



Prepared in cooperation with the Idaho Department of Environmental Quality

# Compilation and Analysis of Multiple Groundwater-Quality Datasets for Idaho

Open-File Report 2018-1079

By Stephen A. Hundt and Candice B. Hopkins

# Purpose

- Combine datasets
  - Several agencies in the State of Idaho regularly collect groundwater-quality data
  - Data housed in at least four separate databases
- Summarize available data
  - What data have been collected, from where, when, and by whom?
- Perform analysis
  - Demonstrate how a large statewide dataset can be analyzed in bulk
  - Learn something about Idaho groundwater quality

# METHODS

# Database Compilation

- Collected data up to September 2016, from:
  - Water Quality Portal
    - U.S. Geological Survey (USGS), Environmental Protection Agency (EPA), Bureau of Reclamation (Reclamation), others
    - Queried through web interface
    - <https://www.waterqualitydata.us/>
  - Idaho Department of Environmental Quality (IDEQ)
    - Made personal data request
  - Idaho Department of Water Resources (IDWR)
    - Made personal data request
- Missing data from:
  - Idaho State Department of Agriculture (ISDA)
    - Large set of pesticide analyses

# Database Compilation

- MS Access Database
  - Samples
    - Water-quality sample results
  - Monitoring Locations
    - Latitude, longitude, name, depth, aquifer, and other details
  - Organization Codes
    - Relates the location identification system used in this database to those used by different agencies
  - Nitrate Priority Areas
    - Location identification, name, and statistics of 2014 nitrate priority areas, as delineated by the Idaho Department of Environmental Quality (2014)

# General Analysis Approach

- Combined some well records
  - Created composite wells based upon location, name, and depth information
  - Only for analysis (and results); not the database
  - In results and shapefile, composite wells have a locationID greater than 30,000
- Removed records without measurement units
  - About 125,000 removed (7 percent of records)
- Did not attempt to:
  - Identify and fix misplaced decimal places, or misreported measurement units
  - Look for samples that may have been non-detects, but were not explicitly reported as such

# General Analysis Approach

- Focused on data collected since 1990
  - 89 percent of samples
- Nitrate priority areas used when summarizing analysis results
  - As delineated by IDEQ (Idaho Department of Environmental Quality, 2014)
  - Just for the examples in this presentation, all well-by-well analysis results available:  
<https://www.sciencebase.gov/catalog/item/5a2724dde4b03852bafc8ebd>
- Depth and aquifer ignored for simplification
- Used R programming language

# Data Availability Summary

- A summary of the database contents
  - Location of wells
  - Summaries of total, by year, statewide, and by 2014 Nitrate priority area
    - Number of samples and site visits
    - Number of wells
    - Number of analytes
    - By agency

# Drinking Water Standard Exceedances

- EPA sets Maximum Contaminant Level (MCL) standards for 75 chemicals, micro-organisms, and radionuclides (U.S. Environmental Protection Agency, 2017)
  - <https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations>
- Summarized number of MCL exceedances:
  - Compared to MCLs as of May 2017
  - For each of the 60 analytes in the database with an MCL
  - For every well
  - Since 1990

# Summary Statistics

- Number, minimum, maximum, mean, median, standard deviation
  - Used NADA R package (Lee, 2017) to account for censored data (non-detects)
    - Kaplan-Meier method
  - 57 analytes with MCLs had sufficient data for these calculations
  - Every well
  - Since 1990

# Trend Analysis

- Mann-Kendall trend test (Helsel and Hirsch, 2002)
  - For strength of trend and statistical significance
    - tau and p-values
  - Used NADA R package to account for censored data
- Akritas-Theil-Sen non-parametric slope
  - Median of all pairwise slopes in dataset
  - Slope represents the rate and direction of concentration change
- 51 analytes with MCLs had sufficient data for these calculations
- Every well
- Since 1990
  - No limit was set on the spacing of data within 26 year window
    - Results include wells without recent data and wells with short data-collection periods

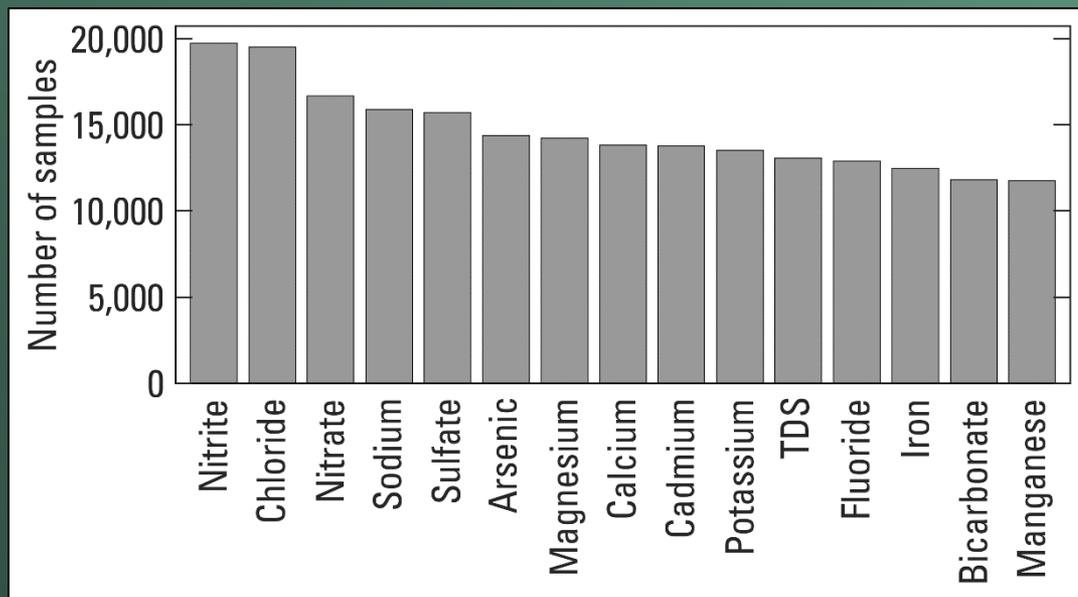
# Correlations

- Grouped all statewide data together for larger sample sizes
  - Rather than separating by region or well
  - Locally specific associations may be missed
- Ran Spearman tests for association between pairs of analytes
  - Resistant to outliers
  - No assumption of linearity
- 63 analytes with sufficient data for these calculations were included
  - Those with MCLs and a few extra of interest

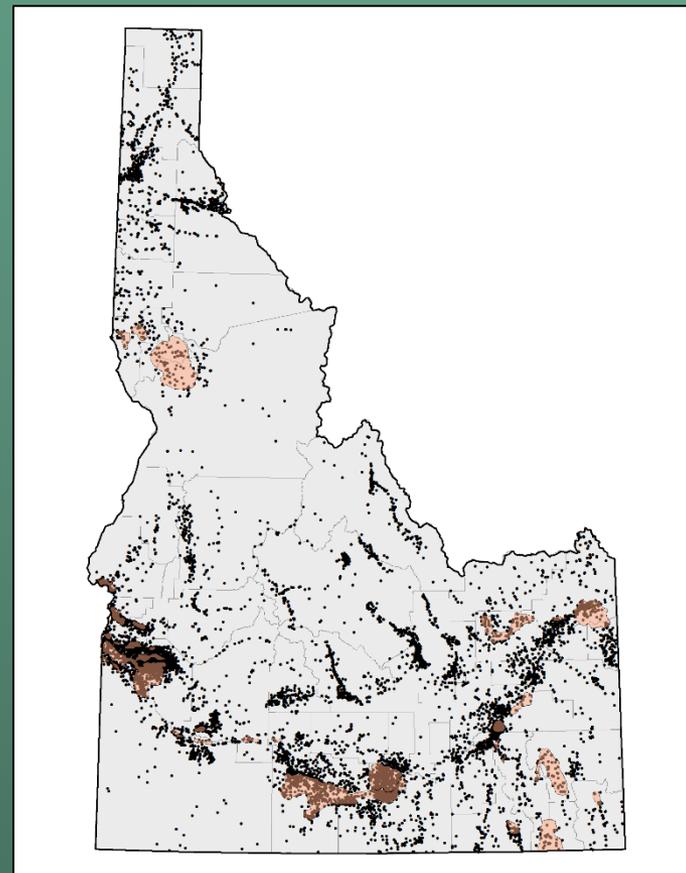
# RESULTS

# Data Compilation Summary

- Distributed throughout State
- 1.9 million sample results
- More than 16,000 wells
- More than 800 analytes



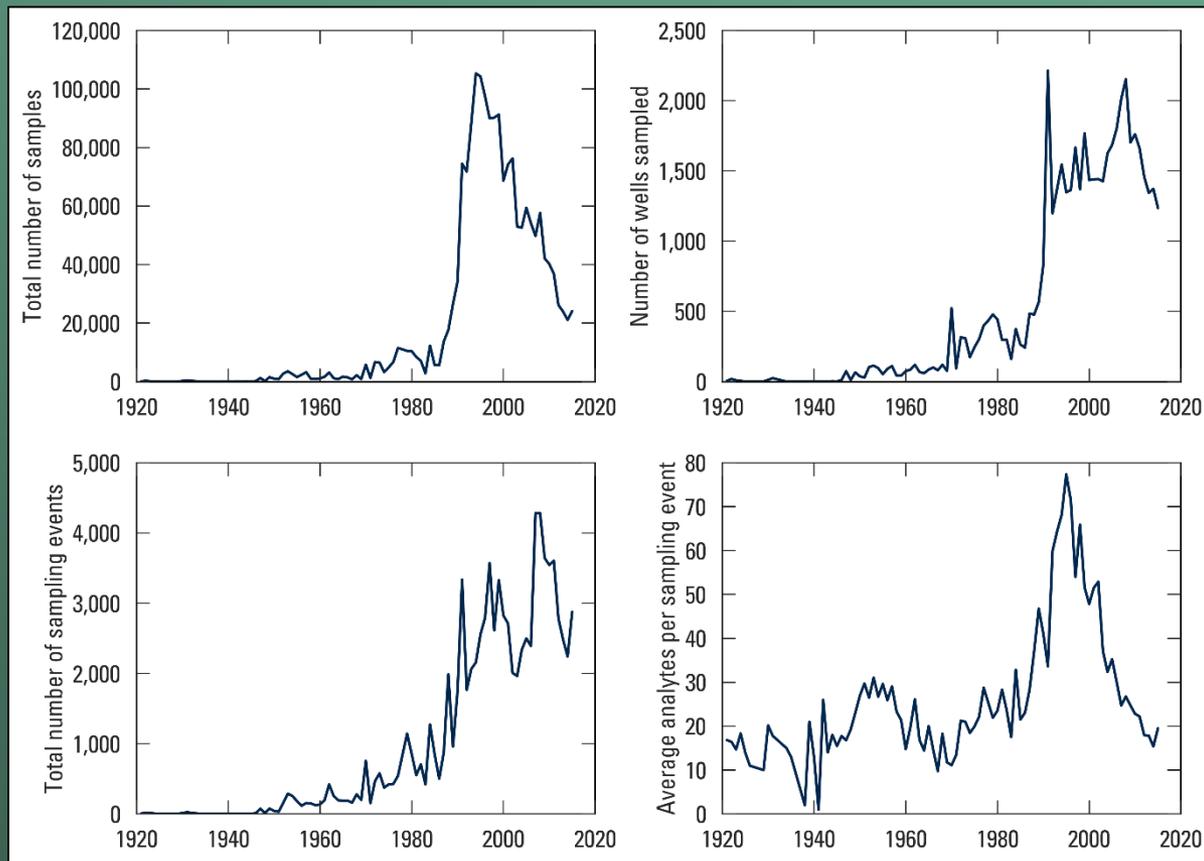
Total samples for top 15 analytes



Well locations and 2014 Nitrate Priority Areas (orange)

