E.O. Lawrence Berkeley National Laboratory
University of California
Environmental Restoration Program

United States Department of Energy

Analysis of Background Distributions of Metals in the Soil at Lawrence Berkeley National Laboratory

By

David Diamond, David Baskin, Dennis Brown, Loren Lund, Julie Najita
and Iraj Javandel

Lawrence Berkeley National Laboratory

ENVIRONMENTAL RESTORATION PROGRAM

June 2002
Revised April 2009

This work was done at the Lawrence Berkeley National Laboratory, which is operated by
the University of California for the U. S. Department of energy
under Contract DE-AC02005CH11231

DISCLAIMER

The background metals concentrations presented in this report were derived from samples collected at the Ernest Orlando Lawrence Berkeley National Laboratory in the Berkeley/Oakland Hills of Alameda County, California. The natural background concentration of a chemical in rock and soil can vary substantially between and even within sites, depending largely on the specific rock units present and the nature of the parent material from which the soil formed. Caution should therefore be exercised when extrapolating the results described in this report to other sites, since the data may not be comparable to background conditions in other areas, particularly those where soil or rock types differ. Reference herein to any specific commercial products, process, or service by trade name, trademarks, manufacture, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California.

TABLE OF CONTENTS

DISCLAIMER ii

1 PURPOSE OF THE BACKGROUND ANALYSIS 1

2 DATA USED IN THE ANALYSIS 2

3 STATISTICAL METHODS 3

3.1 Identification of Outliers 4
3.2 Distribution Shape 5
3.3 Assessment of Depth-Related Concentration Variations 9
3.4 Summary Statistics 10
3.5 Selection of Maximum LBNL Background Levels 12

4 RESULTS 13

4.1 Summary 13
4.2 Antimony 17
4.3 Arsenic 17
4.4 Barium 18
4.5 Beryllium 18
4.6 Cadmium 19
4.7 Chromium 19
4.8 Cobalt 20
4.9 Copper 21
4.10 Lead 21
4.11 Mercury 22
4.12 Molybdenum 23
4.13 Nickel 23
4.14 Selenium 24
4.15 Silver 24
4.16 Thallium 25
4.17 Vanadium 26
4.18 Zinc 26

5 CONCLUSIONS 27

6 ACKNOWLEDGEMENTS 29

7 REFERENCES 30

LIST OF TABLES

Table 1. Summary of Analytical Results for LBNL Background Data Sets 4

Table 2. Summary of Distribution Types and Outliers.....14

Table 3. Summary Statistics for Background Data Sets After Removal of Outliers.....16

Table 4. Comparison of Background Values to Other Background Estimates.....28

Table 5. Upper Estimates of Background Metals Concentrations for LBNL.....29

LIST OF FIGURES

Figure 1. Flow Chart for Evaluating Q-Q Plots and Determining Maximum Background Levels.....7

LIST OF APPENDICES

Appendix A. The Effect of Large Data Sets on the Shapiro-Francia Test A-1

Appendix B. Graphical Representations of Data by Metal.....B-1

Appendix C. Explanation of Statistical Methods.....C-1

Appendix D. Depth Versus Concentration Correlation Coefficients and Scatterplots... D-1

Analysis of Background Distributions of Metals in the Soil at Lawrence Berkeley National Laboratory

1 PURPOSE OF THE BACKGROUND ANALYSIS

As part of its Resource Conservation and Recovery Act (RCRA) Corrective Action Program (CAP), the Lawrence Berkeley National Laboratory (LBNL) Environmental Restoration Program conducted an evaluation of naturally occurring metals in soils at the facility. The purpose of the evaluation was to provide a basis for determining if soils at specific locations contained elevated concentrations of metals relative to ambient conditions. Ambient conditions (sometimes referred to as "local background") are defined as concentrations of metals in the vicinity of a site, but which are unaffected by site-related activities (Cal-EPA 1997).

Local background concentrations of 17 metals were initially estimated by LBNL using data from 498 soil samples collected from borings made during the construction of 71 groundwater monitoring wells (LBNL 1995). These concentration values were estimated using the United States Environmental Protection Agency's (USEPA's) guidance that was available at that time (USEPA 1989). Since that time, many more soil samples were collected and analyzed for metals by the Environmental Restoration Program. In addition, the California Environmental Protection Agency (Cal-EPA) subsequently published a recommended approach for calculating background concentrations of metals at hazardous waste sites and permitted facilities (Cal-EPA 1997). This more recent approach differs from that recommended by the USEPA and used initially by LBNL (LBNL 2002). The purpose of the 2002 report was to apply the recommended Cal-EPA procedure to the expanded data set for metals that was available at LBNL. This revision to the 2002 report has been updated to include more rigorous tests of normality, revisions to the statistical methods used for some metals based on the results of the normality tests, and consideration of the depth-dependence of some sample results. As a result of these modifications, estimated background concentrations for some metals have been slightly revised from those presented in the

original 2002 report. In cases where estimated background concentrations were reduced, site data were reviewed to assess whether significant changes to results of the RCRA CAP findings would result. This assessment indicated that no significant changes in RCRA CAP findings would result from the revisions.

2 DATA USED IN THE ANALYSIS

The Environmental Restoration Program collected and analyzed more than 1,600 soil samples for 17 metals during the RCRA Facility Investigation (LBNL 2000). These samples were collected primarily from soil borings. The borings, which were located throughout the facility, were generally made either for the purpose of collecting soil samples or for the installation of groundwater monitoring wells. A small proportion of the samples was collected from construction excavations. Soil samples were collected from depths ranging from the soil surface to a maximum depth of 180.5 feet.

