5	1.39	
16	7/17/2007	2007	-73.146518	41.683146	PROB_Lake	2.48	4.08	3.49	3.16	6.42
17	8/30/2007	2007	-73.146518	41.683146	PROB_Lake	2.03	3.224	3.17	2.95	8.2
18	7/24/2007	2007	-92.040018	45.580645	PROB_Lake	0.09	2.432	6.99	6.68	6.2
19	9/6/2007	2007	-92.040018	45.580645	PROB_Lake	0.14	4.088	6.47	6.52	6.6
20	7/12/2007	2007	-93.590405	41.292111	PROB_Lake	3.79	30.24	8.66	7.6	1.5
21	8/22/2007	2007	-93.590405	41.292111	PROB_Lake	5.9	148.2	9.13	7.13	0.8
22	7/17/2007	2007	-83.505888	42.61823	PROB_Lake	3.13	4.384	4.96	4.94	9.1
23	8/18/2007	2007	-83.505888	42.61823	PROB_Lake	3.61	3.328	4.69	4.55	4.9

Cons:

Difficult to maintain “version control”

Spreadsheet not yet robust ; changes may result in unexpected operation

To operate spreadsheet:

User makes two selections ...

- 1) Selection of parameter from dropdown menu
- 2) Type of Thresholds TH1 & TH2
either 25th & 75th percentiles OR user-specified values

Graphics are automatically refreshed (few seconds delay)

1) Select Parameter

Select Parameter

Total Phosphorus (ugP/L)

Select Threshold Type

☒ Percentiles
☐ Table Values

View Threshold Table

Thresholds are 25th & 75th percentiles
TH1 = 9 ugP/L
TH2 = 84 ugP/L

2) Select Threshold Type:
percentiles or user-
specified values

User can change custom threshold values in “pop-up” Table

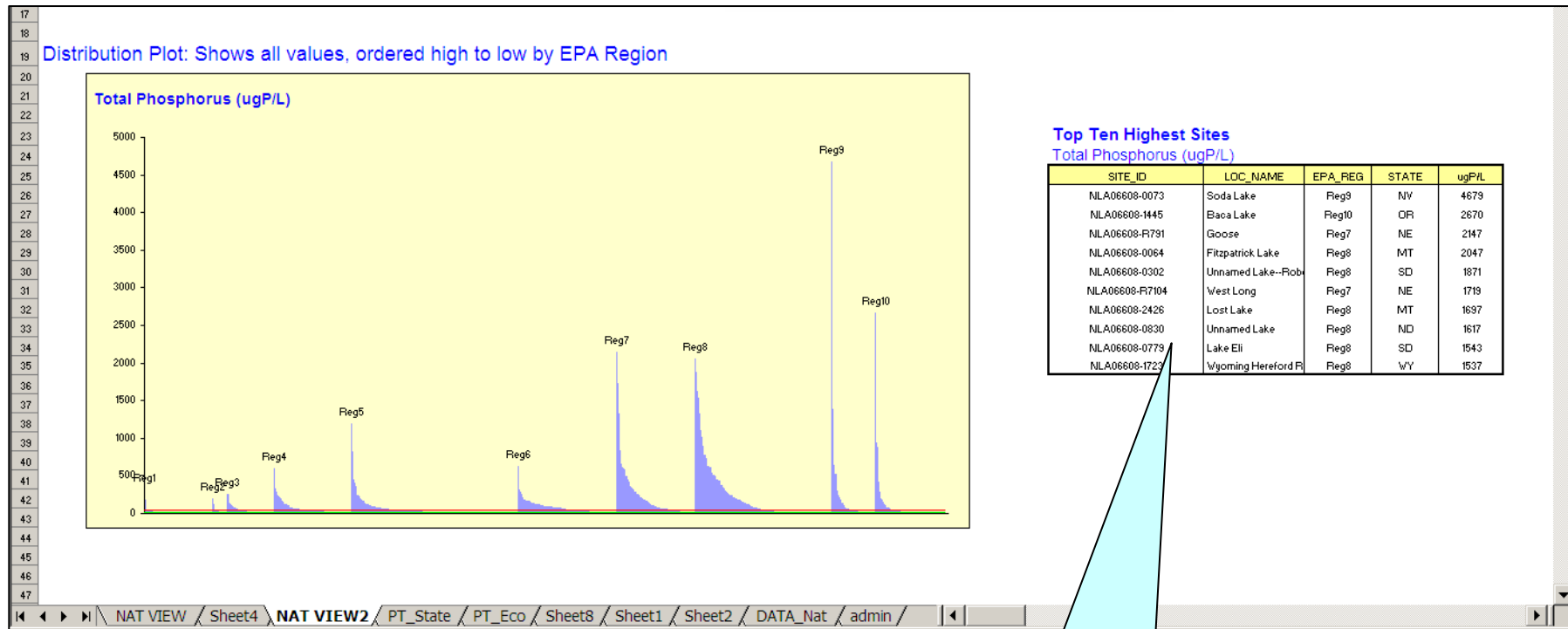
Note: more complex Threshold Tables are possible

Enter custom threshold values TH1 & TH2 here.
These thresholds will then be used to create graphics when “Thresh Type = Table Values” is selected from main page.

Parameter	TH1	TH2	Min	Max	Median
Conductivity (uS/cm @ 25 C)	100	600	0.61	36000	230
Acid Neutralizing Capacity (ueq/L)	500	3000	-63.0	91633	1714
Secchi depth (m)	1	3	0.04	36.7	1.4
pH units	6	8	0	10.3	8.2
Turbidity (NTU)	2	40	0.06	574	3.7
Color (PCU)	4	10	0	165	12
Potassium (mg/L)	1	5	0.05	1412	2.1
Sodium (mg/L)	3	500	0.19	16740	7.1
Calcium (mg/L)	2	50	0.21	486	21.1
Magnesium (mg/L)	4	15	0.06	2466	6.4
Chloride (mg/L)	5	30	0.05	15814	8.3
Sulfate (mg SO4/L)	5	70	0	40086	9.0
Ammonium (mg N/L)	0.01	0.05	0	2.6	0.02
Nitrate (mg N/L)	0.01	0.1	0	5.4	0
Nitrate + Nitrite (mg N/L)	0.01	0.1	0	6.2	0
Total Nitrogen (mg N/L)	0.05	0.2	0.01	26.1	0.57
Total Nitrogen (ugN/L)	50	200	10	26100	571
Total Phosphorus (ugP/L)	5	50	0	4865	25
Silica (mg SiO2/L)	1	10	0.01	91.9	5.6
Chlorophyll a (ug/L)	40	40	0.07	936	7.7
Total Organic Carbon (mg/L)	2	7	0.37	325	5.8
Dissolved Organic Carbon (mg/L)	2	8	0.34	291	5.4
Euphotic zone depth (m)	1	5	0.06	45	2.5
Depth at sampling site (m)	1	10	0.5	97	5.9
Lake area (hectares) from NHD	40	80	0	167490	68
Estimated Mean Depth (m)	1	10	0.2	32.3	2.0
Estimated volume (m3)	100000	20000	19987	2E+10	1244511