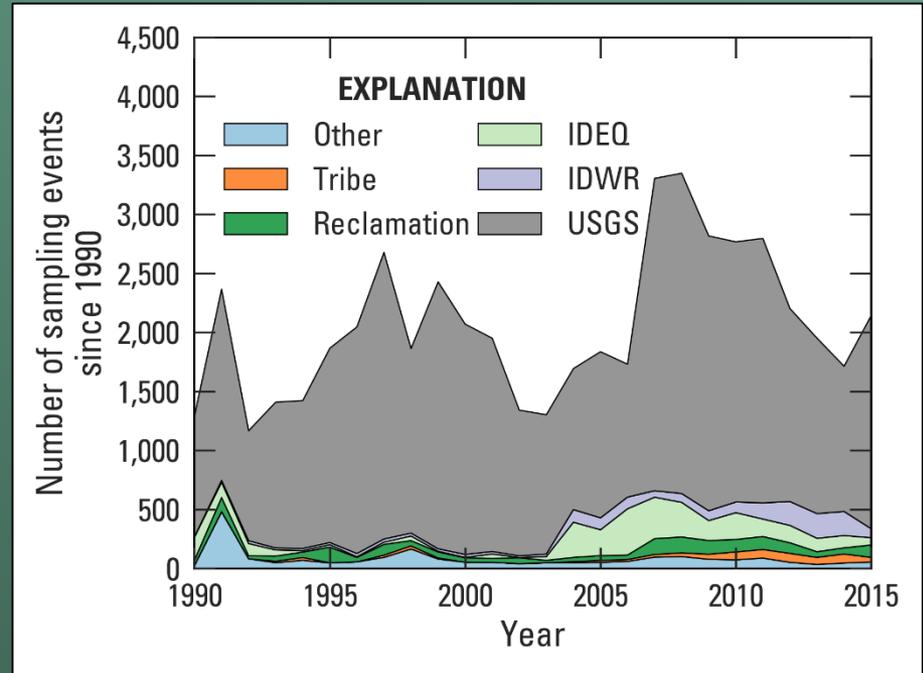
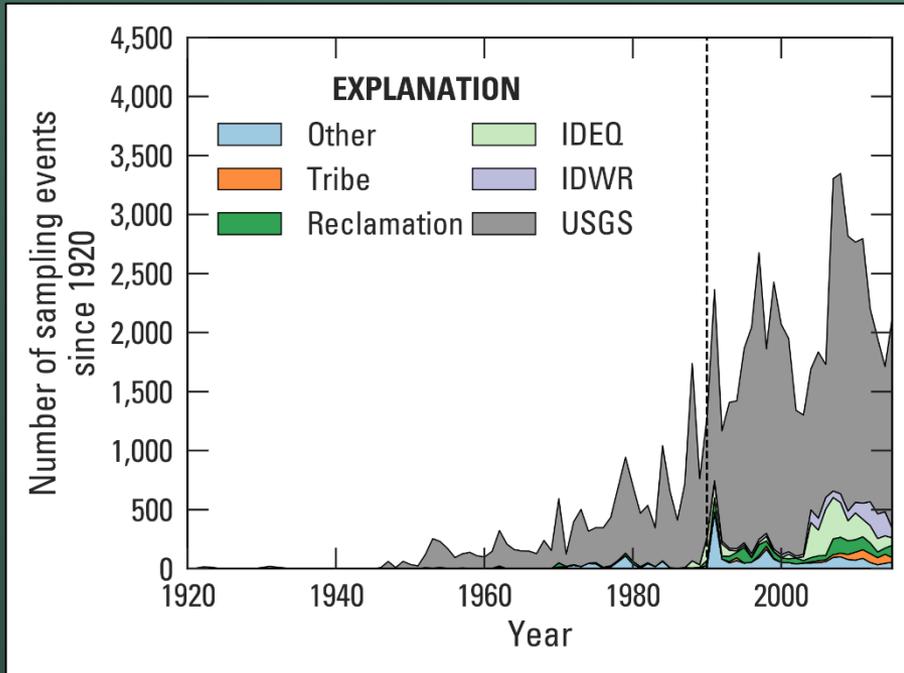
# Data Compilation Summary

- Overall increase in sampling per year through the mid-1990s, with notable recent decreases
- A sampling event refers to site visit during which one or more water samples were collected for analysis
- Decreases in total samples since late 1990s largely result of decrease in analytes tested per event
- Number of wells and sampling events have also decreased more recently



# Data Compilation Summary

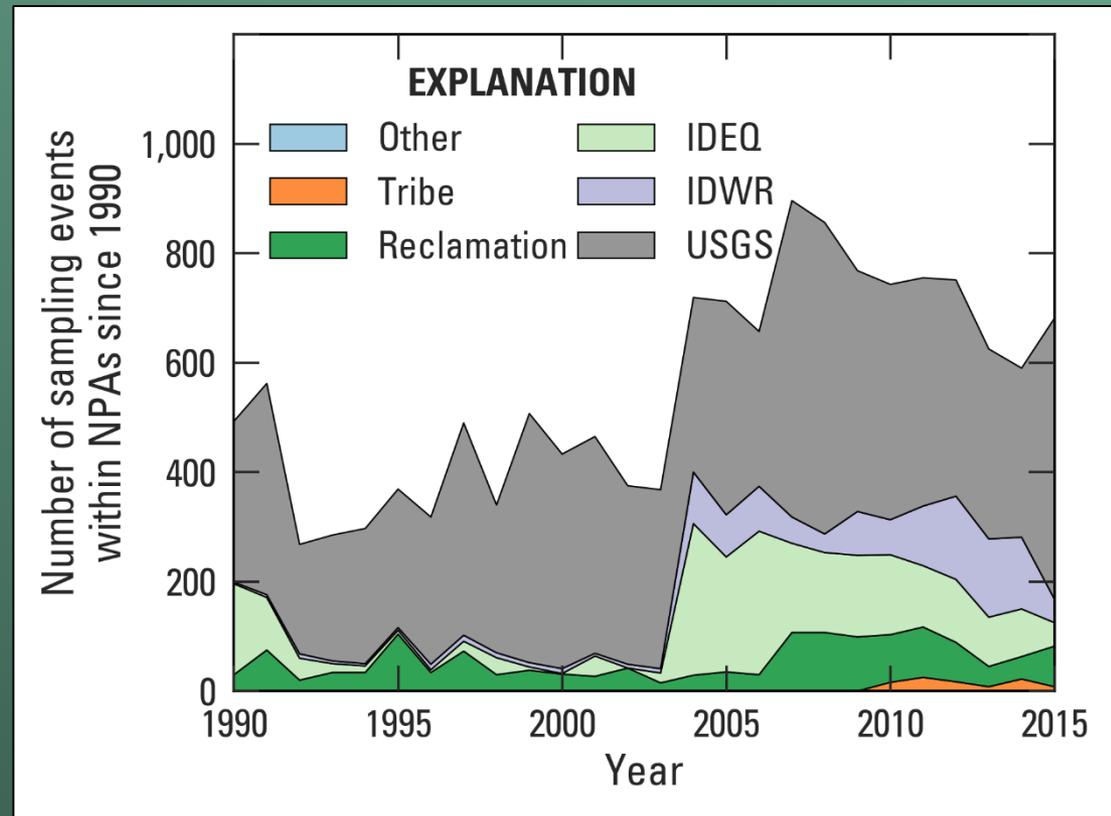
- Most active agencies (not including ISDA)
  - USGS
  - Reclamation
  - IDEQ
  - Tribes
  - IDWR
  - EPA (included in “Other” group in figures)



# Data Compilation Summary

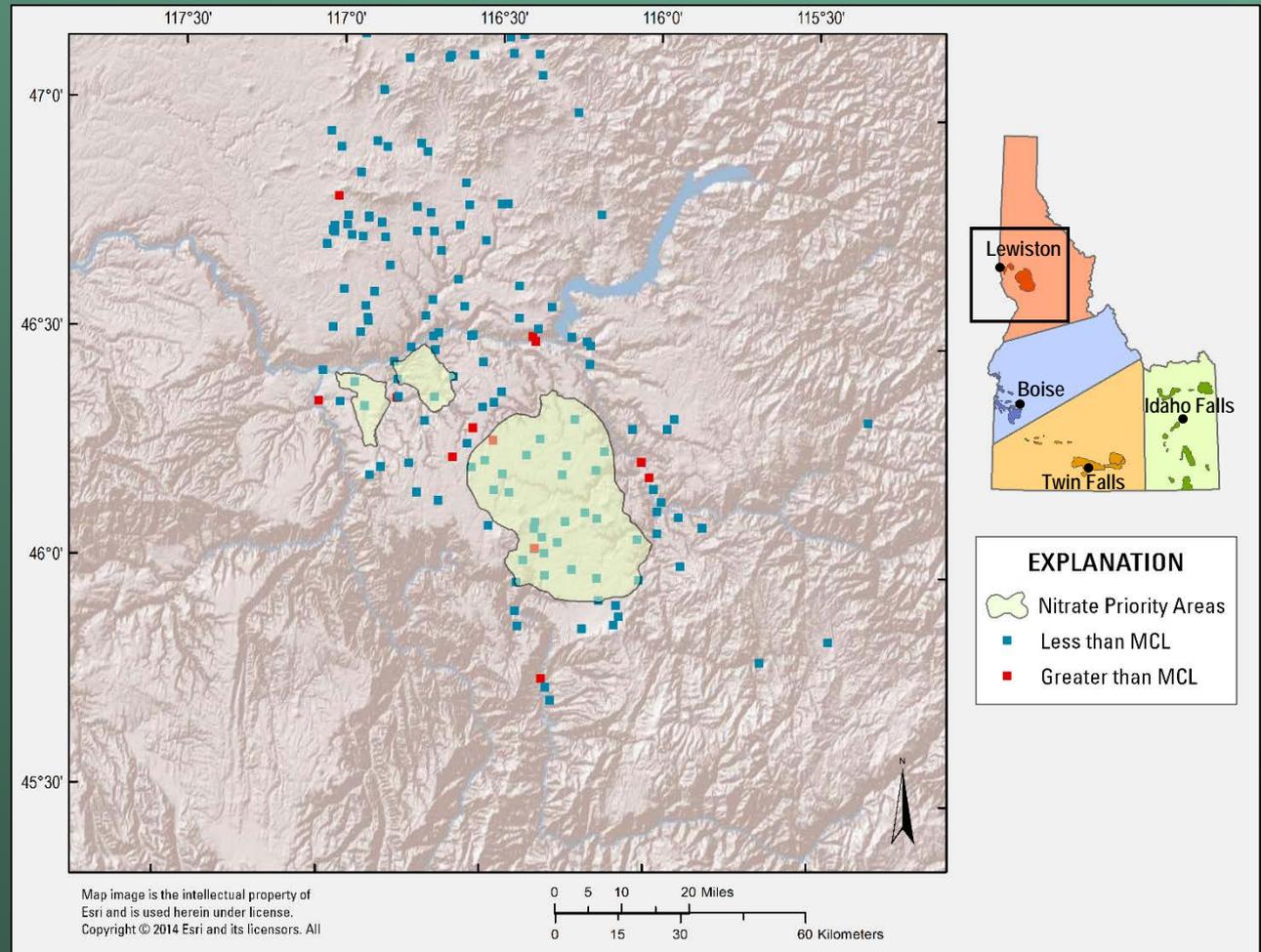
- Around 25 percent of statewide sampling activity has occurred within Nitrate Priority Areas (NPA)

	Statewide	Nitrate Priority Areas
Number of Samples	1,940,714	478,152
Number of Sampling Events	97,201	25,395



# MCL Exceedances

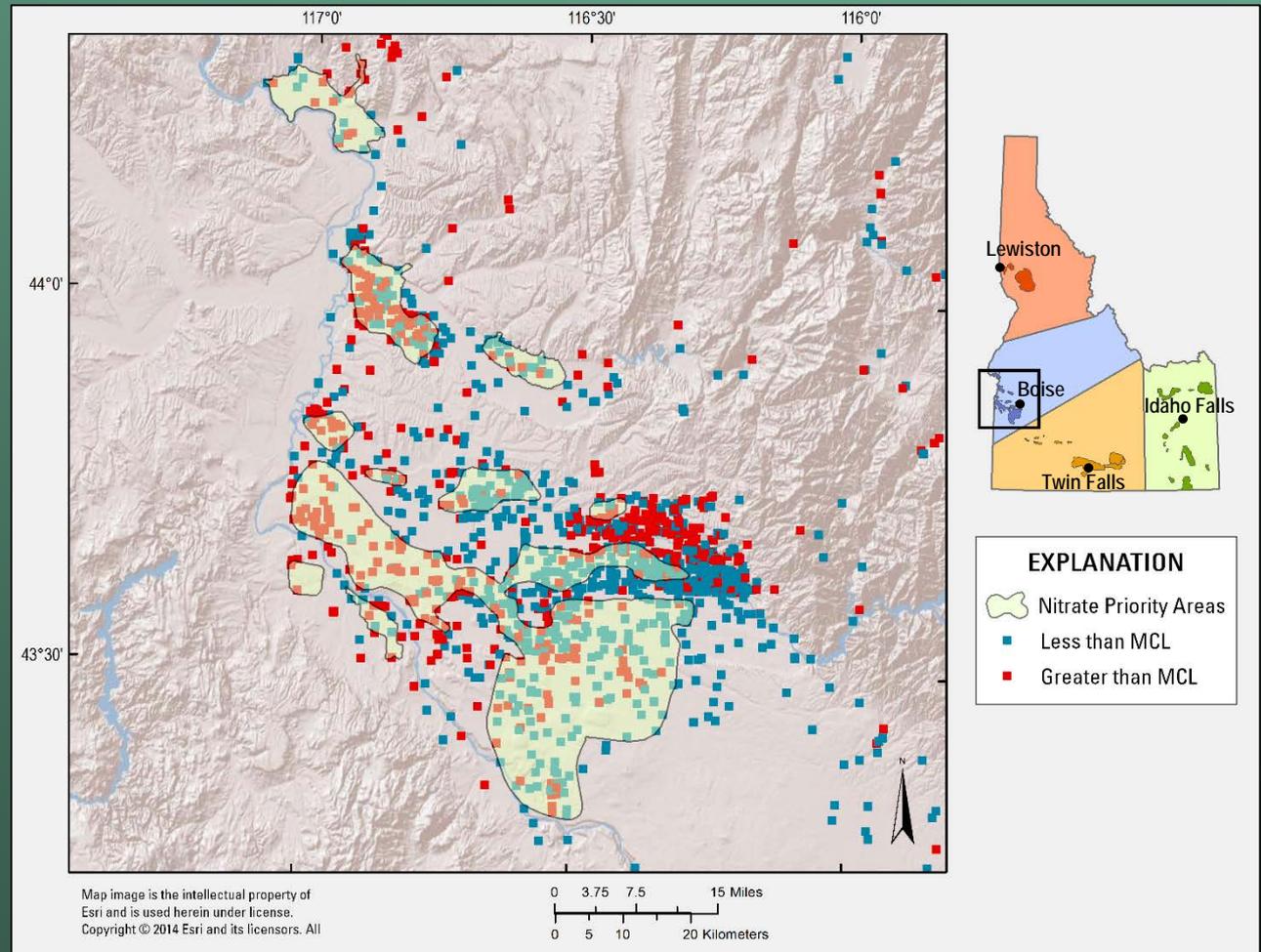
- Arsenic MCL is 0.01 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with arsenic MCL exceedances since 1990 in northern NPAs. Idaho is divided into north, west, south, and east regions for display.