Samples were collected and analyzed according to procedures described in LBNL's Quality Assurance Program Plan and Standard Operating Procedures (LBNL 1994a, 1994b). Metals were analyzed by USEPA Method SW-6010 (inductively coupled plasma optical emission spectrometer) with the exception of mercury, which was analyzed by USEPA Method SW-7471 (atomic absorption spectrometer). All analytical results were compiled in a relational database (4th Dimension, version 6.05, ACI, Inc., Cupertino, California) that contains the metal concentration for each sample as well as other relevant sample descriptors. The descriptors include, for example, a unique sample identifier, the sampling depth, and the geologic formation from which the sample was collected.

Several facilities (e.g., plating shops) at LBNL are known to have handled metals and were specifically investigated during the RCRA Facility Investigation to determine if a release of metals to the environment had occurred. Data from these facilities were excluded from the statistical analysis of background concentrations. In addition, data from several other locations were excluded because, based on the analytical results, there appeared to be some metals contamination.

After exclusion of these data, about 1,400 samples remained for each metal. The number of sample records and the frequency of detection for each metal are summarized in Table 1. For all metals except zinc, at least one sample had a concentration that was less than the reporting limit of the analytical method. These samples are recorded as “non-detects” in the LBNL background metals sample data set. For the purposes of conducting statistical calculations, the “non-detects” were assigned values of one-half their reporting limits where non-detects constituted less than half of the data set, and were assigned the reporting limit where they constituted more than half the data set.

3 STATISTICAL METHODS

The statistical methods used to analyze the data are generally consistent with guidance recommended by Cal-EPA (1997), although some modifications were made because of the large number of samples. Statistical calculations were performed, and graphs of the data were generated, using Microsoft Access, Microsoft Excel, and commercially available statistical software (*Resampling Stats in Excel* [Blank, *et al.*, 1999], S-Plus version 4.5 [MathSoft, Seattle, Washington]).

The LBNL background metals data sets described in the preceding section were evaluated using the statistical analysis steps described below.

Table 1. Summary of Analytical Results for LBNL Background Data Sets

Metal	Number of Observations	Number of Detections	Number of Non-detects	Detection Frequency (%)
Antimony	1,397	166	1,231	12
Arsenic	1,397	1210	187	87
Barium	1,397	1392	5	100
Beryllium	1,397	656	741	47
Cadmium	1,395	418	977	30
Chromium	1,403	1401	2	100
Cobalt	1,397	1392	5	100
Copper	1,400	1398	2	100
Lead	1,398	1043	355	75
Mercury	1,406	217	1,189	15
Molybdenum	1,394	200	1,194	14
Nickel	1,399	1392	7	99
Selenium	1,397	304	1,093	22
Silver	1,397	201	1,196	14
Thallium	1,398	192	1,206	14
Vanadium	1,397	1395	2	100
Zinc	1,396	1396	0	100

3.1 Identification of Outliers

As the first step in the statistical analysis, graphical plots of the LBNL background data set for each metal were examined to determine whether outliers were present. Outliers are defined as sample values that are unusually large (or small) and that are obvious deviations from the background distribution. Outliers may arise from sampling or analytical errors (e.g. cross contamination of a field sample or miscalibration of an analytical instrument), transcription errors, classification errors (e.g., a sample that should have been associated with a facility known to have handled metals was assigned to a nearby location instead), or the presence of contaminated samples in the background data set. The purpose of this evaluation was to eliminate, to the extent possible, the effect of such errors on the estimation of background concentrations. Section 4 includes metals-specific discussions of the outlier records removed from the data set.

3.2 Distribution Shape

After removal of outliers, graphical plots of the background data set (Appendix B) were used to identify the shape of the data distributions for individual metals. Histograms, box plots, and quantile-quantile (Q-Q) plots were plotted for each metal. These plots were used to assess distribution symmetry and to compare the observed distribution to known parametric probability distributions such as the normal distribution. In addition, box plots were created for each of the main geologic units present at LBNL. The geologic units are the Great Valley Group, consisting of marine mudstones, shales and sandstones, the Orinda Formation, consisting of non-marine mudstones, siltstones and sandstones, and the Moraga Formation and Mixed Unit, which consist primarily of volcanic rocks. In addition, a separate category was created for unclassified units (typically artificial fill, young alluvium, soil, etc) and multiple formations (i.e. samples where the formation could not be determined or material containing fragments from multiple formations). These geologic units are described in significantly more detail in the RFI Report (LBNL, 2000).

Although use of a single distribution test (e.g., the Shapiro-Francia or Shapiro-Wilk tests) is the method recommended by Cal-EPA (1997), application of such a test can lead to inaccurate conclusions because it may classify very small, possibly random, deviations from the expected distribution as statistically significant differences in cases where the sample size is large (Gibbons 1994), as it is in this study. The adjoining text box and Appendix A provide additional discussions of this problem. The Jarque-Bera test (Jarque and Bera 1980), which is based on the skewness and kurtosis of the distribution, was therefore used to assess normality since it is considered a more reliable method than the Shapiro-Francia and Shapiro-Wilk tests when the dataset is large. This approach was carried out by using SolverStat Excel Add-In (Comuzzi and others 2004) and Statistics with Java (Satman 2008).

Effect of Large Sample Sizes on Statistical Tests

Due to the large sample sizes used in this study, tests for deviations from normality such as the Shapiro-Francia (1972) test, the Shapiro-Wilk test, and other tests of this nature are not appropriate and give misleading information about the distribution being sampled. This is a well-known phenomenon in hypothesis testing and occurs because most test statistics have standard errors that decrease inversely with the square root of the sample size. Hence, for data sets with a large number of samples, very small deviations are detected as significant by the test. This can result in an incorrect rejection of the null hypothesis. Namely, the statistical test indicates that the data are not normally distributed when in fact they are.