Distribution plots

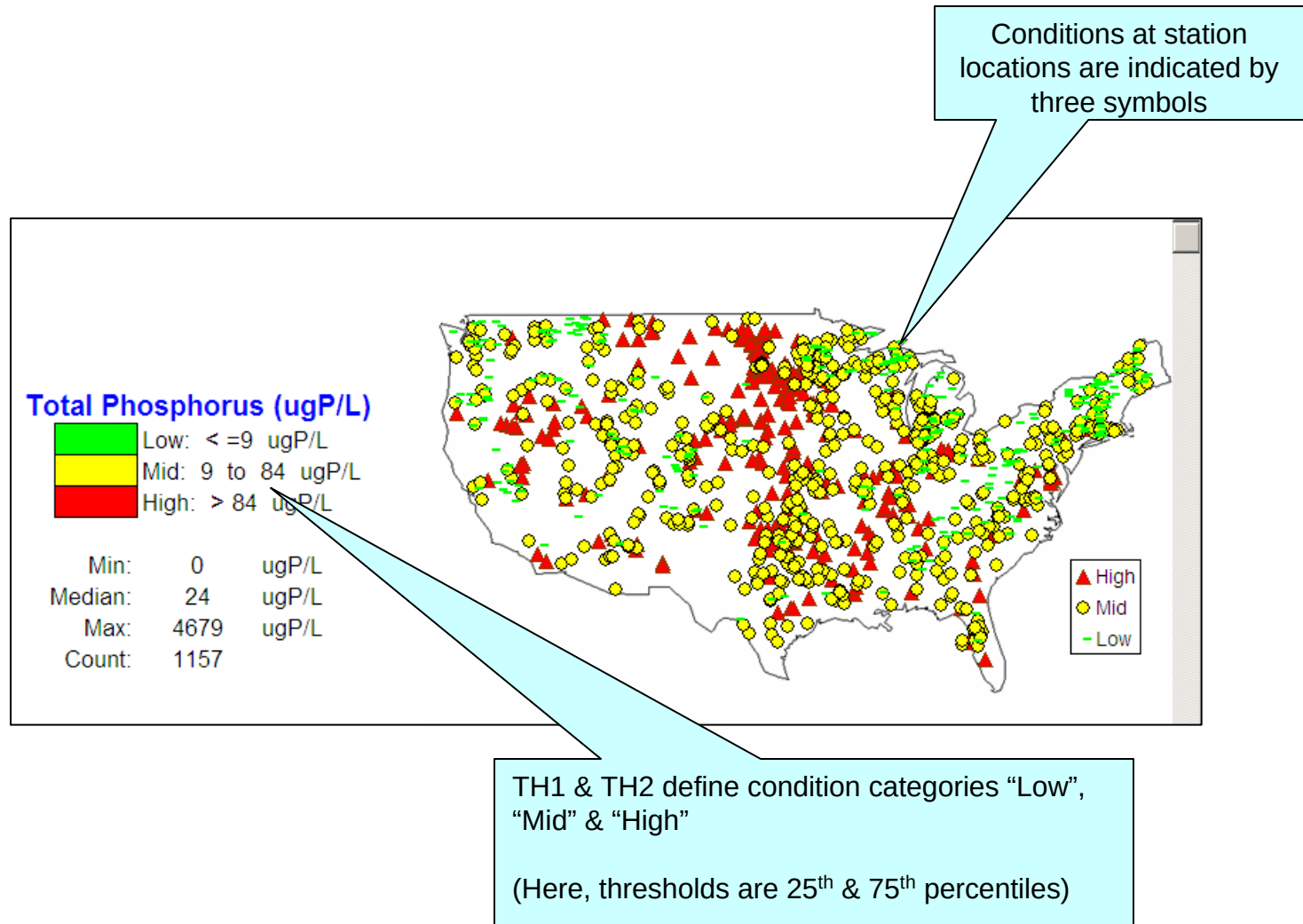
Displays all values in database, organized high to low by EPA Region
Displays range of data values and “shape” of distribution
Provides regional context



Top Ten Sites with highest values are identified

GIS on the cheap...

Selected Thresholds define 3 categories: “Low”, “Mid” & “High”
Station locations and conditions are indicated on map



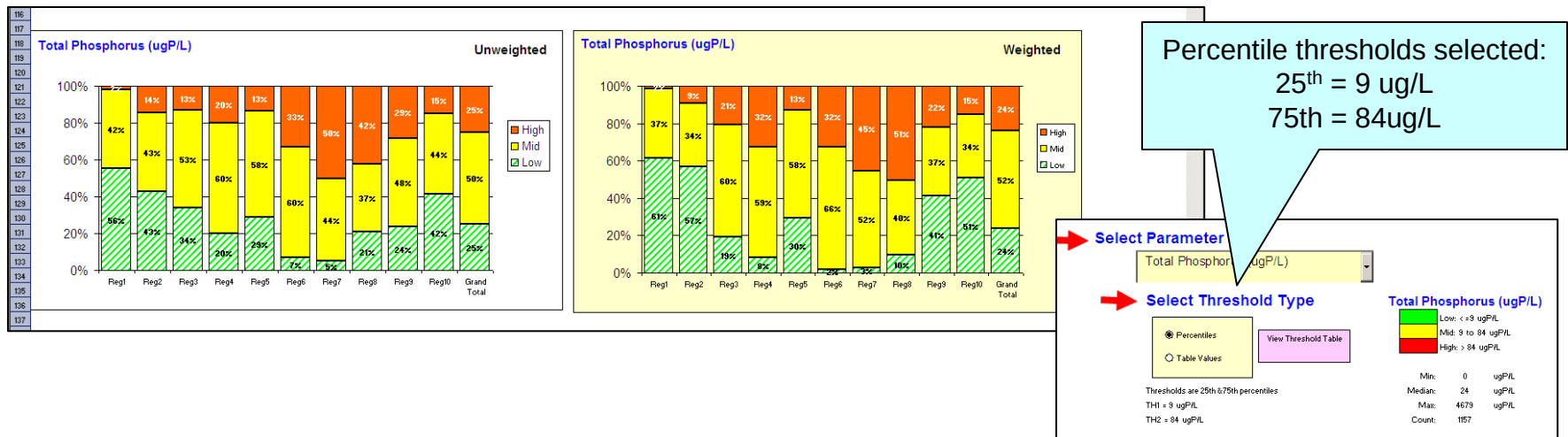
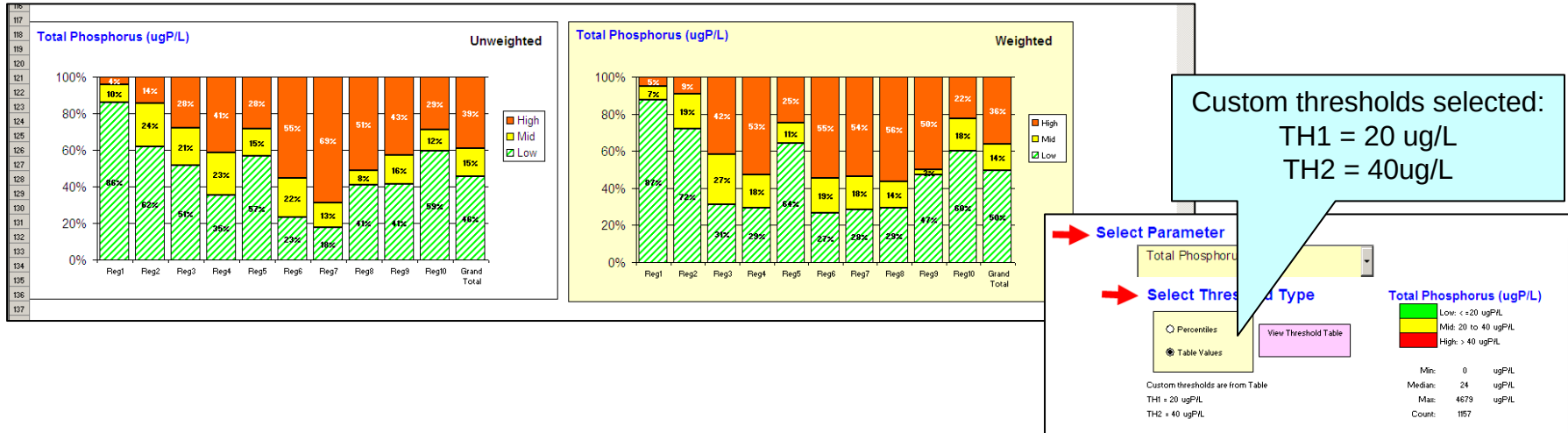
Percentage Plots