# MCL Exceedances

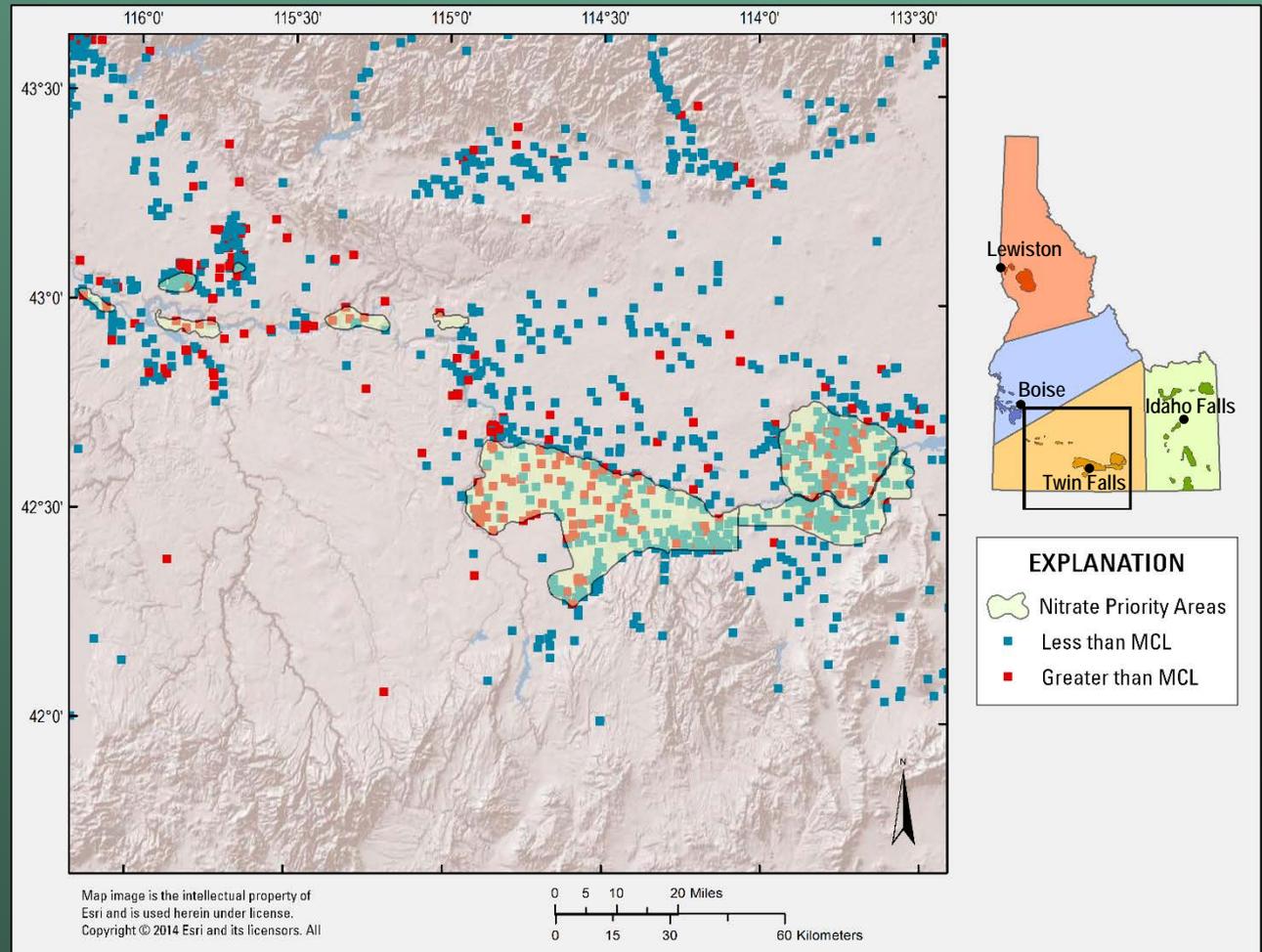
- Arsenic MCL is 0.01 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with arsenic MCL exceedances since 1990 in western NPAs. Idaho is divided into north, west, south, and east regions for display.

# MCL Exceedances

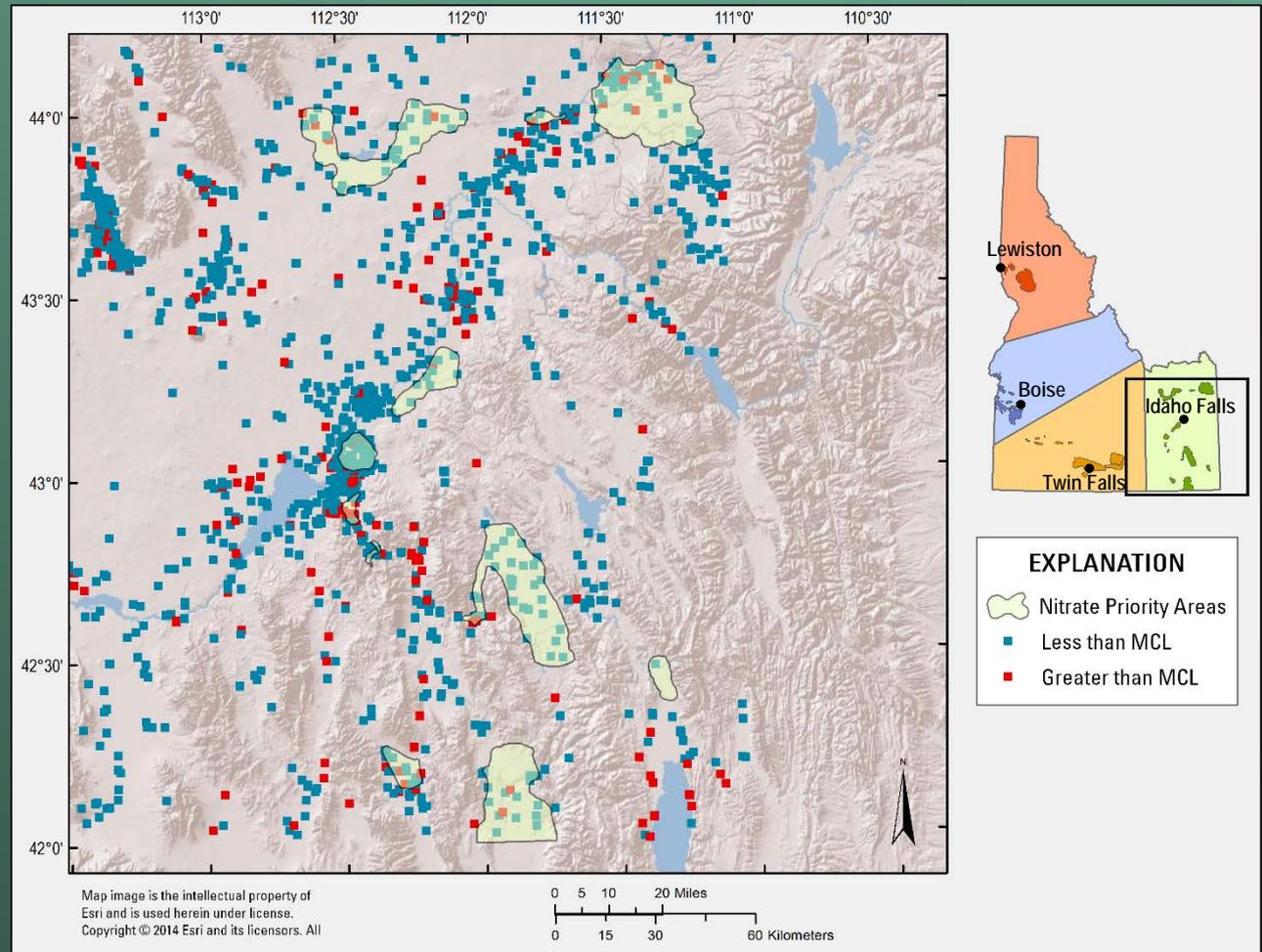
- Arsenic MCL is 0.01 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with arsenic MCL exceedances since 1990 in southern NPAs. Idaho is divided into north, west, south, and east regions for display.

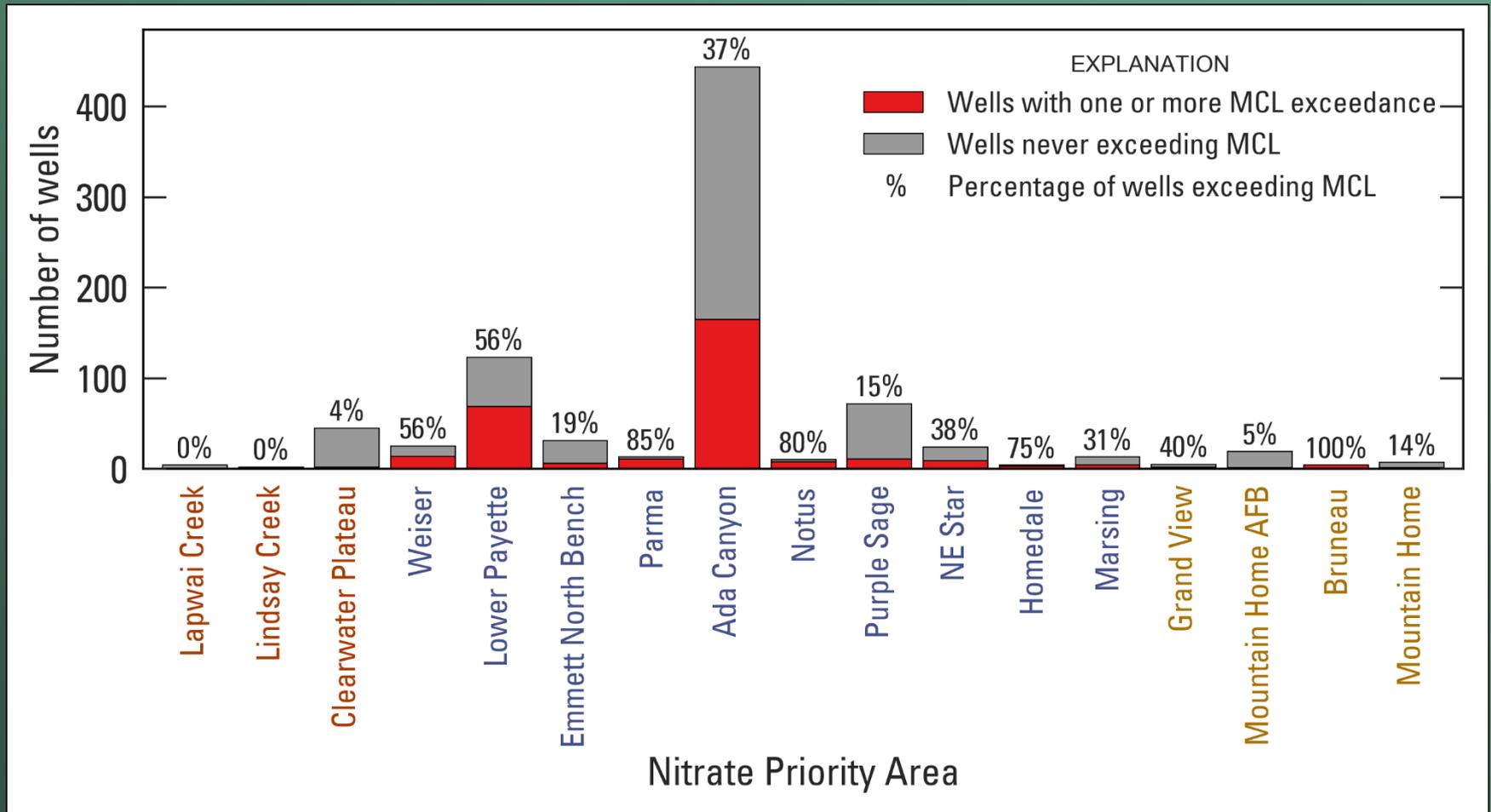
# MCL Exceedances

- Arsenic MCL is 0.01 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with arsenic MCL exceedances since 1990 in eastern NPAs. Idaho is divided into north, west, south, and east regions for display.