Based on the symmetry of the distribution and the conformity of Q-Q plots with a normal or lognormal distribution, the data distribution for each metal was classified as potentially normal, lognormal, or of a type for which the distribution could not be determined. The finding was subsequently tested using the Jarque-Bera method. The flow chart shown on Figure 1 demonstrates how the Q-Q plots were used to classify the distributions.

- For data sets where the histograms and density estimate plots appeared symmetrical, and the raw data in the Q-Q plot formed a straight line, the distributions were considered to be potentially normally distributed and consistent with a single population. The data sets were then subjected to the Jarque-Bera test. For those data sets where the test did not reject the null hypothesis of normality, the entire distribution was then used to calculate LBNL background, using parametric methods, as discussed in Section 3.3.

- For data sets where histograms and density estimate plots appeared to be asymmetrical, and Q-Q plots were non-linear, the data were evaluated to determine whether they were distributed lognormally. These data sets were first transformed by calculating the natural log of each value in the distribution ("log-transformed") and were then replotted on histograms, density estimates, and Q-Q plots. The distribution was classified as potentially lognormal and consistent with a single population if it appeared symmetrical and if points in the Q-Q plot followed a straight line. The transformed data sets were then subjected to the Jarque-Bera test. For those data sets where the test did not reject the null hypothesis of normality, the entire distribution would have been used to calculate LBNL background using parametric methods. However, all of the transformed data sets failed the test for normality so this method was not used.

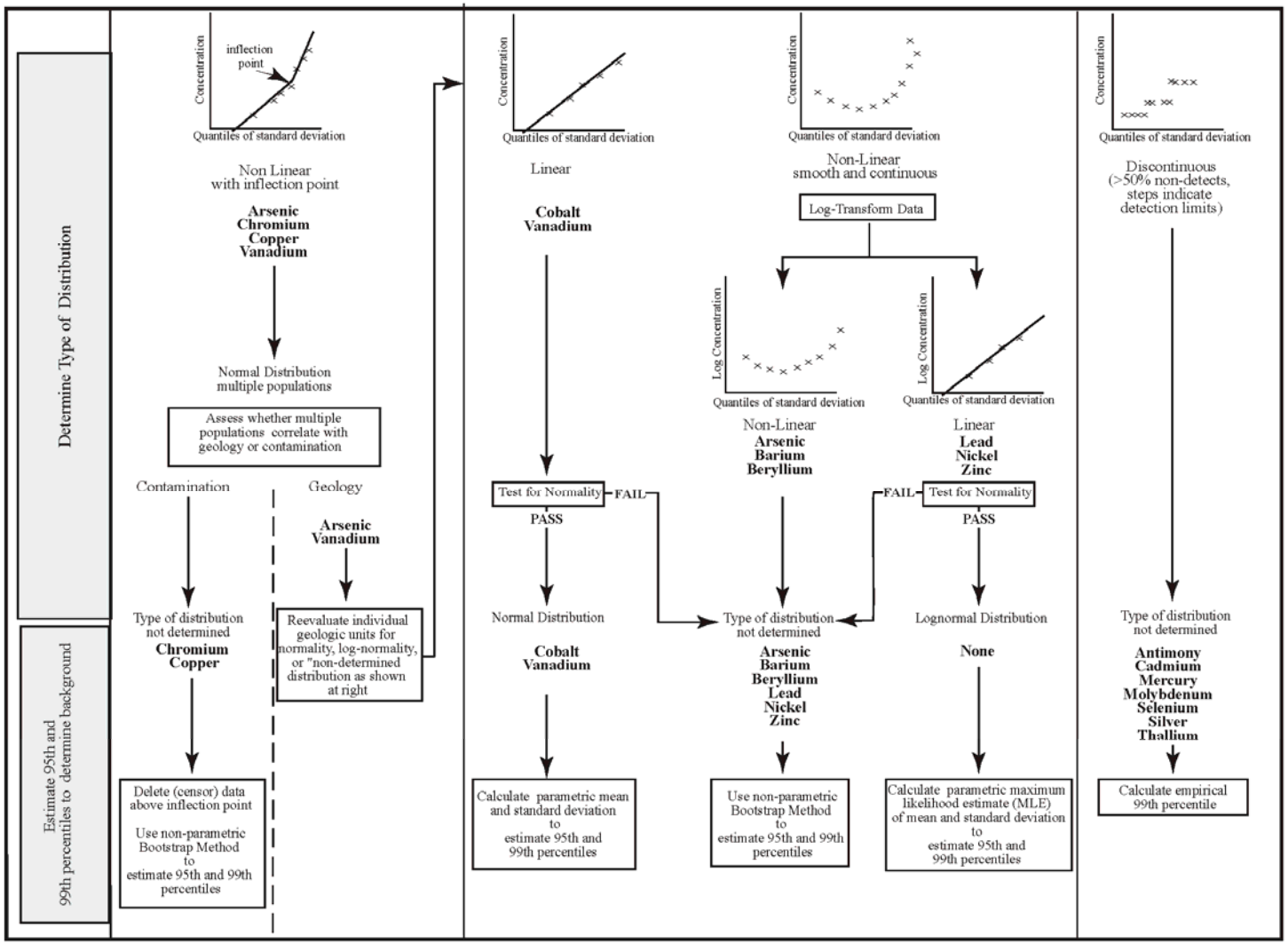


Figure 1. Flow Chart for Evaluating Q-Q Plots and Determining Maximum Background Levels

In some cases, the Q-Q plots were non-linear for both raw and log-transformed data, or the data set failed the Jarque-Bera test. Three types of non-linearity were observed: distributions that were continuous but non-linear with no distinct linear segments; distributions that were continuous but with two linear segments; and, distributions that were discontinuous (discontinuities were caused by a high proportion of non-detects in the data set). See the adjoining text box for additional discussion of the procedures used to interpret non-linear distributions. The distributions that were neither normal nor lognormal were treated as follows:

Interpretation of Non-Linear Distributions

Cal-EPA (1997) guidance specifies that where multiple distributions are present in ambient background data sets, the ambient distribution with the lowest range of concentrations should be used to determine site-specific background (except when soils at a site are derived from different parent materials with different metals distributions). Two (or more) distributions are generally interpreted to be present in a data set when the upper end of a Q-Q plot shows a sharp break in slope. In addition, a long tail at higher concentrations shown in a histogram or a density estimate may also indicate multiple distributions.