Indicates fraction of sites or resource in High, Mid, or Low categories

Unweighted plots reflect measured sites only

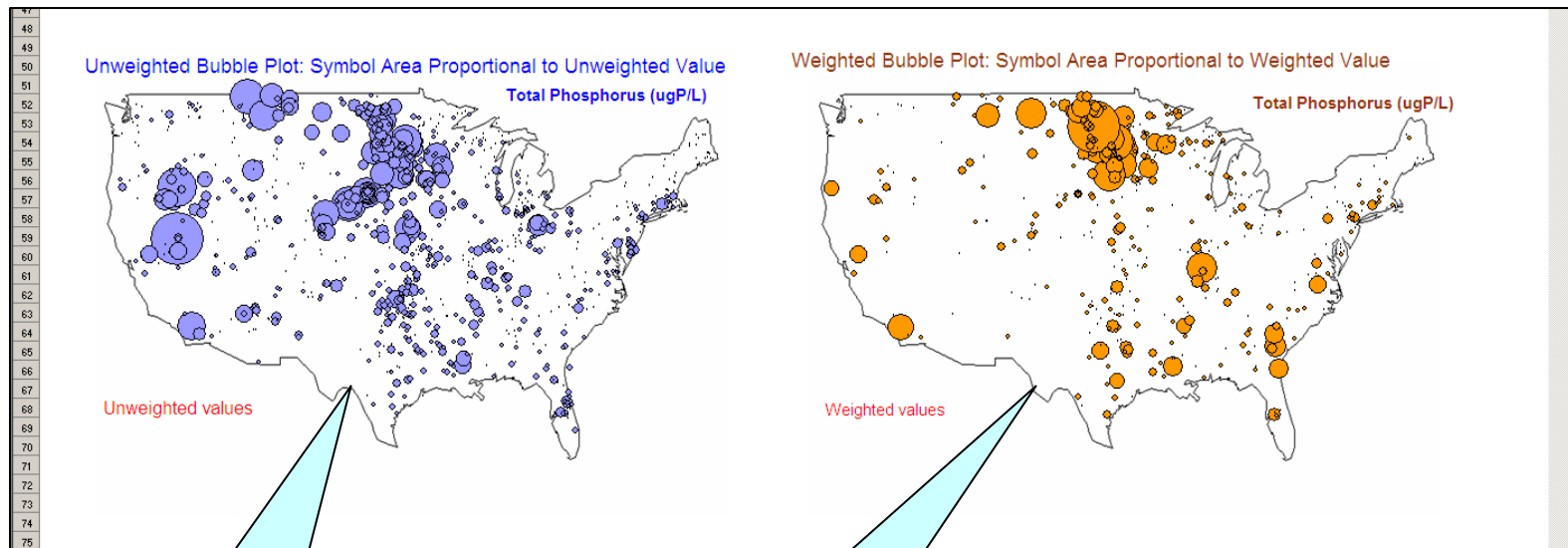
Weighted plots represent all lakes in nation

Note: examples here show effect of changing thresholds TH1 & TH2



Bubble plots

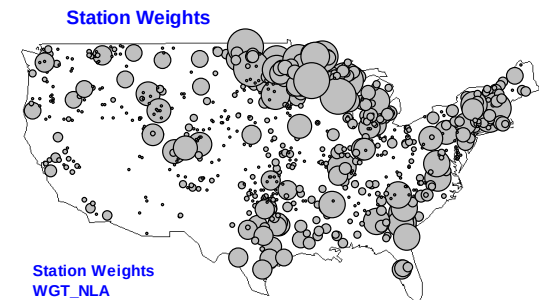
Symbol area is proportional to station value
Separate plots show weighted and unweighted values



Unweighted plot indicates values as measured

Weighted plot indicates how values are used in weighted calculations

This plots shows the station weights themselves. Heavily weighted sites represent a larger population than lightly weighted sites.

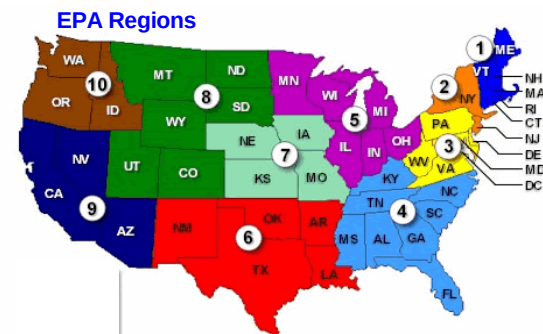


Basic stats visualized

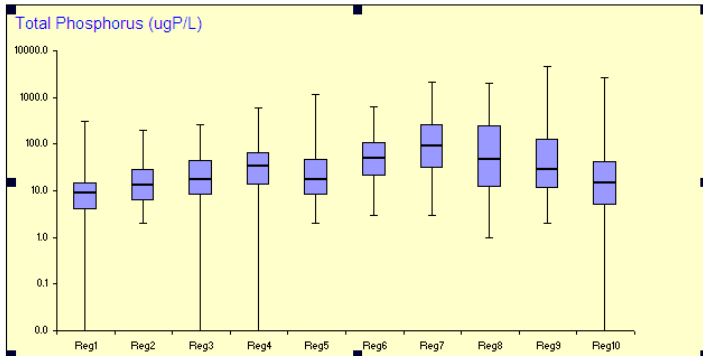
Bar & Whisker plots

Averages & Std Deviations

Weighed and Unweighed averages



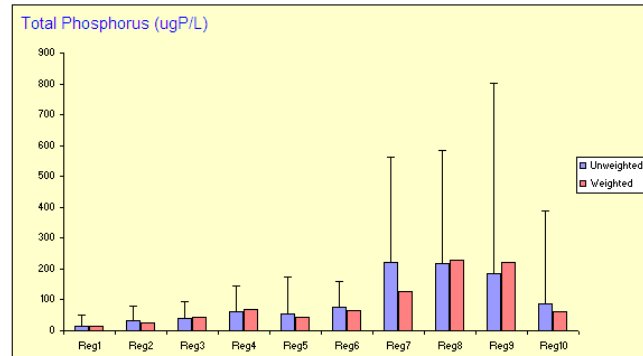
Box & Wisker Plot:
Shows Min, Max, Median, 25th & 75th percentiles