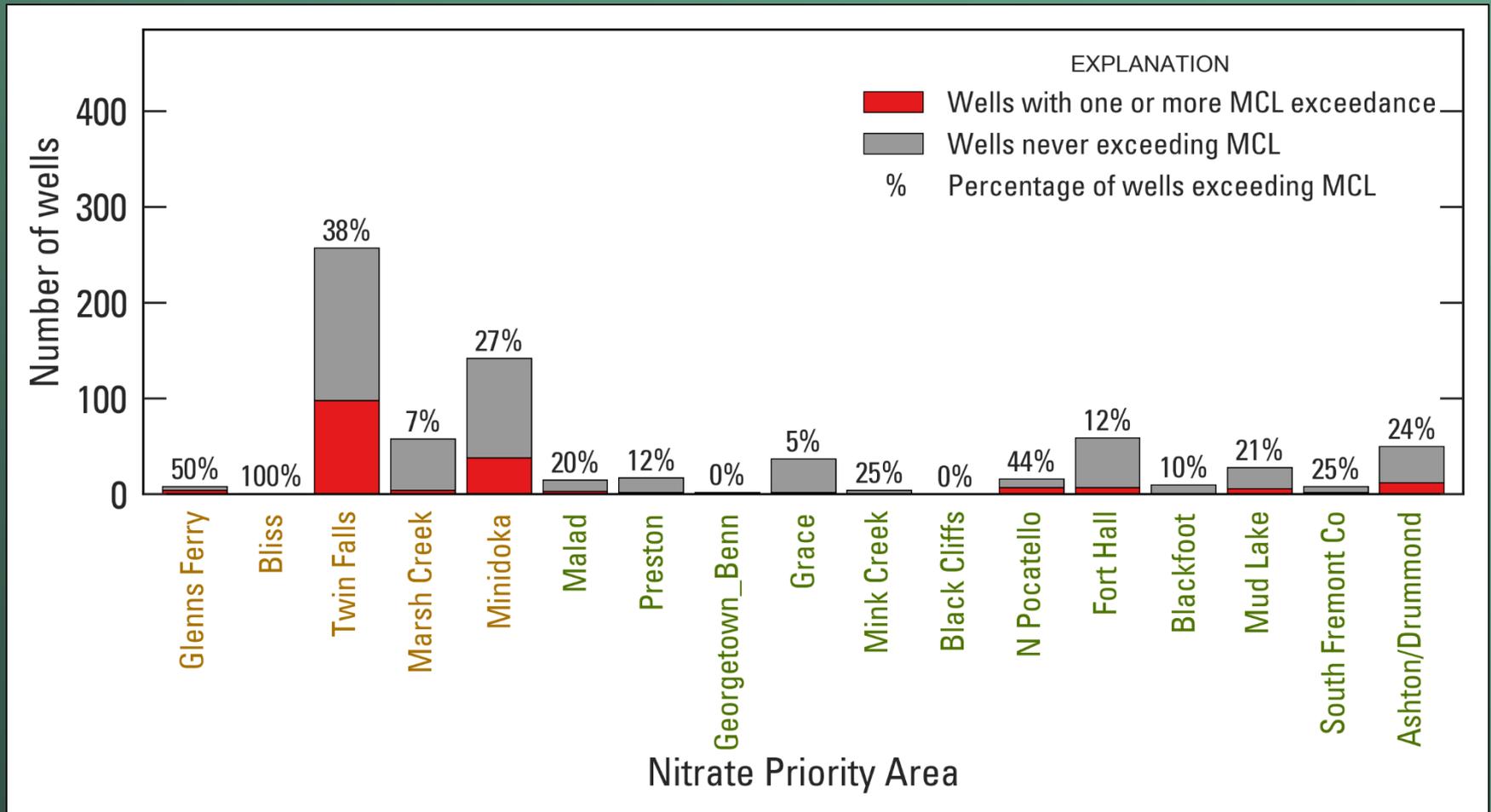
# MCL Exceedances



Wells with arsenic MCL exceedances since 1990. Nitrate Priority Area names are colored by regional groups as shown on the previous maps.



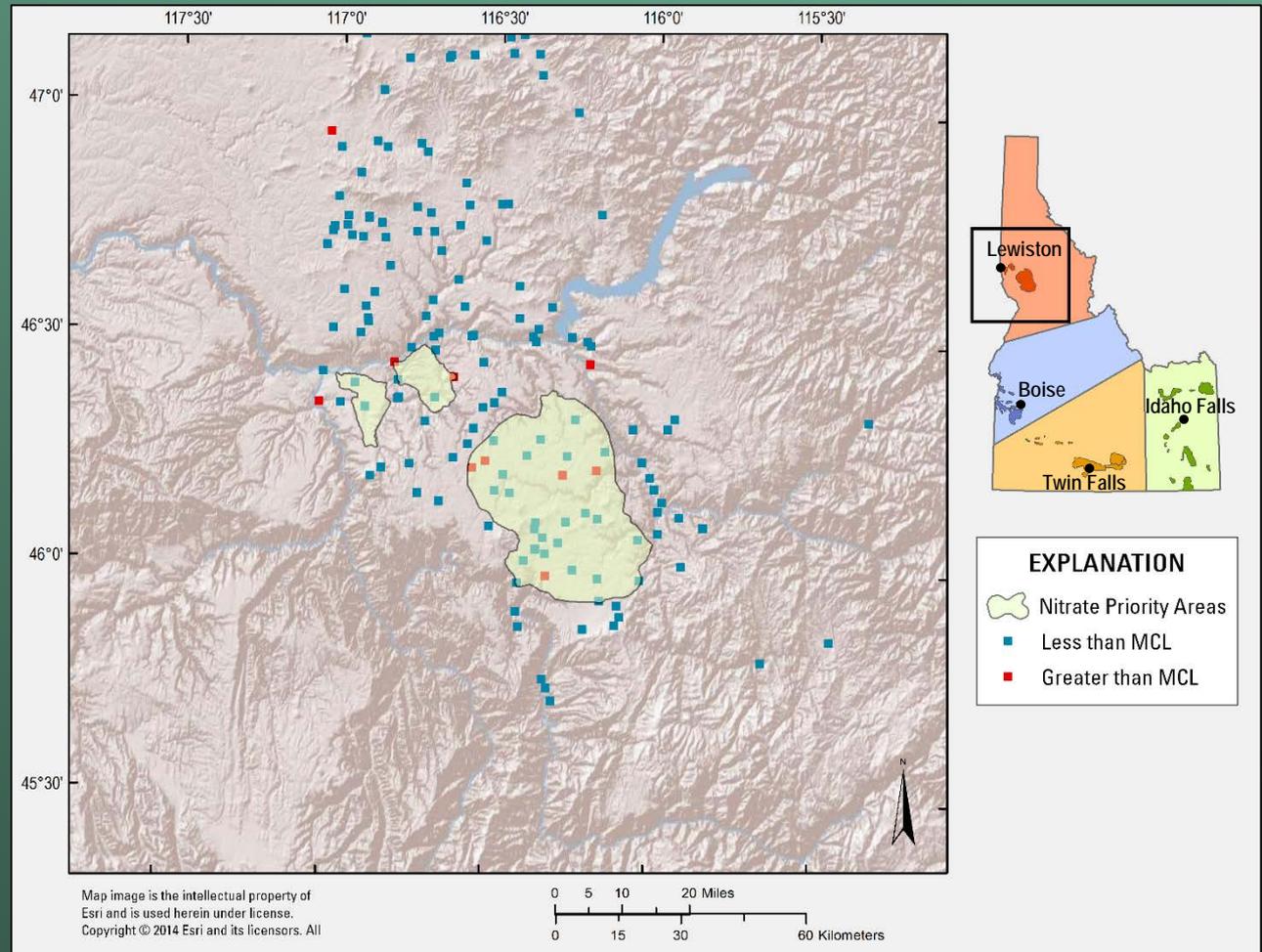
# MCL Exceedances



Wells with arsenic MCL exceedances since 1990. Nitrate Priority Area names are colored by regional groups as shown on the previous maps.

# MCL Exceedances

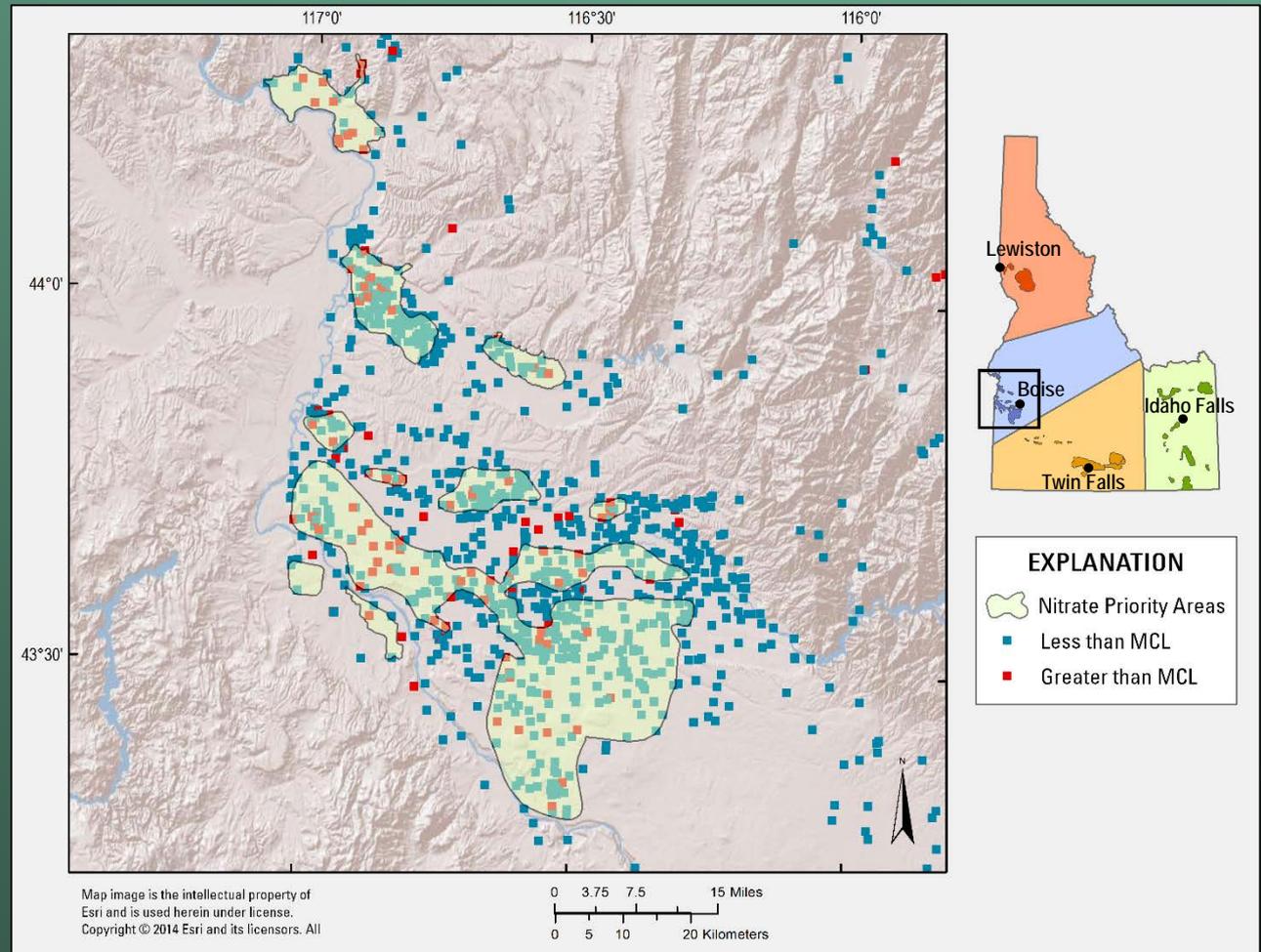
- Nitrate MCL is 10 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with nitrate MCL exceedances since 1990 in northern NPAs. Idaho is divided into north, west, south, and east regions for display.

# MCL Exceedances

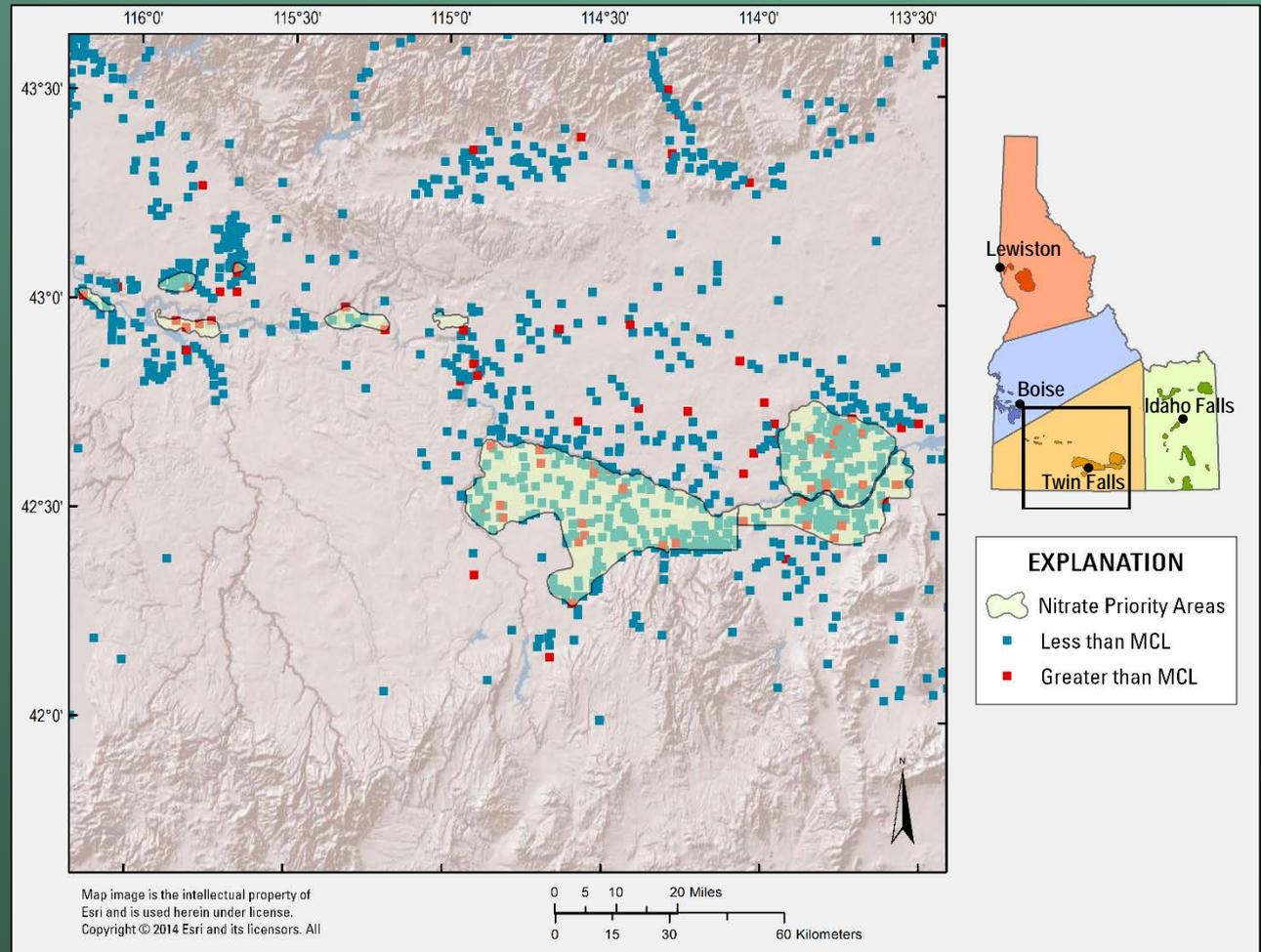
- Nitrate MCL is 10 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with nitrate MCL exceedances since 1990 in western NPAs. Idaho is divided into north, west, south, and east regions for display.