For metals where Q-Q plots suggested the presence of multiple distributions that could not be correlated with LBNL site geology, the distribution with the lowest range of concentrations was used to determine the site-specific background concentration.

- For data sets with a Q-Q plot that formed a continuous non-linear distribution for both raw and log-transformed data, the entire distribution was interpreted to represent the background distribution (of possibly overlapping populations from different geologic units). For these distributions a non-parametric statistical method ("bootstrapping" described in Section 3.4 and Appendix C) was used to calculate LBNL background.

- For data sets with Q-Q plots that formed a continuous distribution of two linear segments the data set was interpreted to represent two distinct populations (i.e., each linear segment represents a separate population within the distribution). In these cases, the data were examined to ascertain whether the multiple populations were associated with systematic differences between geologic units. For two metals, chromium and copper, since no such association could be made, the background distribution was assumed to be the distribution with the lowest concentration values. The distribution of the lowest concentration values was re-examined to determine which statistical method (parametric or non-parametric) was most appropriate for determining the LBNL background for that metal.

- Data sets with discontinuous Q-Q plots, generally had a high proportion (i.e., >50 percent) of non-detect samples. Per USEPA (2000), distribution shapes were not determined for data sets with more than 50 percent non-detects. An

empirical method of determining LBNL background was used for these unclassified distributions as described in Section 3.5.

3.3 Assessment of Depth-Related Concentration Variations

The primary depth-related variations in metals concentrations are expected to be due to the occurrence of different rock and soil types at different depths at most sampling locations, and were addressed as described in Section 3.2 above. Secondary depth-related factors that could potentially create depth-related variations in chemical composition independent of geologic rock/soil type are:

- Due to the immobility and low solubility of most metals, near-surface rock and soil in the developed areas of LBNL may have been impacted by pervasive anthropogenic contamination (particularly lead contamination) derived from building materials and coatings, road runoff, and atmospheric fallout;

- Near-surface (typically upper 1 to 2 feet) native rock and soil may have been subjected to weathering processes for a period sufficiently long to result in chemical compositions that differ from underlying unweathered materials. However, such a correspondence is unlikely to apply to most of the developed areas of LBNL, and would also be difficult to discern in the sample database, because cut and fill operations have removed or buried the weathered zone at most locations;

If the secondary depth-related factors discussed above affect metals concentrations, they would generally be expected to be observable on the Q-Q plots as departures to high values from linear distributions. Such a deviation was only noted in the chromium dataset and the data contributing to the deviation were therefore excluded from the analysis, as discussed further in Section 4.7. Nevertheless, observation of the Q-Q plots may not result in identification of all depth-related multiple populations, especially for metals where distributions were interpreted to be non-normal. To assess this possibility, depth versus concentration scatterplots were examined and depth/concentration correlation coefficients were calculated for each metal (Appendix D). Correlation coefficients were very low (absolute value less than 0.2) for all metals. From inspection of the scatterplots (presented in Appendix D), only lead appears to have a

possible depth/concentration relationship, in that all samples with concentrations exceeding 35 mg/kg (10 samples out of 1,395) occur at depths of 5.1 feet or less, suggesting that a small part of the lead database includes anthropogenic contamination, as is typical of older developed areas. This finding is addressed in Section 4.10.

3.4 Summary Statistics

Summary statistics were calculated for the distribution of each metal using spreadsheet software (i.e., Microsoft Excel 97) after conducting the following operations:

- Removing outliers
- Substituting one-half the detection limit for each non-detect value for datasets with less than 50% non-detect values
- Substituting the detection limit for each non-detect value for datasets with more than 50% non-detect values
- Determining the shape of the distribution
- Selecting the distributions with the lowest concentration ranges as background for datasets with more than one population.

The summary statistics included:

- Arithmetic means, arithmetic standard deviations, and medians for the following three parameters: detection limits; detected concentration values; and all concentration values.
- Estimates of the 95th and 99th percentiles of each population.

The arithmetic mean is the sum of the observed values divided by the number of values observed. The standard deviation is a measure of how much the results differ from the arithmetic mean. The median is the middle of a distribution; half the values in the distribution are greater than the median and half are less than the median.

The upper (95th and 99th) percentiles of the population were estimated by one of the following methods, depending on the type of distribution (refer to Appendix C for further details and example calculations):

For normally distributed metals, the sample average ($\overline{X}$) and the sample standard deviation (*s)* were used as estimators of the population mean (μ) and population standard deviation (σ). These sample statistics were then used to calculate estimated upper percentile values of each population for both the 95th percentile ($\overline{X}$ + 1.645 *s*) and for the 99th percentile ($\overline{X}$ + 2.33 *s*).