Data for above plots

	Reg1	Reg2	Reg3	Reg4	Reg5	Reg6	Reg7	Reg8	Reg9	Reg10
count	99	21	68	111	240	143	113	198	63	101
min	0	2	0	0	2	3	3	1	2	0
25%ile	4	6	8	13	8	21	31	12	11	5
median	9	13	17.5	33	17	50	89	46.5	28	15
75%ile	14.5	28	44.5	64.5	46.5	106.5	255	251.75	127.5	43
max	315	197	255	590	1184	619	2147	2047	4679	2670

Unweighted Average & Std Deviation
Weighted Average

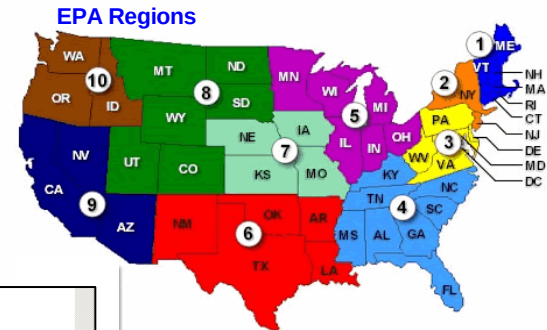


	Reg1	Reg2	Reg3	Reg4	Reg5	Reg6	Reg7	Reg8
unwgt avg	15.82	31.76	38.71	60.80	54.05	76.50	220.62	216.60
stdev	36.16	48.99	54.18	83.56	121.73	83.40	343.18	367.80
weighted avg	16.32	24.74	45.22	68.52	42.78	65.62	128.39	228.79

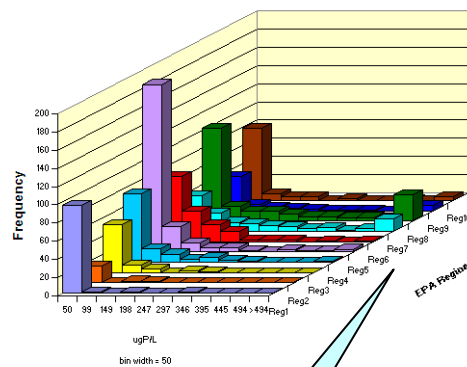
Histograms by region

Compares distributions among regions

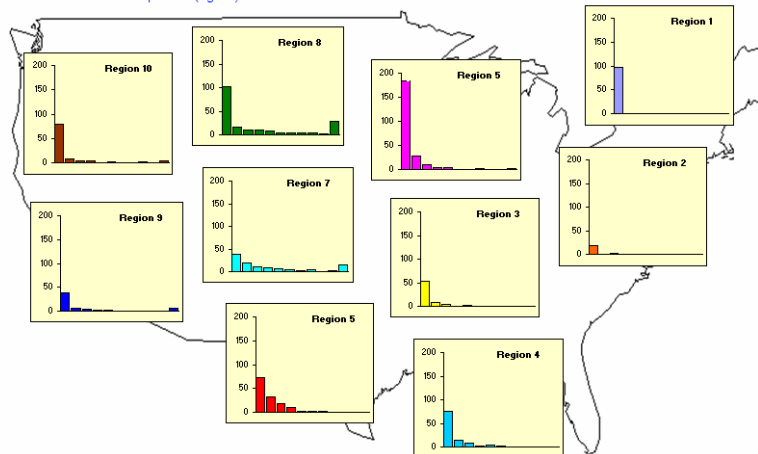
Highlights differences among parameters



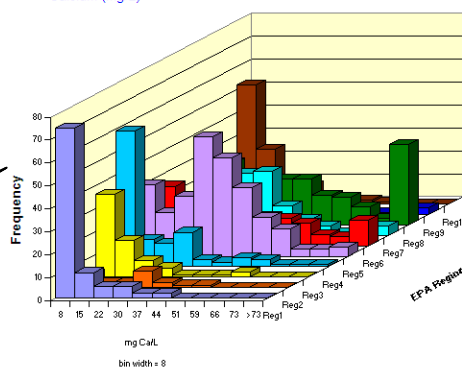
Total Phosphorus (ugP/L)



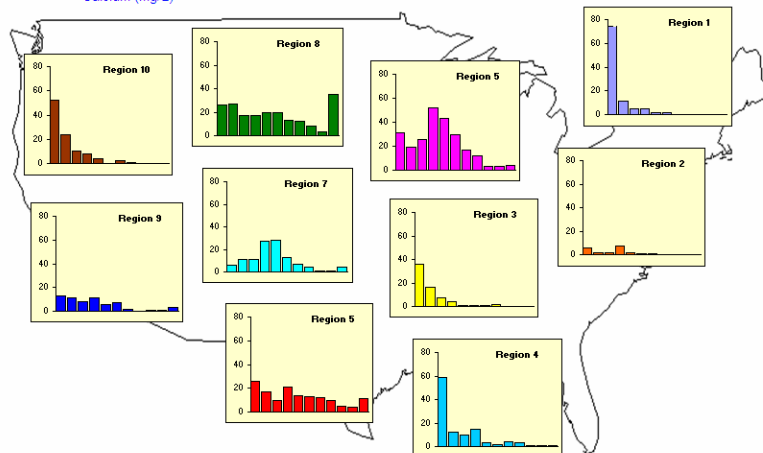
Total Phosphorus (ugP/L)



Calcium (mg/L)



Calcium (mg/L)