# MCL Exceedances

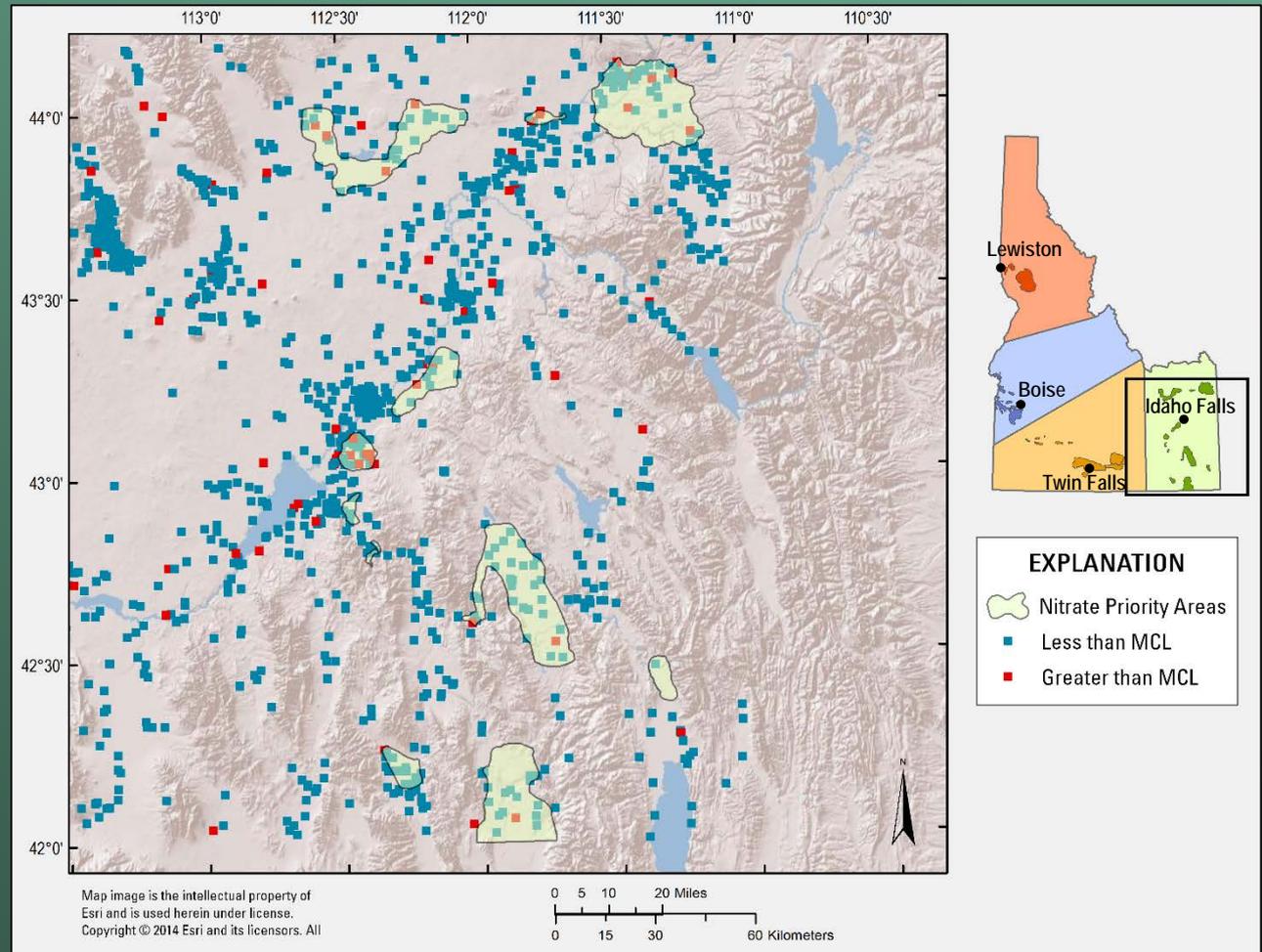
- Nitrate MCL is 10 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with nitrate MCL exceedances since 1990 in southern NPAs. Idaho is divided into north, west, south, and east regions for display.

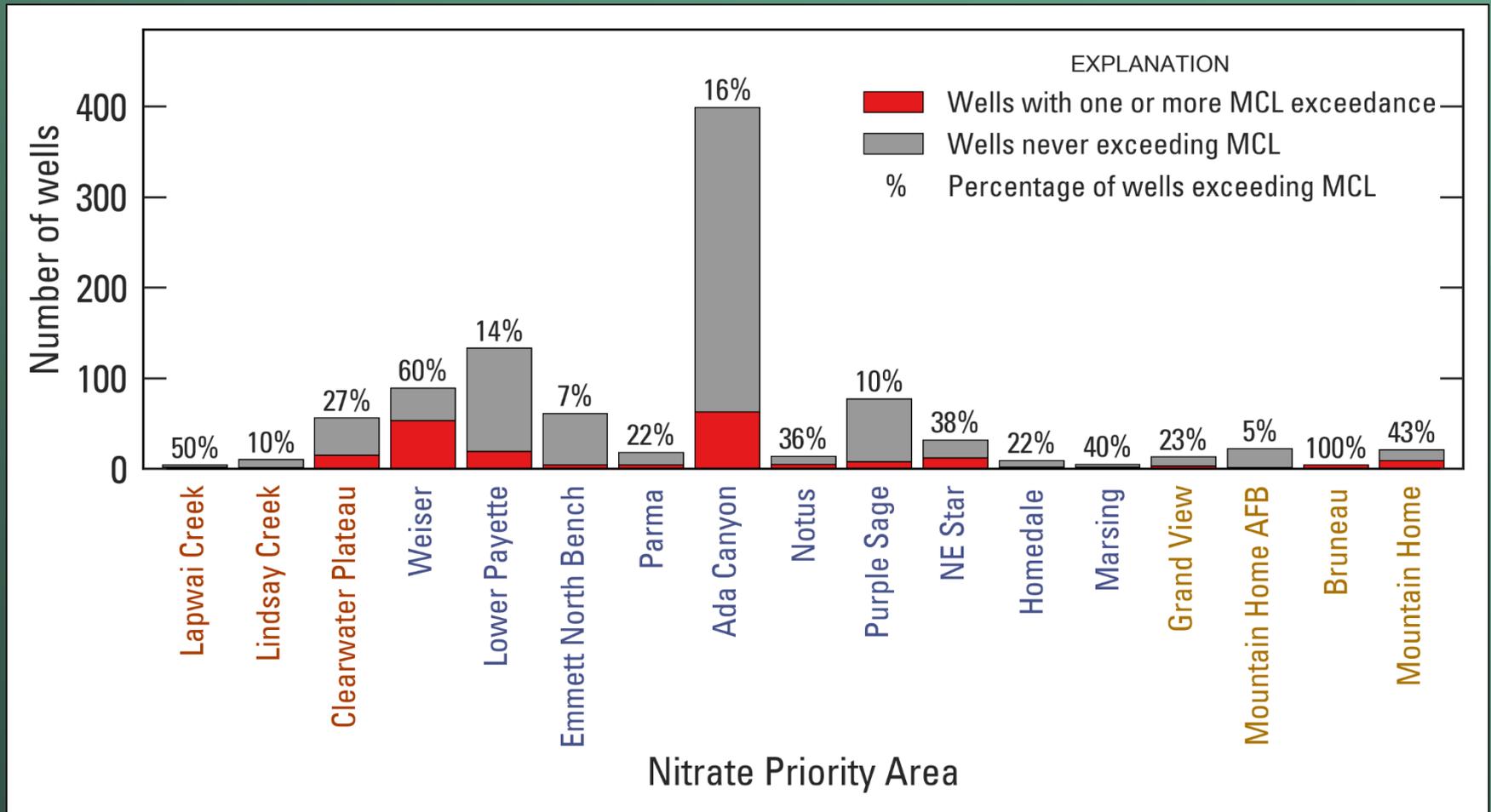
# MCL Exceedances

- Nitrate MCL is 10 mg/L
- Since 1990
- Ever exceeded?
  - Yes or No



Wells with nitrate MCL exceedances since 1990 in eastern NPAs. Idaho is divided into north, west, south, and east regions for display.

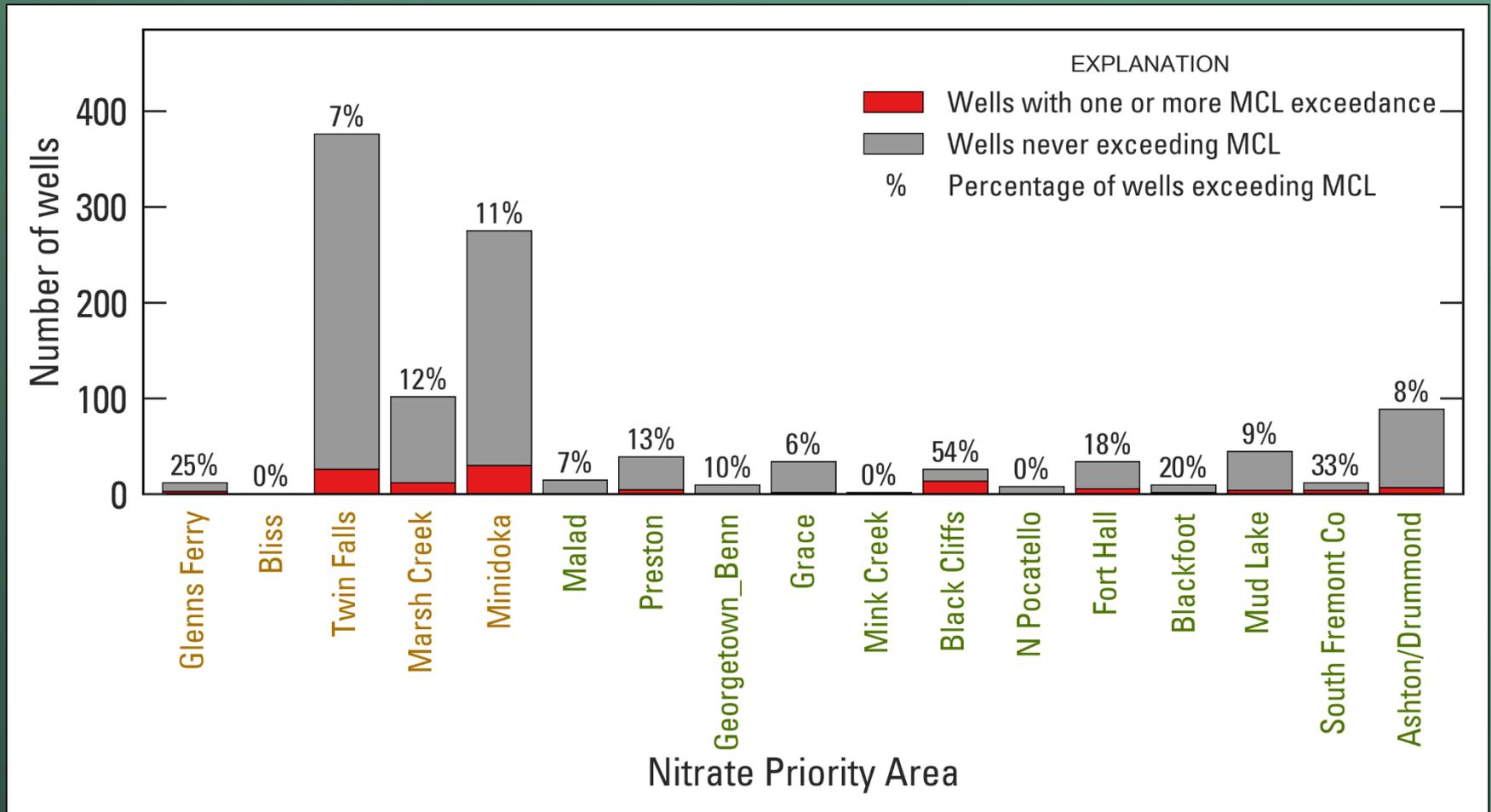
# MCL Exceedances



Wells with arsenic MCL exceedances since 1990. Nitrate Priority Area names are colored by regional groups as shown on the previous maps.