Generally, for lognormally distributed metals, values for μ and σ are estimated by the method of maximum likelihood. The maximum likelihood estimate (MLE) for μ is the arithmetic mean of the log concentrations ($\hat{\mu}$), whereas the MLE for σ is the sample standard deviation of the log concentrations ($\hat{\sigma}$). These MLEs are then used to calculate estimated upper percentile values for the 95th percentile (exp[$\hat{\mu}$+1.645$\hat{\sigma}$]) and the 99th percentile (exp[$\hat{\mu}$+2.33$\hat{\sigma}$]). However, all metals with apparent lognormal distributions failed the Jarque-Bera test for lognormality. As a result, the calculation of 95th and 99th percentile estimates using this methodology was not permissible.

For metals where both log-transformed and non-transformed Q-Q plots show non-linear continuous distributions, it is suspected that more than one population is present. In these cases, estimates for μ and σ were determined by the method described by Cal-EPA (1997). Upper percentiles were estimated by bootstrapping (a nonparametric statistical method) using the commercially available add-in for Microsoft Excel, *Resampling Stats in Excel* (Blank, *et al.,* 1999). Bootstrapping is a technique that resamples observations from data in order to compute non-parametric estimates and standard errors. Here, bootstrapping was used to compute estimates for the 95th and 99th percentiles of the concentration distributions (Efron and Tibshirani 1993). All values are reported to two significant digits.

For metals with a high percentage (i.e., >50 percent) of non-detects and a discontinuous distribution, an empirical method (Section 3.5) was used to determine the LBNL background concentration.

3.5 Selection of Maximum LBNL Background Levels

A maximum LBNL background level was selected for each metal based on the results of the statistical analyses discussed above. The selected maximum LBNL background level is the concentration value against which site concentration data are compared to determine whether the data represent site contamination. Sample concentrations greater than the maximum background levels are categorized as likely site contamination, whereas sample concentrations less than or equal to the maximum background levels are categorized as ambient conditions.

As discussed in Cal-EPA, 1997, simple estimates of upper percentiles (e.g. the 95th or 99th percentiles) are appropriate measures of the upper range of ambient conditions for large, well-distributed data sets. Due to the very large size of the LBNL data sets, calculated estimates of the 99th percentile of the data sets were used as the maximum LBNL background level for metals with low numbers of non-detects and well-defined distributions. For data sets that did not appear to have well-defined normal or lognormal distributions but had only a low proportion of non-detects, upper percentile estimates were calculated using bootstrapping.

For metals with a large number of non-detects (>50 percent) and/or poorly defined distributions, the statistical parameters (mean and standard deviation) used in calculating percentile estimates are considered unreliable. Therefore, empirical 99th percentile values were calculated as follows: 1) all non-detect values were replaced with the analytical method reporting limit (in place of one-half the reporting limit used for the statistical calculations discussed above); 2) the data were ranked and the 99th percentile value was selected based on the ranking order; 3) Non-detect values with reporting limits that exceeded the empirical 99th percentile values may have resulted in anomalously high 99th percentile values. Therefore, all non-detect values with reporting limits greater than or equal to the 99th percentile values were deleted from the dataset, and the empirical 99th

percentiles recalculated. Empirical 95th percentile values were not calculated because the large number of non-detect values that exceeded the apparent 95th percentile values indicated a high degree of uncertainty for this parameter. The 99th percentile empirical value for each metal was then used as the maximum LBNL background level.

4 RESULTS

4.1 Summary

All metals had at least one outlier that was excluded from the background statistical analyses. The criteria that were used to exclude outliers and the number of excluded sample values are listed in Table 2. In most cases the outliers represented unusually large values. However, for nickel and selenium, the outliers included very small values, many or all of which were non-detects. More detailed information about the values that were excluded is contained in the metal-specific discussions in this section.

Appendix B shows data plots of the sample distributions after removal of identified outliers. Review of the data plots indicated that more than one distribution appeared to be present in several metals data sets. For two of the metals, chromium and copper, it was unclear whether the multiple distributions were due to differences between the geologic classifications and/or to regional or local anthropogenic sources of these metals (e.g., air pollution, piping, or paint). Therefore, to ensure that the chromium and copper data did not include potentially contaminated samples, only those data representing the distribution that contained the lowest set of concentration values was used to calculate the 95th and 99th percentiles for these two metals.

The distribution type for each metal is listed in Table 2. Eight of the 17 metals had more than 50 percent non-detects (Table 2), so their distribution types could not be determined and percentile values were calculated empirically. Of the remaining nine metals, one (cobalt) fit a normal distribution and two (vanadium and copper) appeared to constitute two populations. For each of these two datasets, a small number of samples appeared to represent a secondary population with relatively high concentrations, apparently indicating the presence of contamination. After exclusion of the data

representing the secondary population, the vanadium dataset fit a normal distribution. None of the metals fit a lognormal distribution. Concentrations of the remaining metals and the copper data representing the primary population did not fit either distribution and/or failed tests for normality and lognormality, so a non-parametric approach was used to generate percentile estimates.