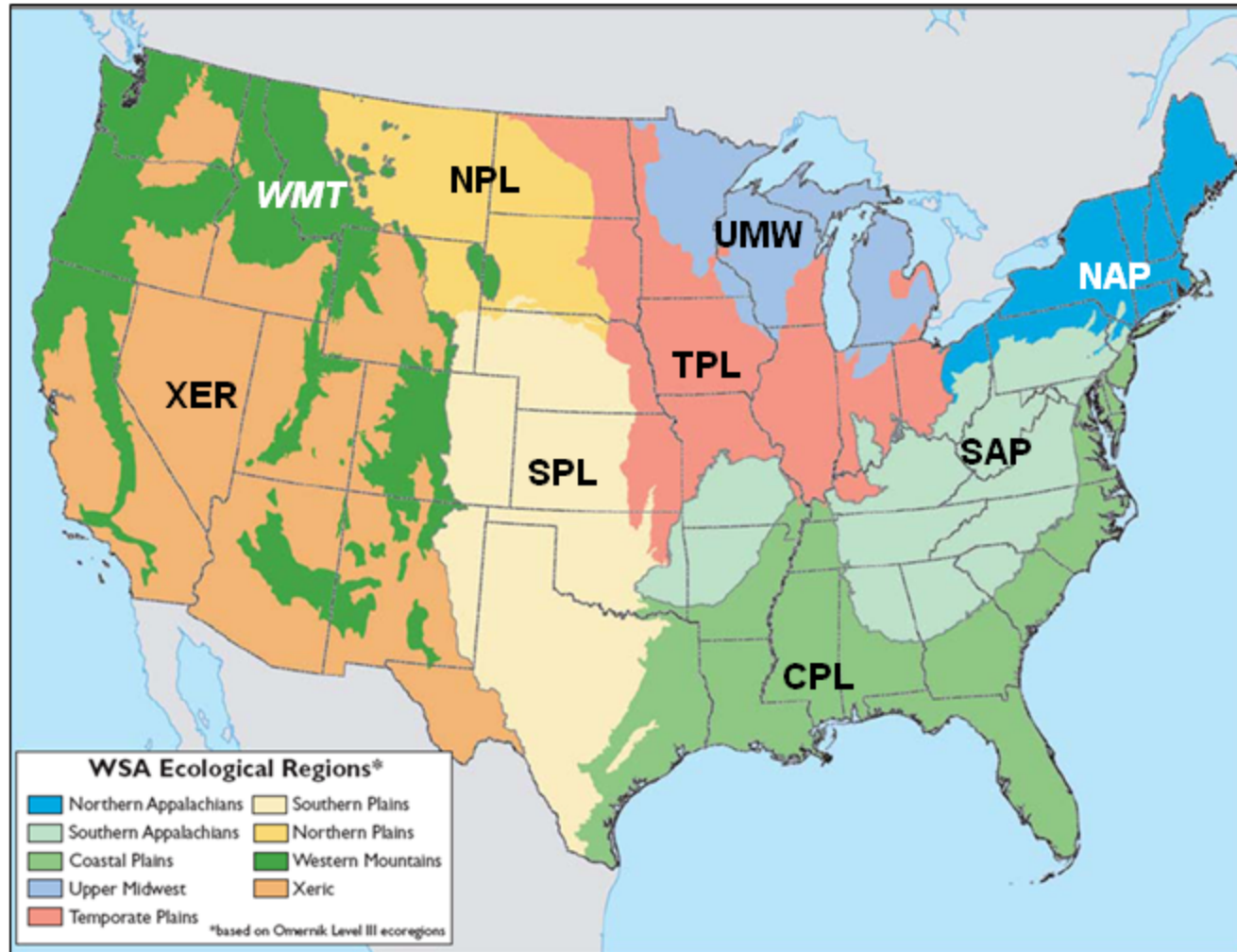


Total
Phosphorus
selected

Calcium
selected

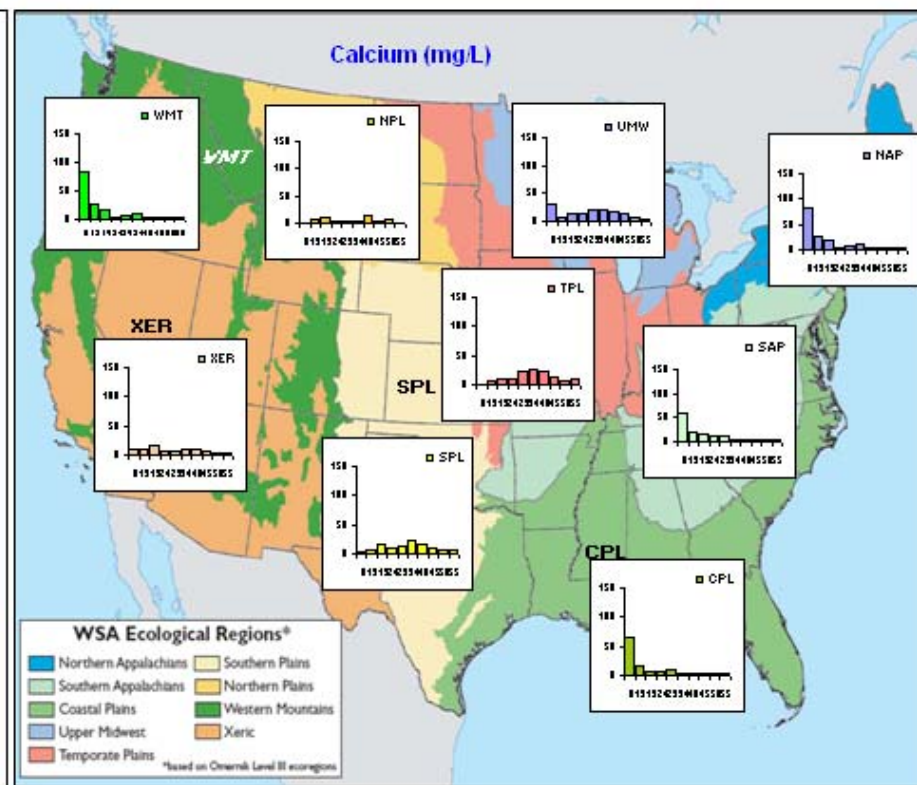
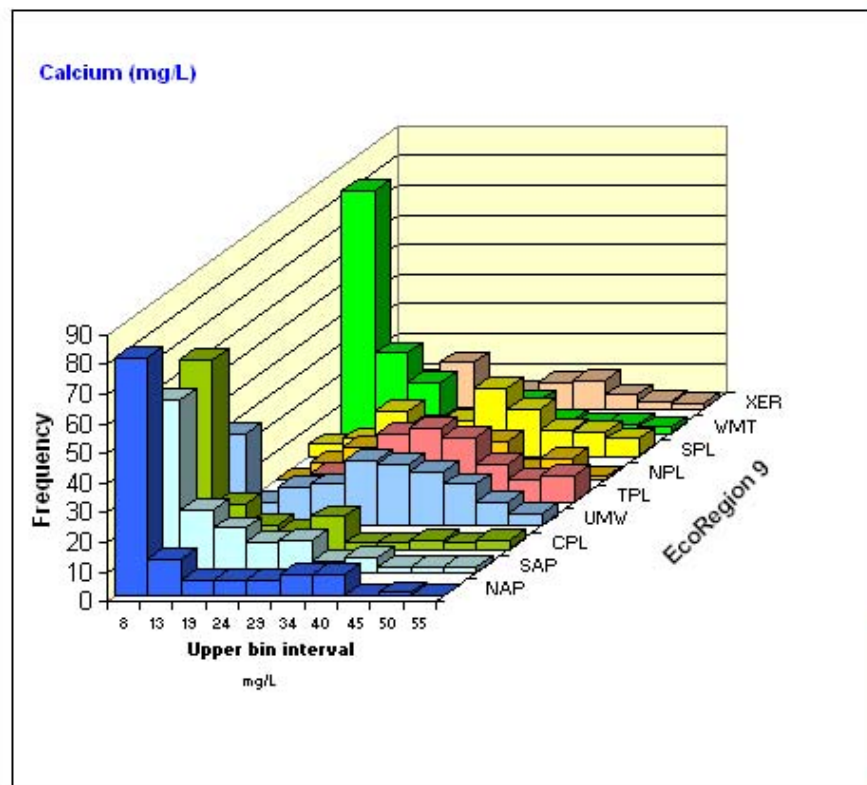
Similar Analysis based on Nine WSA Ecoregions