# MCL Exceedances



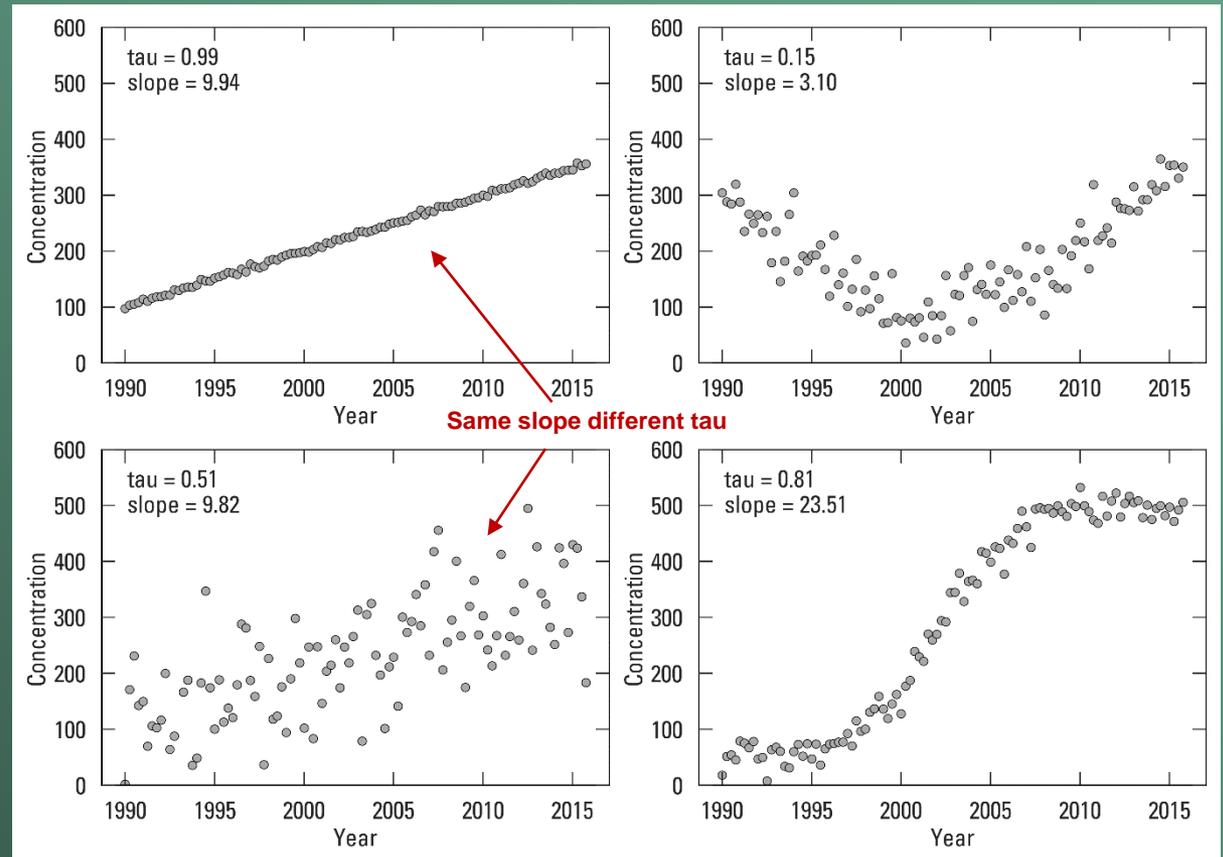
Wells with arsenic MCL exceedances since 1990. Nitrate Priority Area names are colored by regional groups as shown on the previous maps.

# MCL Exceedances

- Exceedance results for 75 analytes are available at <https://doi.org/10.5066/F72V2FBG>.

# Results – Trends

- Kendall's  $\tau$ 
  - strength of monotonic trend
  - linearity not assumed
- $p$ -value
  - statistical significance of  $\tau$
- Slope
  - magnitude of trend
  - Akritas-Theil-Sen nonparametric line
- Will miss breaks or switches in trend direction
- No assumption of linearity



Synthetic examples of time-series to demonstrate features of the trend statistics

# Trends

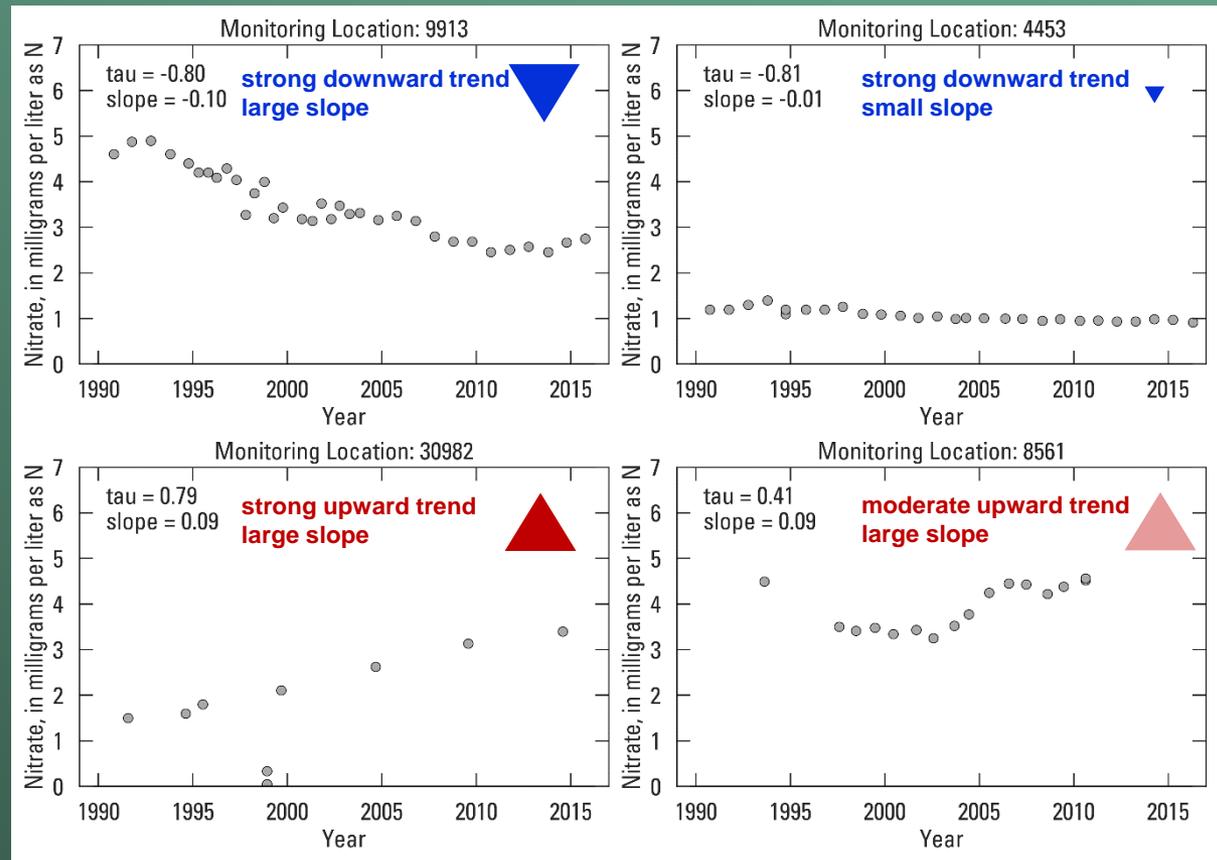
- Color of triangle indicates direction of trend:

- Blue: decreasing concentration
- Red: increasing concentration

- Size indicates magnitude of slope

- Opacity indicates magnitude of  $\tau$

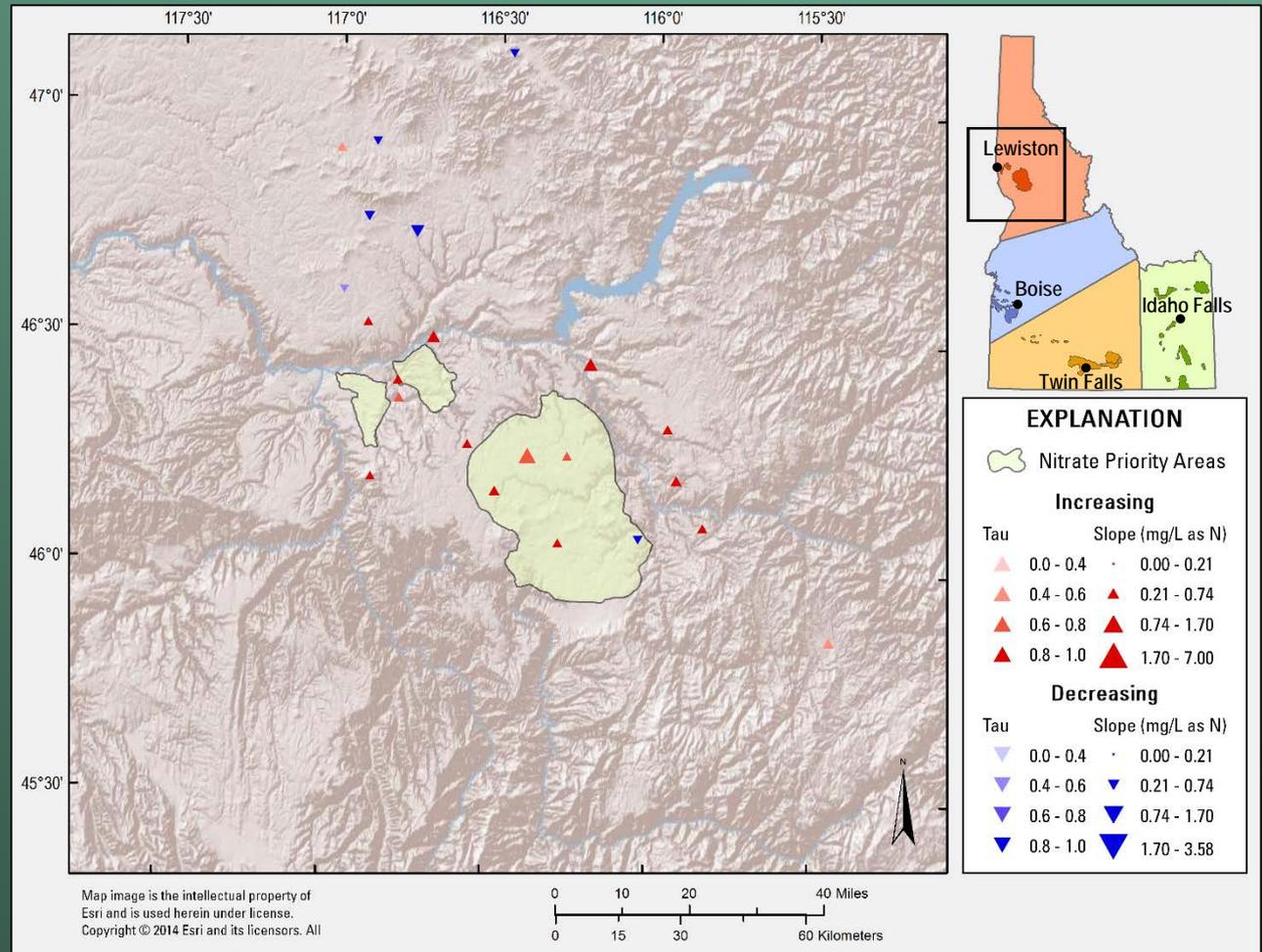
- Stronger color: higher magnitude, stronger trend
- Weaker color: lower magnitude, weaker trend



Real examples of concentration time-series to demonstrate how of  $\tau$  and slope are symbolized on the following maps

# Trends

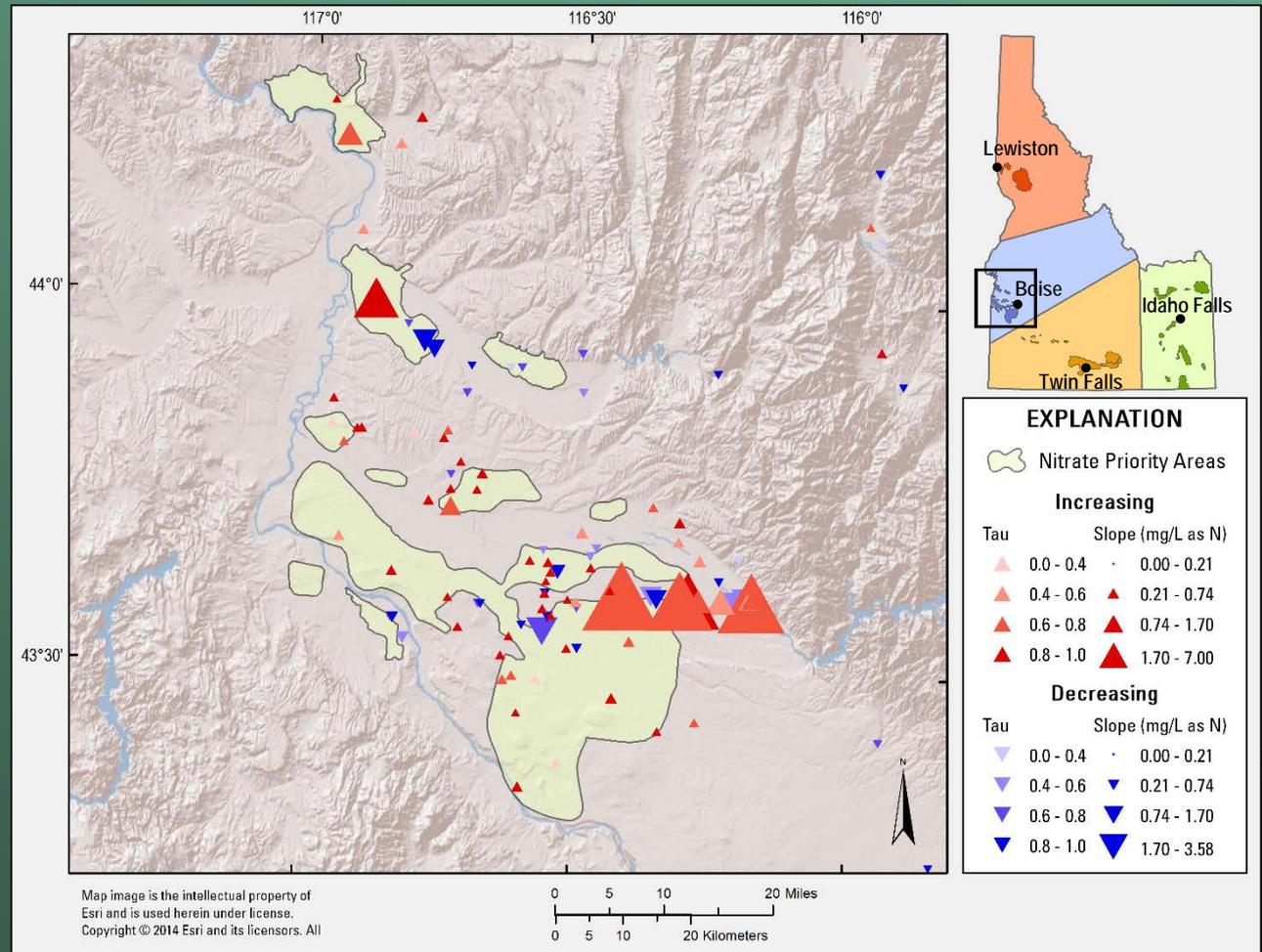
- Includes locations with:
  - More than 3 detections
  - $p < 0.1$
- 9 percent of all locations meet these criteria for Nitrate



Nitrate trends since 1990 in eastern NPAs. Idaho is divided into north, west, south, and east regions for display.