Table 2. Summary of Distribution Types and Outliers

Metal	Criteria for Removing Outliers (mg/kg)	Total Outliers Excluded	Outliers that were Non-Detects	After Removal of Outliers		
				Distribution Type [1]	Number of Observations[3]	Number of Non-detects[3]
Antimony	> 35	2	1	ND-N	1,395/1124	1,230/960
Arsenic (all)	> 50	2	0	ND-F	1,395	187
Arsenic (Great Valley Group)	> 50	1	0	ND-F	178	3
Arsenic (other units)	> 50	1	0	ND-F	1217	184
Barium	> 500	7	0	ND-F	1,390	5
Beryllium	> 1.5 and MDL ≥ 5	3	1	ND-N	1394/1,128	740/474
Cadmium	> 8	3	0	ND-N	1,392	977
Chromium	> 400	3	0	ND-F	1,400	2
Cobalt	≥ 30	10	0	Normal	1,387	5
Copper	≥ 200	4	0	ND-F	1,396	2
Lead	> 100	3	0	ND-F	1,395 (1023)[4]	355 (272)[4]
Mercury	≥ 1.3	5	0	ND-N	1,401	1,189
Molybdenum	≥ 25	4	0	ND-N	1,390390/1,120	1194/924
Nickel	≤ 5	7	7	ND-F	1,392	0
Selenium	≤ 0.1 or > 10	21	12	ND-N	1,376	1,081
Silver	> 10	4	0	ND-N	1,393/1,366	1,196/1,176
Thallium	> 60	4	1	ND-N	1,394/1,371	1,205/1,182
Vanadium	> 120	2	0	Normal[2]	1,395	2
Zinc	> 200	4	0	ND-F	1,392	0

[1] Because the background data sets are large (> 1,000 samples) the determination of distribution type was based on a visual inspection of the plotted data and tests for normality after outliers had been excluded, as described in the text.

[2] Based on the distribution of the lowest population of an apparent two-population data set.

[3] Where two quantities separated by "/", the second quantity indicates the number of observations or non-detects remaining after removal of non-detects with high reporting limits, as explained in Section 3.4.

[4] Second quantity in parentheses indicates the number of observations or non-detects remaining after removal of data points representing samples shallower than 5.1 feet, as explained in Section 3.3.

ND-N – Not determined due to high percentage of nondetects (see Ttble 1).

ND-F – Not determined due to failed tests for normality or and lognormality.

MDL –Method detection limit.

Metal-specific information is presented in the following sections. A summary of the background statistics (i.e., range, arithmetic mean, median, standard deviation, 95th and 99th percentiles) for each metal is provided in Table 3. Plots of the individual metals data sets are shown in Appendix B.

Table 3. Summary Statistics for Background Data Sets After Removal of Outliers

	Reporting Limits for Non-Detects [1] (mg/kg)			Detected Concentration Values [1] (mg/kg)			Arithmetic Mean (mg/kg)	Standard Deviation (mg/kg)	Median (mg/kg)	Percentiles		Percentile Estimation Method[2]
Metal	**min**	**med**	**max**	**min**	**med**	**max**				**95th**	**99th**	
Antimony	0.01	5	5	0.7	1.6	22	-- [3]	--	--	--	< 6	Empirical
Arsenic (all)	0.033	0.5	2.0	0.3	4.8	42	5.5	5.4	4	17	28	Bootstrap
Arsenic (Great Valley Group)	0.2	0.2	0.25	1.2	10	42	11.6	7.2	10	28	42	Bootstrap
Arsenic (other units)	0.033	0.5	2.0	0.3	4.2	30	4.7	4.4	3.5	14	24	Bootstrap
Barium	0.5	5	5.0	1.7	1	490	130	71	120	280	410	Bootstrap
Beryllium	0.004	0.5	2.0	0.06	0.5	1.2	-	-	-	--	1.0	Empirical
Cadmium	0.009	0.5	5.0	0.05	1.1	7.5	--	--	--	--	5.6	Empirical
Chromium (total) [4]	0.5	0.5	0.5	1.7	56	144	58	26	56	100	120	Bootstrapping
Cobalt	0.5	5	5.0	0.92	14	29	14	4.8	14	22	25	Parametric
Copper [4]	0.25	0.6	1.0	2.2	31	69	32	13	31	58	63	Bootstrapping
Lead	0.043	5	25	0.66	7	84	7.0	6.5	5.4	17	43	Bootstrapping
Lead (deep) [6]	0.043	2.5	10	0.66	7	31	6.4	4.9	5.1	16	24	Bootstrapping
Mercury	0.0001	0.2	0.5	0.023	0.1	0.82	--	--	--	--	0.42	Empirical
Molybdenum	0.016	2.5	25	0.26	1.7	14	--	--	--	--	4.8	Empirical
Nickel	NA[5]	NA	NA	6.0	57	380	68	46	57	164	272	Bootstrapping
Selenium	0.2	0.5	5.0	0.25	1.7	9.1	--	--	--	--	4.9	Empirical
Silver	0.012	1	10.0	0.2	0.5	7.7	--	--	--	--	2.9	Empirical
Thallium	0.098	5	50	0.16	6	20	--	--	--	--	10	Empirical
Vanadium	0.5	0.5	0.5	0.79	46	120	46	19	46	77	90	Parametric
Zinc	NA	NA	NA	3.8	61	190	64	23	61	110	140	Bootstrapping

[1] min – minimum, med – median, max – maximum

[2] Estimation methods . (see text for further explanation):
Bootstrap: calculated using bootstrapping method (also referred to as resampling).
Parametric: calculated for normal distributions.
Empirical: value based on ranked sampled data.

[3] "--" - not calculated because of the high percentage of non-detects (see Table 2).

[4] Summary statistics calculated on concentration values less than 150 mg/kg and 70 mg/kg for chromium and copper, respectively.

[5] NA -- not applicable because metal was detected 100-percent of the time.

[6] Deep lead data set excludes shallow data (<5.1 feet) due to potential anthropogenic contamination.