Wadeable Stream Assessment nine aggregated Omernik level 3 ecoregions



Histograms of Nine WSA Ecoregions

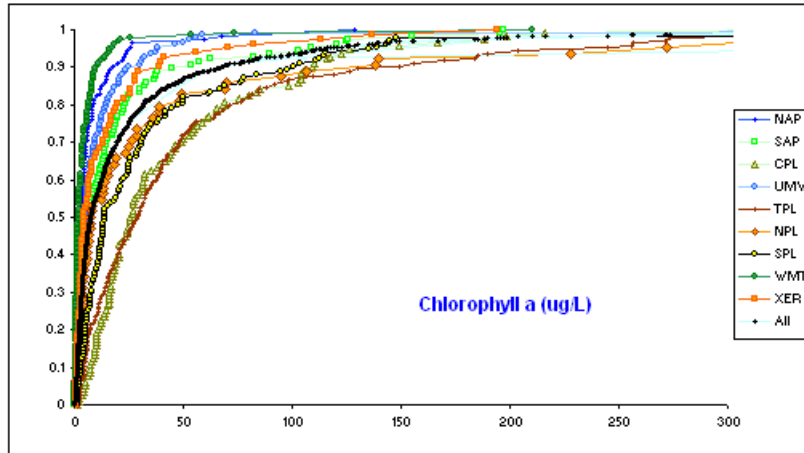
Wadeable Stream Assessment nine aggregated Omernik level 3 ecoregions



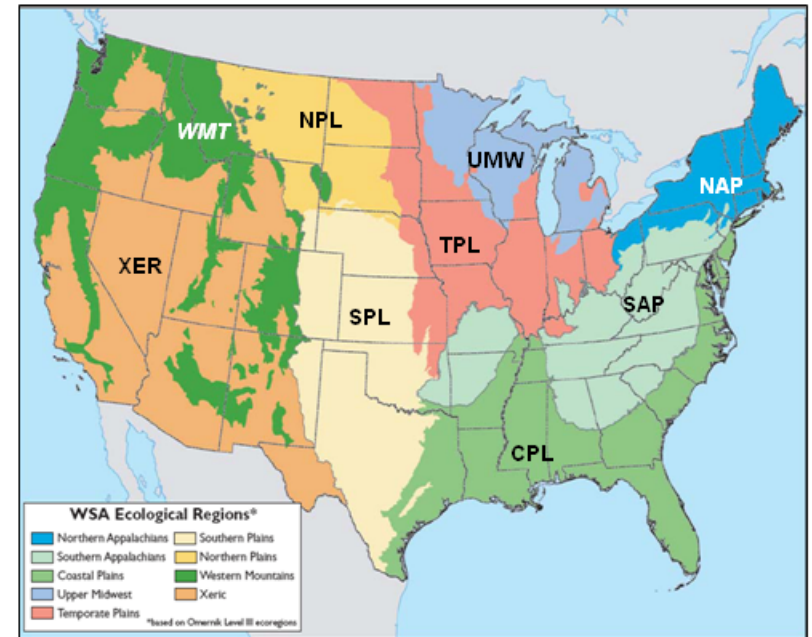
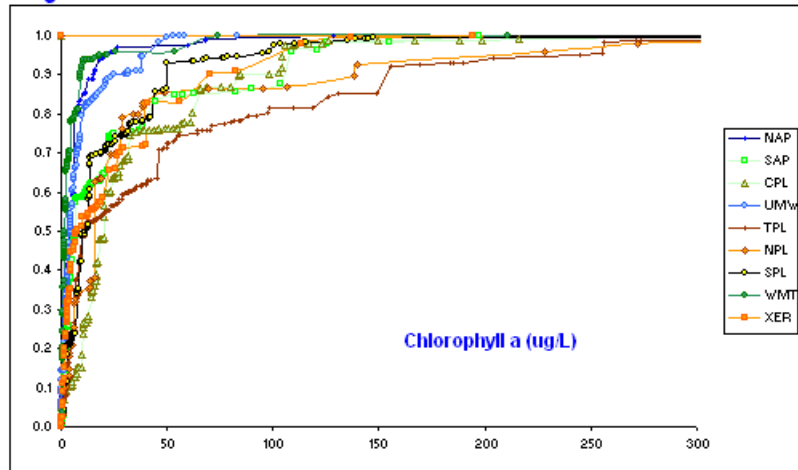
Continuous Distribution Functions (CDF)

Weighted and Unweighted CDFs

Unweighted CDF

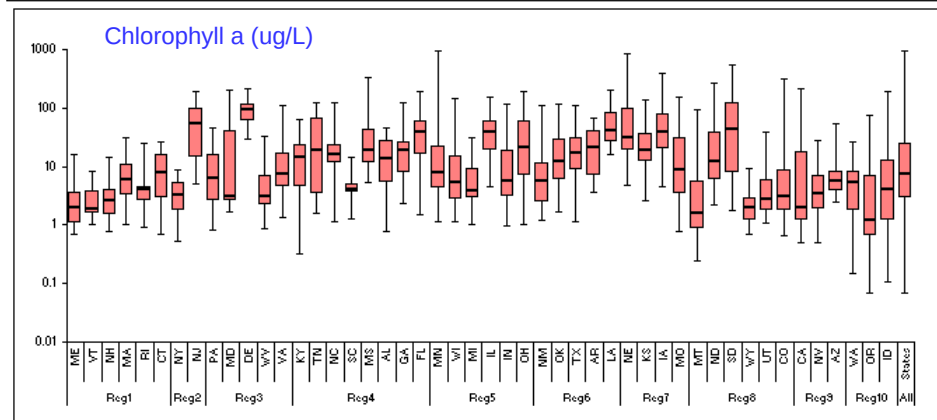
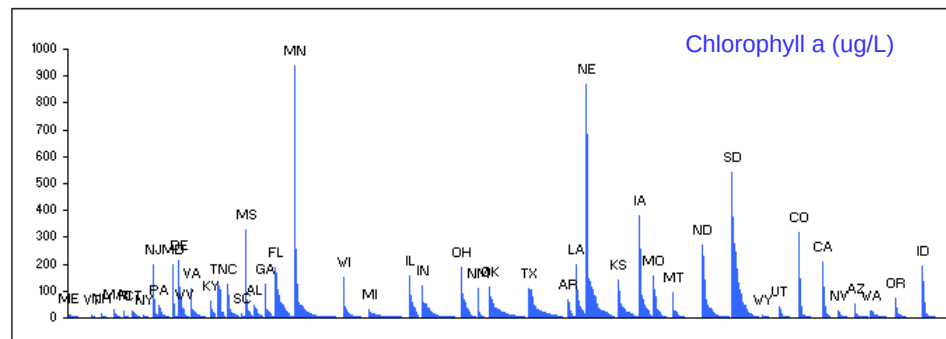
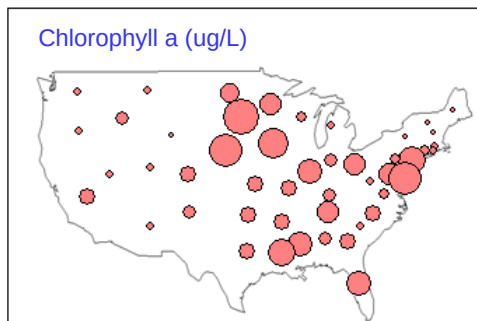