# Trends

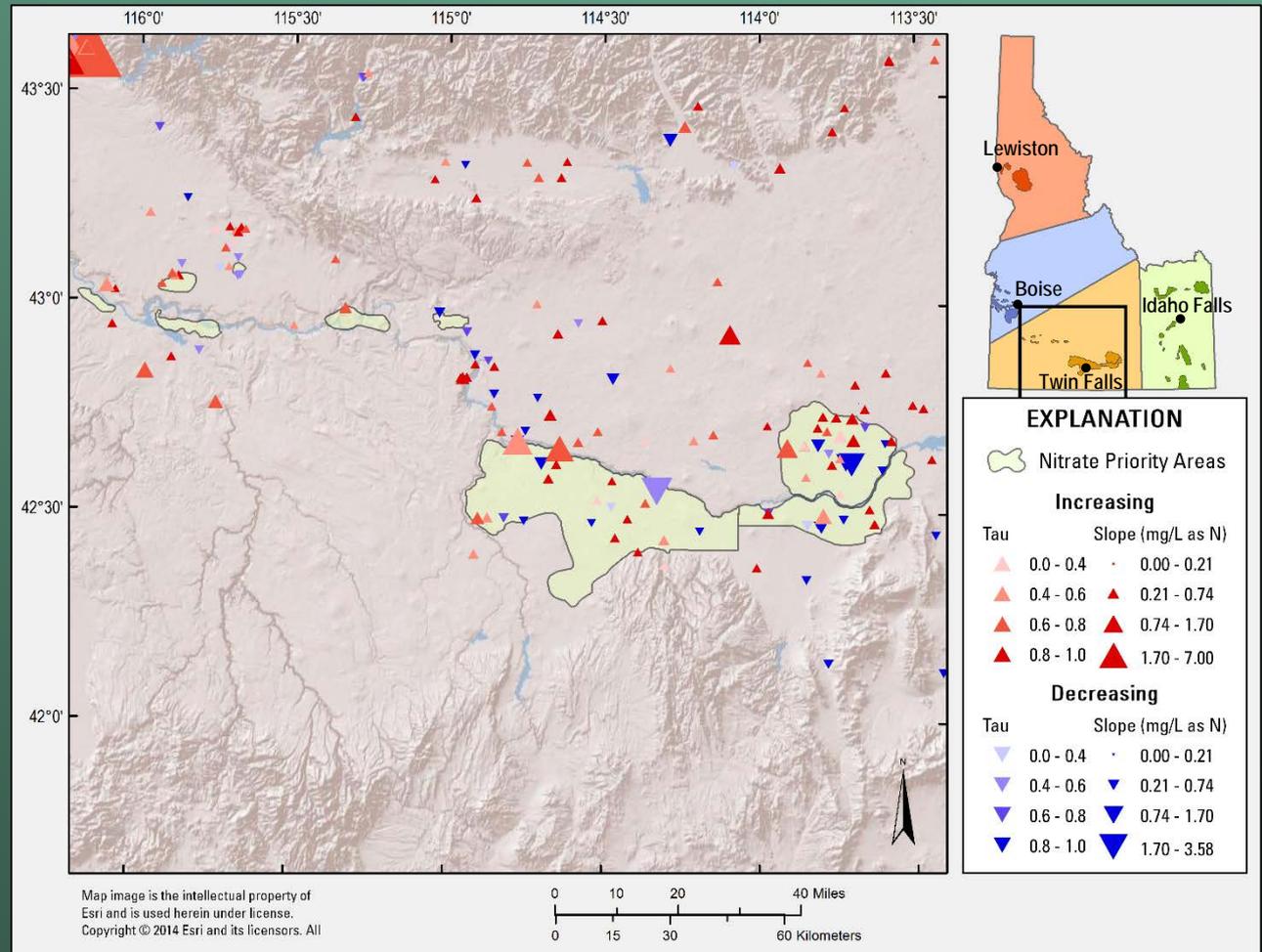
- Includes locations with:
  - More than 3 detections
  - $p < 0.1$
- 9 percent of all locations meet these criteria for Nitrate



Nitrate trends since 1990 in western NPAs. Idaho is divided into north, west, south, and east regions for display.

# Trends

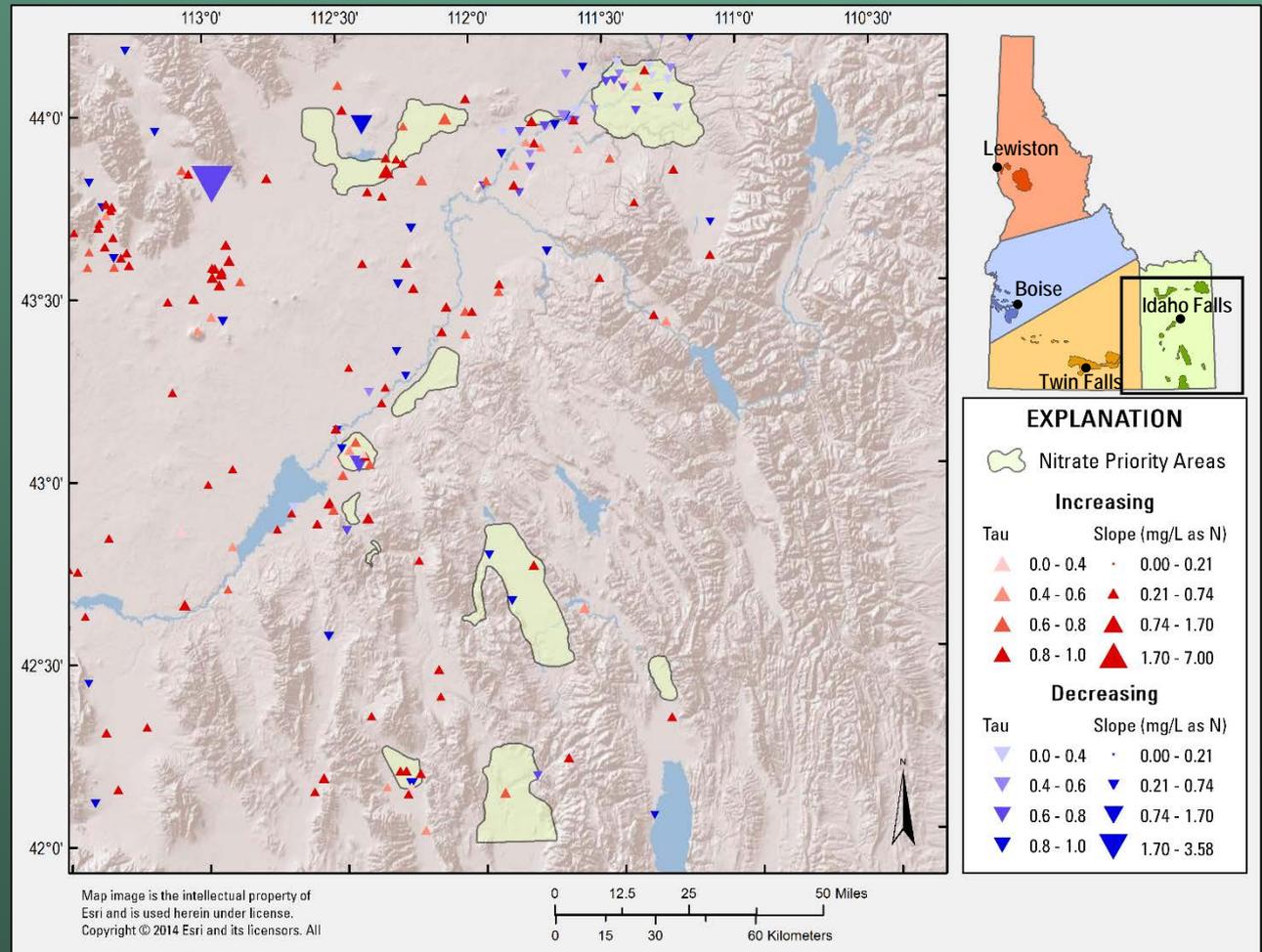
- Includes locations with:
  - More than 3 detections
  - $p < 0.1$
- 9 percent of all locations meet these criteria for Nitrate



Nitrate trends since 1990 in southern NPAs. Idaho is divided into north, west, south, and east regions for display.

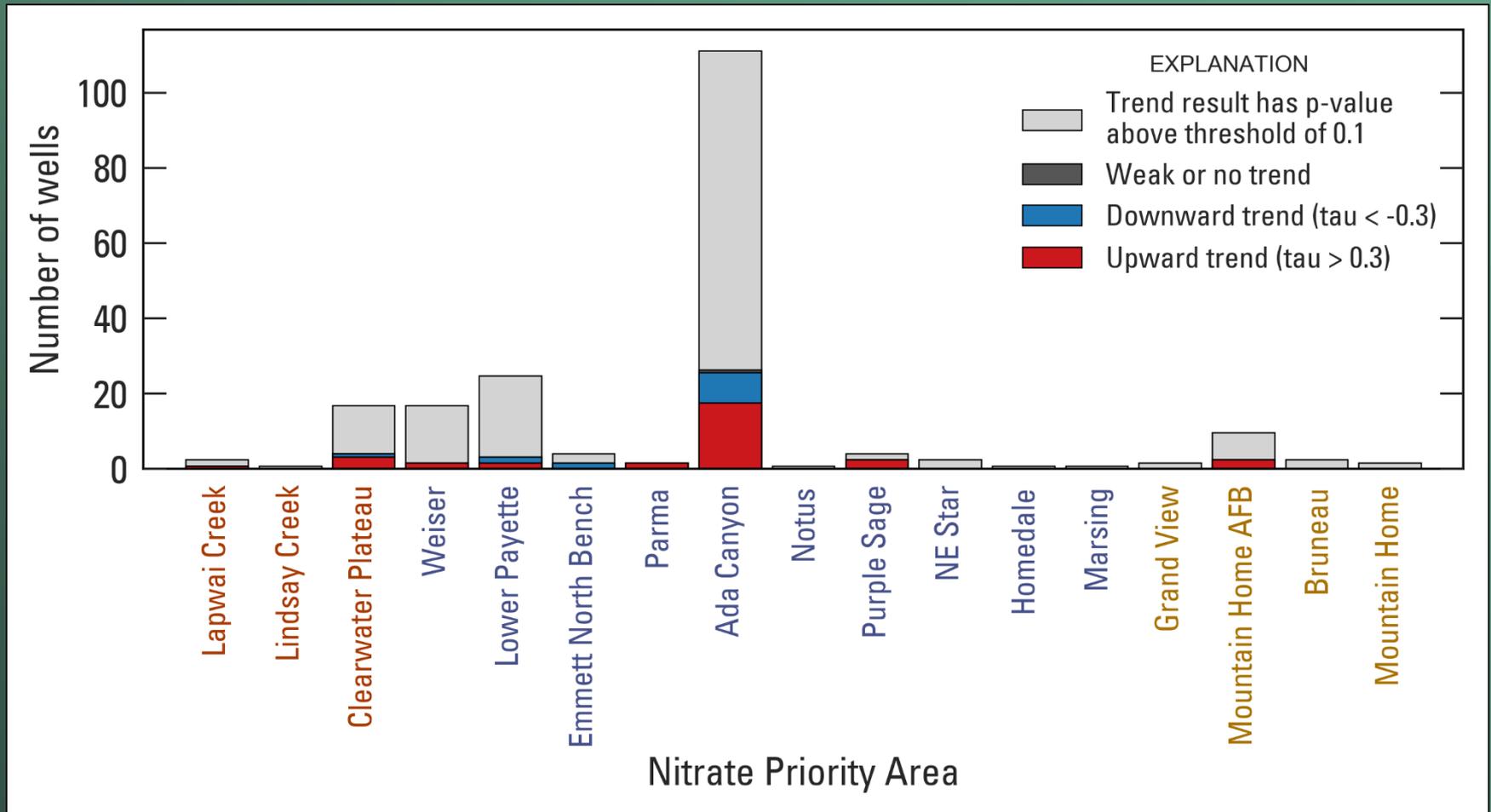
# Trends

- Includes locations with:
  - More than 3 detections
  - $p < 0.1$
- 9 percent of all locations meet these criteria for Nitrate