4.2 Antimony

One antimony sample with a concentration greater than 35 mg/kg was excluded as an outlier from the background analysis. This sample was collected at a depth of about 70 feet below ground surface (bgs) and thus is unlikely to represent contamination. One non-detect sample with a method detection limit of 50 mg/kg was also excluded as an outlier. Due to the large proportion of non-detects in the data set, no reliable statistical parameters could be calculated. Therefore, a value equal to the reporting limit was substituted for each non-detect result and the empirical 99th percentile of the resulting dataset was calculated. However, the data point corresponding to the 99th percentile was a reporting limit (20 mg/kg) substituted for a non-detect result. Only one sample with detected antimony concentrations exceeded this non-detect value and 19 data points at this value were reporting limits substituted for non-detect values. Therefore, all substitutions for non-detect results that were greater than or equal to this 99th percentile value were excluded from the data set and the 99th percentile recalculated. After this deletion, the data point corresponding to the 99th percentile was also a reporting limit (10 mg/kg) substituted for a non-detect result. Only five samples with detected antimony concentrations exceeded the recalculated 99th percentile value and 252 data points at this value were reporting limits substituted for non-detect values. Therefore, all substitutions for non-detect results that were greater than or equal to this value were also excluded and the 99th percentile recalculated again. The resulting 99th percentile value (6 mg/kg) was used as the LBNL maximum background value for antimony since the majority of data points exceeding that value represented detected concentrations.

4.3 Arsenic

Differences in arsenic concentration ranges were apparent for the different geologic classifications as shown by box plots of the data (Appendix B). In particular, concentrations measured in the Great Valley Formation are about double those of the other units. The variation in median arsenic concentrations between the geologic classifications may partly explain why the entire background data set fits neither a normal nor a lognormal distribution.

Due to the noticeable difference between arsenic concentrations for the Great Valley Group and for the other geologic units, separate data sets were created for these two classifications. Upon separating the arsenic data into these two data sets, two data points were determined to represent outliers in the separated data sets. These data points are 63 mg/kg for the Great Valley Group (next lowest value 42 mg/kg) and 51 mg/kg for the combined other geologic units (next lowest value 30 mg/kg). These points were removed from the arsenic data set and the arsenic statistics were calculated. After removal of the outliers, the data set did not appear to fit either a normal or a lognormal distribution. Because the data fit neither distribution, the non-parametric bootstrap method was used to estimate the 95th and 99th percentiles.

4.4 Barium

Seven barium samples with concentrations greater than 500 mg/kg were excluded as outliers from the background analysis. After removal of the outliers, the data set did not appear to fit either a normal or a lognormal distribution. Because the data fit neither distribution, the non-parametric bootstrap method was used to estimate the 95th and 99th percentiles.

Box plots of the data by geologic formation show a great degree of overlap, although the Moraga/Mixed classification has a lower median concentration than the other three categories. The variation in median barium concentrations between the geologic classifications may partly explain why the entire background data set does not fit a normal or lognormal distribution.

4.5 Beryllium

Three beryllium samples were excluded as outliers from the background analysis. Two of these samples had concentrations greater than 1.5 mg/kg and one was a non-detect sample with an elevated reporting limit. The two samples with detects greater than 1.5 mg/kg were collected at depths of 30 and 75 feet bgs and thus are unlikely to represent contamination.

After exclusion of the identified outliers, 53% of the samples remaining in the background data set were non-detects. Due to the large proportion of non-detects, no reliable statistical parameters could be calculated. Therefore, a value equal to the reporting limit was substituted for each non-detect result. However, the highest 19 values in the dataset were reporting limits substituted for non-detect results. Therefore, these non-detect results were deleted from the dataset. The empirical 99th percentile of the resulting dataset was calculated. However, the data point corresponding to the 99th percentile as well as 246 others were reporting limits substituted for the non-detect results. Therefore, all non-detect results that had detection limits greater than or equal to this 99th percentile value were excluded from the data set and the 99th percentile recalculated. The resulting 99th percentile value was used as the LBNL maximum background value for beryllium.

Box plots of the data divided by geologic classification show a large degree of overlap, although the Moraga/Mixed classification has a lower median concentration than the other classifications. The variation in median beryllium concentrations between the geologic classifications may partly explain why the entire background data set does not fit a normal or lognormal distribution.

4.6 Cadmium

Three cadmium samples with concentrations greater than 8 mg/kg were excluded as outliers from the background analysis.

Due to the large proportion of non-detects in the data set, no reliable statistical parameters could be calculated. Therefore, the empirical 99th percentile was used as the LBNL maximum background value for cadmium. There were 16 sample results with detected concentrations of cadmium greater than the 99th percentile, indicating a high degree of certainty in selecting the 99th percentile value for cadmium.

4.7 Chromium

Three chromium samples with concentrations greater than 400 mg/kg were excluded as outliers from the background analysis. All three appear to be unrelated to

contamination because samples collected both above and below them in the same soil borings had much lower concentrations than the samples identified as outliers. The outlier samples were collected at depths between 11 and 25 feet bgs.

After removal of the outliers, the plots of the chromium data show evidence of more than one distribution. For example, a fairly long tail above 200 mg/kg is observed in the density estimate of the distribution and the histogram, and the Q-Q plot shows a departure from linearity at about 150 mg/kg. This "knee" in the Q-Q plot suggests that low concentrations are normally distributed from a single population but that high concentrations are from another distribution(s). Among the 26 samples that exceed 150 mg/kg, 10 were collected from a single area near Building 75. Because of the departure from linearity in the Q-Q plot at 150 mg/kg and because one-third of the samples above this concentration are from the same unit, summary statistics and the 95th and 99th percentiles for chromium were calculated on the data set below 150 mg/kg. Although the distribution appeared to be normally distributed, the data below 150 mg/kg failed the Jarque-Bera test for normality; therefore the 95th and 99th percentiles for chromium were calculated by bootstrapping.

Box plots of sample data from the different geological classifications indicate that differences in median concentrations between geologic units are small. However, the Orinda formation and Unclassified/Multiple classification appear to contain many of the higher concentrations above 150 mg/kg, which were excluded from the dataset, as discussed above.