Weighted CDF



View and Analyze Data at the State and Local Level

Graphics comparing conditions among states



Close-up Regional View: New England States (Reg 1)

State condition in context of neighboring states
User can employ local or regional thresholds

Make Selections ←

Select Parameter: Chloride (mg/L) Select EPA Region: Region 1

☒ Show only first visits to a station

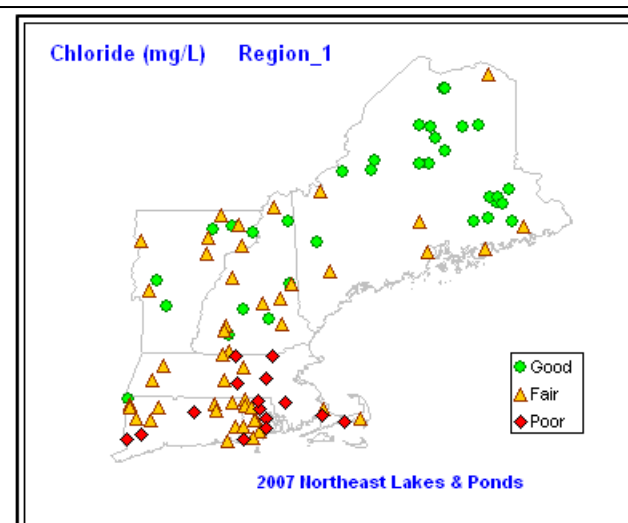
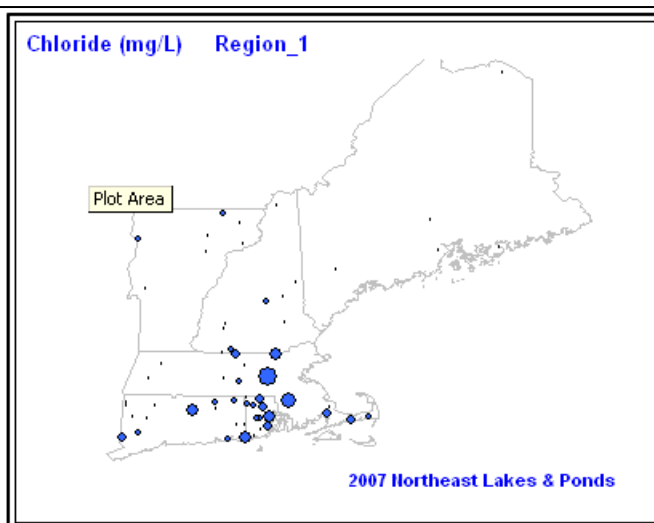
→ **Top Ten "worst sites"**

EPA_REG: Region 1

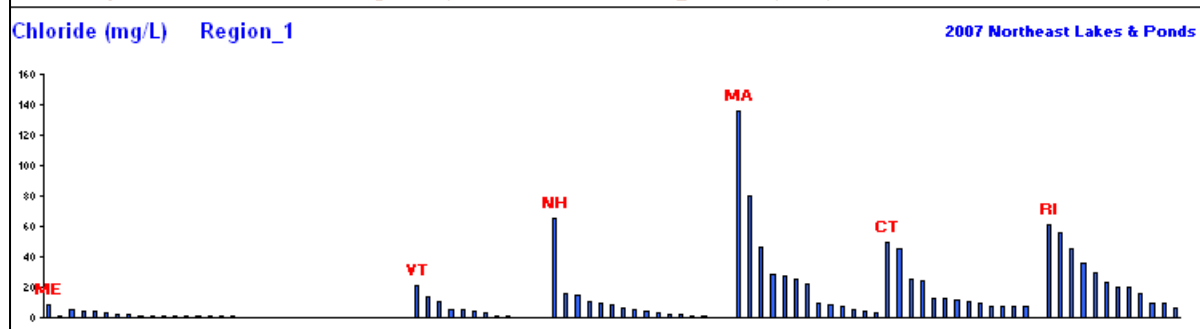
Check boxes to temporarily disable dominating values.

Check boxes of extreme values.	Max Value	LOC_NAME	ST	Total
☐		Gristmill Pond	MA	135.59
☐		Norton Reservoir	MA	80.42
☐		Long Pond	NH	64.877
☐			MA	27.529
		Gorton Pond	RI	60.98
		Chapman Pond	RI	55.323
		Union Pond	CT	49.711
		Dick's Pond	MA	45.81
		Lake Kenosia	CT	45.759
		Slack Reservoir	RI	45.447
		Slatterville Reservoir	RI	40.45

Press Here To Reset All Disabled Sites



Distribution plot: Presents values in decreasing order by state. Data are located in PT_DistBubble (below)