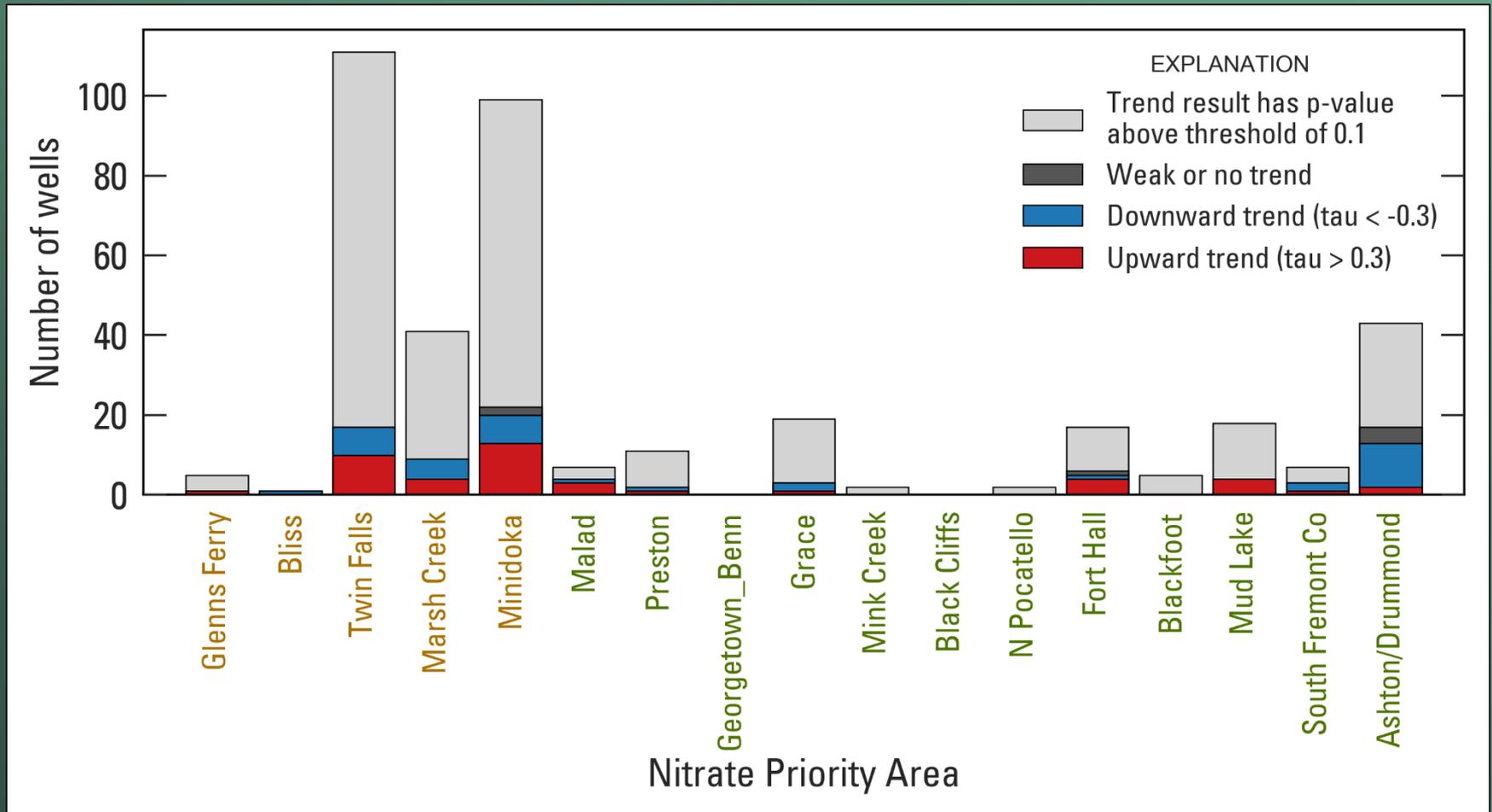
Nitrate trends since 1990 in eastern NPAs. Idaho is divided into north, west, south, and east regions for display.

# Trends



Nitrate trends since 1990 in wells with more than three detections. NPA names are colored by regional groups as shown on the previous maps.

# Trends



Nitrate trends since 1990 in wells with more than three detections. NPA names are colored by regional groups as shown on the previous maps.

# Trends

- Trend results for 63 analytes are available at <https://doi.org/10.5066/F72V2FBG>.

# Correlations

- Spearman rank correlation test
  - *rho*: strength of monotonic association
  - *p*: statistical significance
- 63 analytes

## Correlations with nitrate, $p < 0.1$

Positive Correlations		Negative Correlations	
Analyte	Spearman $\rho$	Analyte	Spearman $\rho$
uranium	0.72	cadmium	-0.44
TDS	0.59	iron	-0.36
chloride	0.52	aluminum	-0.35
sulfate	0.42	manganese	-0.30
tetrachloroethylene	0.32	carbon tetrachloride	-0.28
barium	0.28	thallium	-0.24
atrazine	0.25	beryllium	-0.22
simazine	0.22	antimony	-0.19
arsenic	0.21	mercury	-0.17
selenium	0.21	zinc	-0.16
radium-226	0.20	chromium	-0.15
alachlor	0.18	silver	-0.13
trichloroethylene	0.15	copper	-0.02
$\alpha$ particle	0.13		
nitrite	0.09		
flouride	0.09		

# Correlations

A selection of highly correlated pairs;  $p < 0.1$

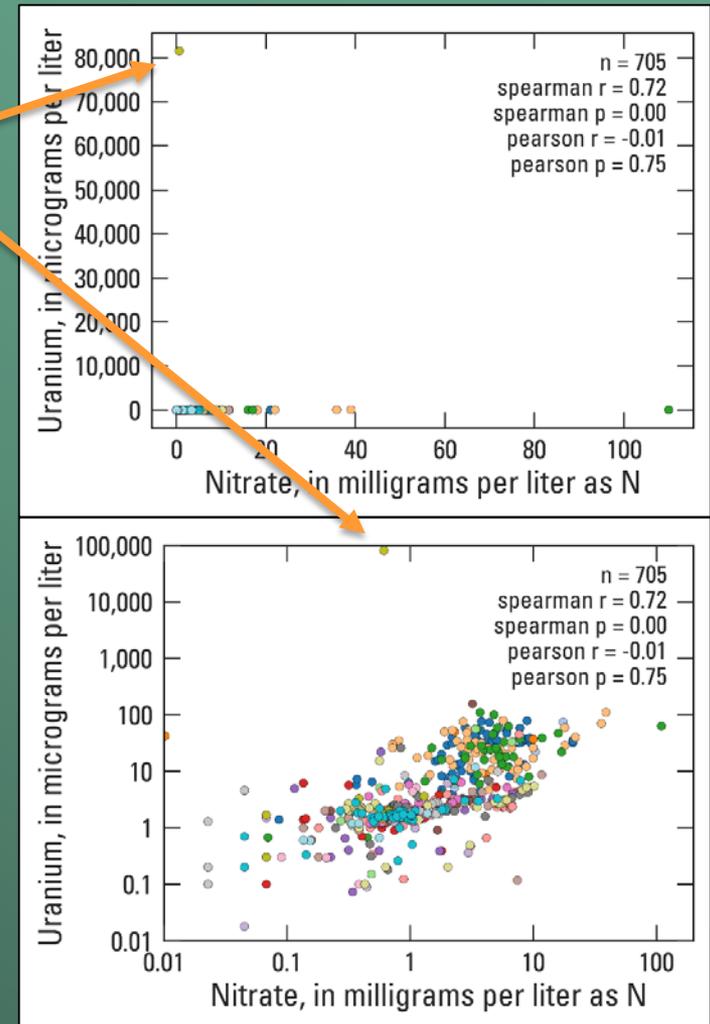
Analyte 1	Analyte 2	Spearman $\rho$
carbon tetrachloride	trichloroethylene	0.91
sulfate	TDS	0.83
chloride	TDS	0.80
benzene	toluene	0.77
cadmium	zinc	0.76
uranium	TDS	0.74
uranium	atrazine	0.70
chloride	sulfate	0.67

Analyte 1	Analyte 2	Spearman $\rho$
cyanide	sulfate	0.66
antimony	copper	0.66
selenium	sulfate	0.63
cadmium	selenium	0.58
aluminum	zinc	0.57
selenium	TDS	0.57
antimony	arsenic	0.54
antimony	silver	0.53

# Correlations

- Spearman is resistant to outliers compared to Pearson
- Outliers could be natural result of long-tailed distributions or misreported values

outlier

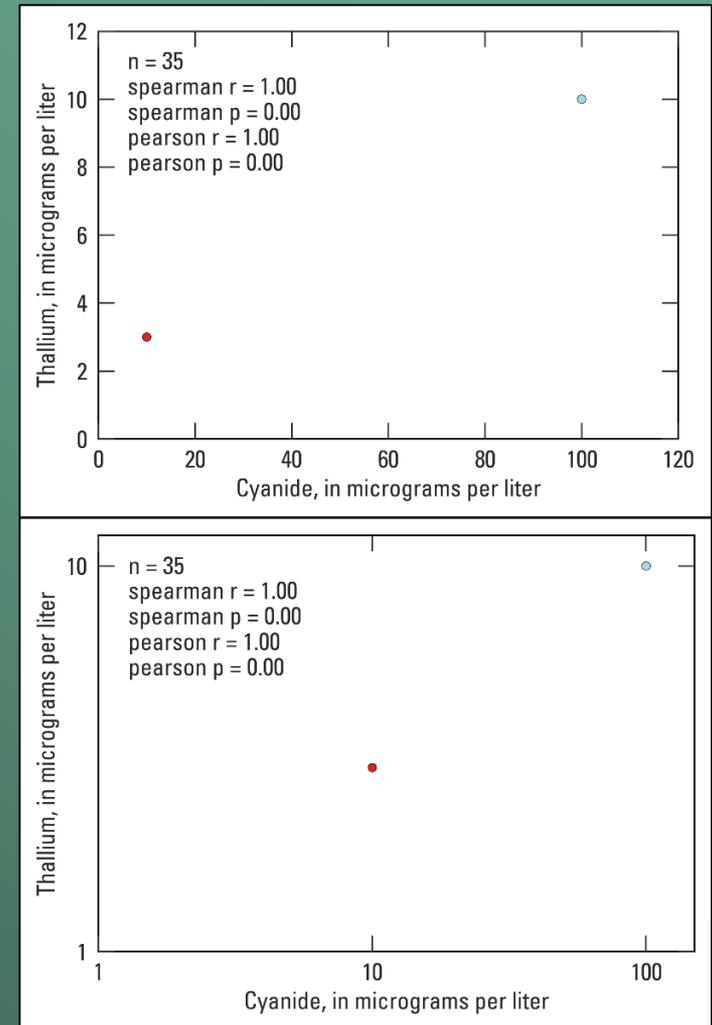


Uranium and Nitrate concentrations. Colors represent different sampling locations.

Results

# Correlations

- However, some spurious results appear
- May result from non-detections that were reported at concentrations equal to (or a fraction of) the detection limit, rather than flagged as not detected
- Smaller sample sizes more prone to producing spurious results

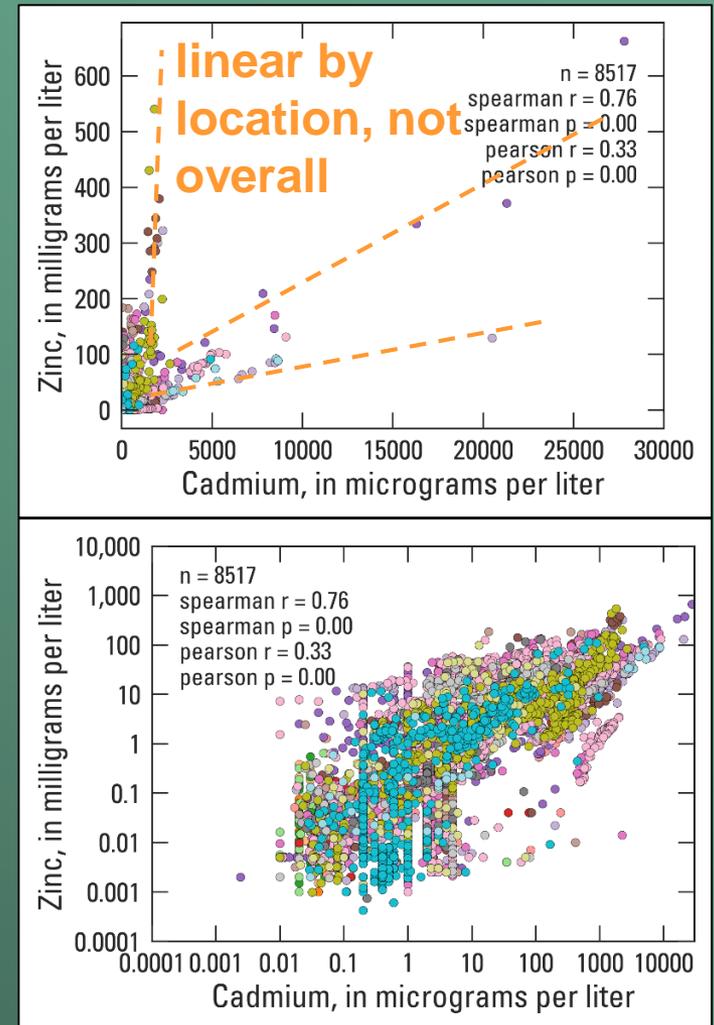


Thallium and Cyanide concentrations. Colors represent different sampling locations.

Results

# Correlations

- Grouping all together may miss or understate regional correlations that do not occur everywhere
  - Every color on the figures represents a different sampling location



Zinc and Cadmium concentrations. Colors represent different sampling locations.

Results

# Correlations

- Correlation results for 63 analytes are available at <https://doi.org/10.5066/F72V2FBG>.

# Links to database and results

- Data release:  
<https://doi.org/10.5066/F72V2FBG>
  - Database
  - Monitoring location shapefile
  - Exceedances
  - Trends
  - Correlations
- For R code: [shundt@usgs.gov](mailto:shundt@usgs.gov)