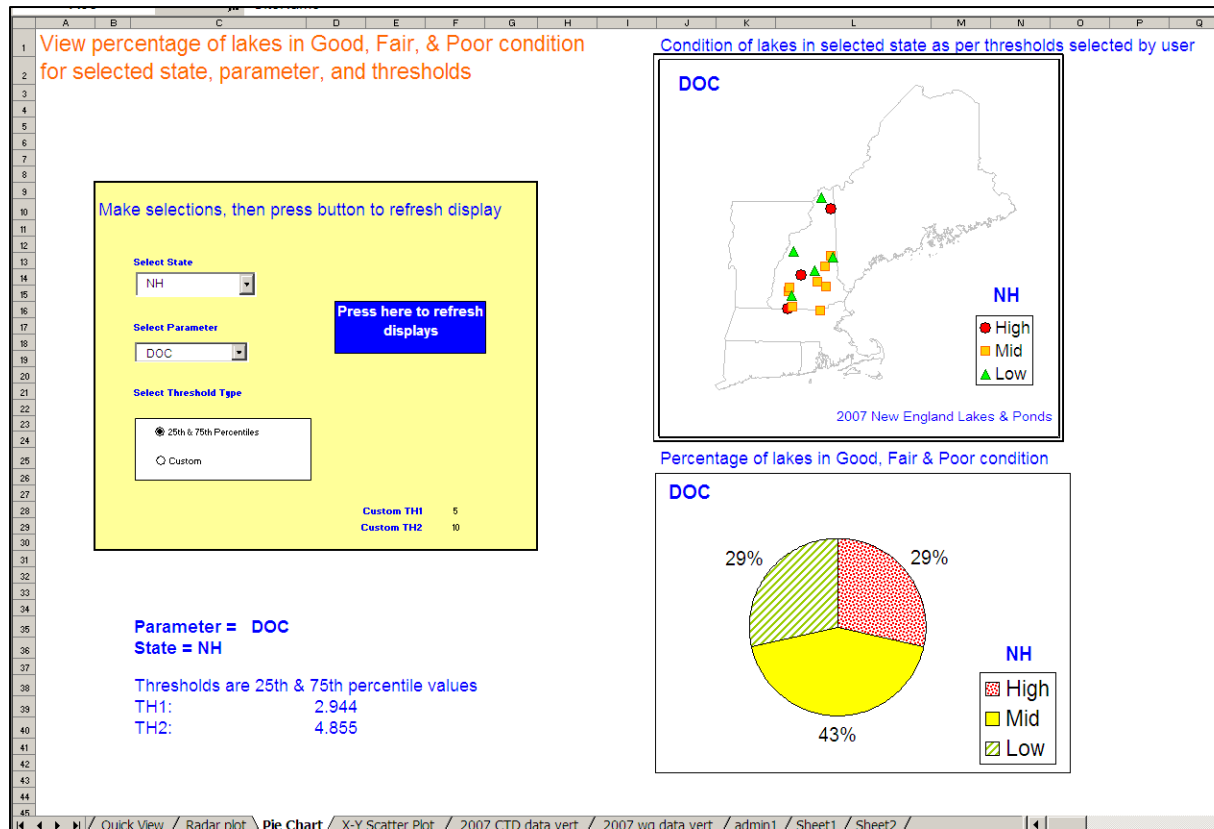
☒ Percentiles of National Data

☐ Custom Thresholds

Analyte CL_PPM
ChemName Chloride (mg/L)
Units mg CL/L
25%ile 2.238
75%ile 24.626
Custom T1 5
Custom T2 50

Close-up State View:

View probabilistic assessment data in one state
Calculate statistics that may be useful in state reports

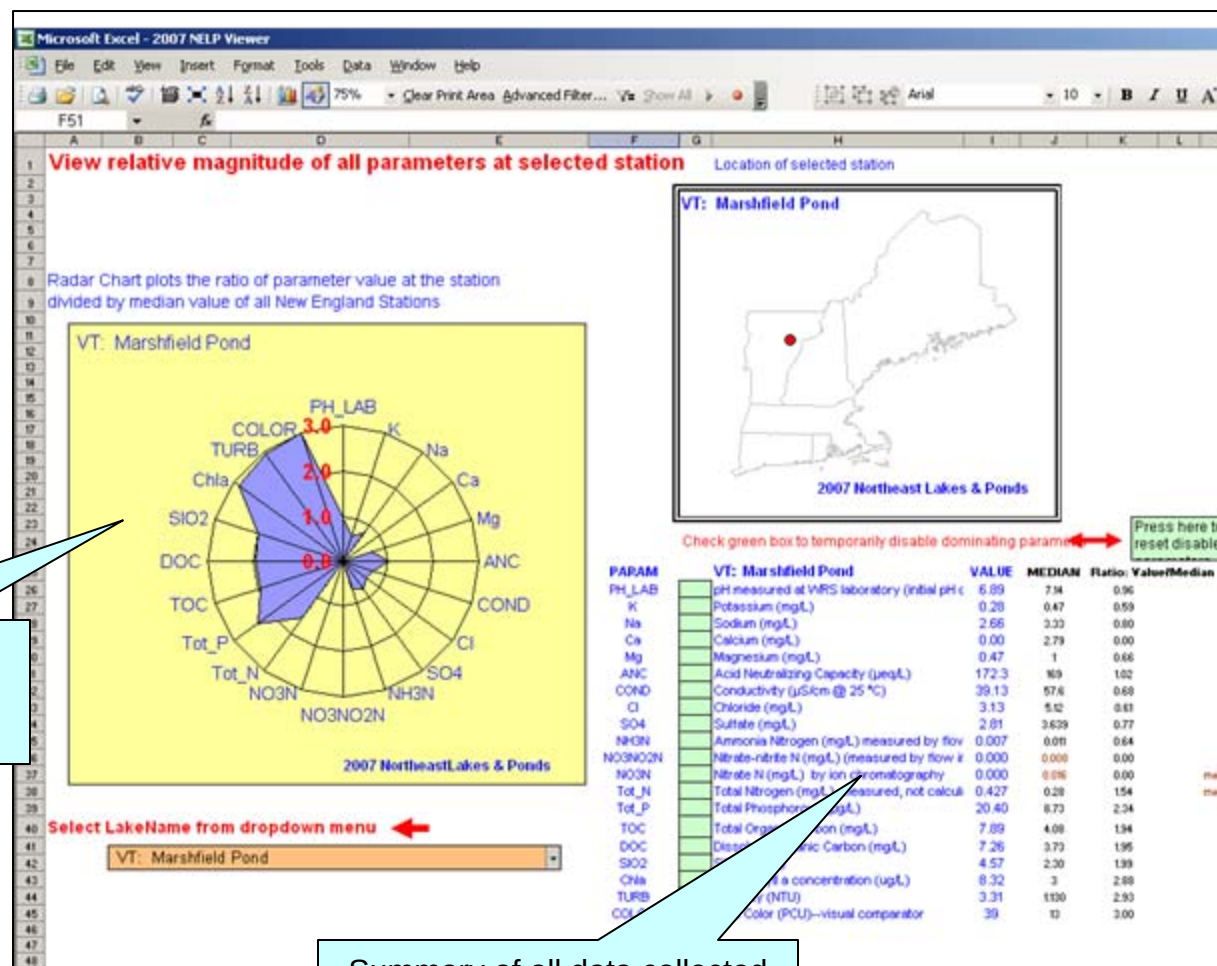


Next Step:

Add non-probabilistic local monitoring data
Compare with NLA data

Close-up Lake View:

View values of all parameters measured at a single lake
“Radar plot” compares lake data to regional averages
(ratio of station values to New England median values)



Summary: Excel Data Exploration Tool

Provide a simple tool to visualize & analyze NLA data

- All data available on spreadsheet**

 - Probabilistic NLA chem data**

 - Other data sets can be added**

- Offers graphical views of detailed data**

 - “Distribution plots”**

 - Condition maps (Low, Mid & High)**

 - Percentage plots**

 - Bubble plots**

 - Basic stats**

 - Histograms**

 - CDFs**

- Interactive operation**

 - User's choice of thresholds to define condition**

 - User can adapt Excel graphs & plots**

Summary: Continued

Tool provides views at several scales

Large scale:

National

EPA Regions

WSA Ecoregions

Regional scale:

EPA Regions

User-defined regions

State scale:

Focus on single state

Compare with neighbors

Local scale:

Focus on single lake

Summary: Continued

Potential benefits of tool

Familiar Excel environment

Simple operation

User chooses thresholds

Stats, Tables, graphics, & data available for use in reports

Useful for QA of data

Can add other data for comparison with NLA

Challenges

Tool not “bullet-proof”

Difficult to maintain version-control of data

Long term maintenance and support

Feedback is sought and appreciated

Is this approach advisable?

Who would use such a tool?

How would they use it?

Any component particularly useful or not?

What's missing?

Contact:

John Kiddon; kiddon.john@epa.gov; 401-782-3044

Hal Walker; walker.henry@epa.gov; 401-782-3134