4.8 Cobalt

Ten cobalt samples with concentrations equal to or greater than 30 mg/kg were excluded as outliers from the background analysis. The ten samples were collected at depths between 0.8 and 70 feet bgs from different locations. .

After removal of the outliers, the plots of the cobalt data show excellent agreement with a normal distribution and passed the Jarque-Bera test for normality. The histogram and the frequency distribution have a classic "bell" shape and the Q-Q plot is

linear. Because the data fit a normal distribution the maximum likelihood estimation method was used to calculate the 95th and 99th percentiles. There was a large amount of overlap between samples assigned to different geologic classifications.

4.9 Copper

Four copper samples with concentrations equal to or greater than 200 mg/kg were excluded as outliers from the background analysis. The four samples were collected at depths between 2 and 85 feet bgs from different locations. .

After removal of the outliers, plots of the copper data show evidence of more than one distribution. For example, the Q-Q plot shows a departure from linearity at about 70 mg/kg. This result suggests that low concentrations are normally distributed from a single population but that high concentrations are from another distribution(s). Plots of the lognormally transformed data set were asymmetrical and nonlinear. Because of the departure from linearity in the Q-Q plot at 70 mg/kg, summary statistics and the 95th and 99th percentiles for copper were calculated on the data set below 70 mg/kg. Although the distribution appeared to be normally distributed, the data below 70 mg/kg failed the Jarque-Bera test for normality; therefore the 95th and 99th percentiles for copper were calculated by bootstrapping. Differences between sample data from different geologic classifications were small.

4.10 Lead

Three lead samples with concentrations greater than 100 mg/kg were excluded as outliers from the background analysis. The outlier samples were collected at depths between 0 and 5 feet bgs from possibly contaminated areas.

After removal of the outliers, plots of the lead data suggest that the data may be lognormally distributed. However, the log-transformed data failed the Jarque-Bera test for normality; therefore the 95th and 99th percentiles for lead were calculated by bootstrapping.

There was a large degree of overlap between the data plots for the different geologic classifications, although the Unclassified/Multiple Formation samples contain slightly higher concentrations. This result may occur because this category includes samples of surficial soil units that would be more susceptible than generally deeper rock units to regional anthropogenic sources of lead such as lead-based paint and air emissions from vehicles that used leaded gasoline. Also, as noted in Section 3.3, the depth/concentration scatterplot (Appendix D) showed that lead concentrations exceeding 39 mg/kg were only present in soil samples collected from 5.1 feet or shallower. This finding suggests that the lead data consists of two components, natural background plus pervasive near-surface anthropogenic contamination. To address this finding, separate lead background estimates were made for both the complete original background dataset, and for a truncated dataset that excludes values corresponding to samples from depths of 5.1 feet and less. The estimates of 95th and 99th percentiles for the complete dataset were 17 mg/kg and 43 mg/kg, respectively, and are interpreted to represent both natural and anthropogenic background. These values are considerably greater than the 95th and 99th percentile values of 16 and 24 mg/kg, respectively, for the truncated dataset which are interpreted to represent natural background. The new truncated dataset contains 1,023 data points, as compared to the original complete dataset of 1,395 data points.

It should be noted that for the purposes of this background study (i.e. discerning contamination derived from waste management practices at LBNL, such relatively pervasive surface contamination can be considered to be “anthropogenic background”.

4.11 Mercury

Five mercury samples with concentrations equal to or greater than 1.3 mg/kg were excluded as outliers from the background analysis. The outlier samples were both collected at depths of about 30 feet bgs.

After removal of outliers, 85% of the remaining data were non-detects. Due to this large proportion of non-detects, no reliable statistical parameters could be calculated. The empirical 99th percentile of the dataset was 0.43 mg/kg. There were 14 sample results with greater concentration values than the 99th percentile, only one of which was a non-

detect sample. The resulting 99th percentile value was used as the LBNL maximum background value for mercury.

4.12 Molybdenum

Three molybdenum samples with concentrations greater than or equal to 25 mg/kg and one non-detect result with a detection level of 25 mg/kg were excluded as outliers from the background analysis. The outliers were collected at depths between 4 and 165.5 feet bgs, each from different locations.

After exclusion of the identified outliers, 86% of the samples remaining in the background data set were non-detects. Due to the large proportion of non-detects, no reliable statistical parameters could be calculated. Therefore, a value equal to the reporting limit was substituted for each non-detect result and the empirical 99th percentile of the resulting dataset was calculated. However, the data point corresponding to the 99th percentile was a reporting limit (10 mg/kg) substituted for a non-detect result, and 20 of the 24 sample results with concentration values greater than the 99th percentile were also associated with non-detect results. Therefore, all non-detect results that were greater than or equal to this 99th percentile value were excluded from the data set and the 99th percentile recalculated. The resulting 99th percentile value was also a non-detect data point (5 mg/kg), and 250 of the 261 sample results with concentration values greater than the 99th percentile were also associated with non-detect results. Therefore, all substitutions for non-detect results that were greater than or equal to this value were also excluded and the 99th percentile recalculated again. The resulting 99th percentile value was used as the LBNL maximum background value for molybdenum.

4.13 Nickel

Seven outliers were excluded from the background calculations for nickel. All were non-detects with reporting limits less than or equal to 5 mg/kg.

After removal of the outliers, the plots of the nickel data suggest that the data may be lognormally distributed. However, the log-transformed data failed the Jarque-Bera test

for normality; therefore the non-parametric bootstrap method was used to estimate the 95th and 99th percentiles.

Box plots of sample data from the different geological classifications indicate that there is a large degree of overlap between the data sets of the different geologic classifications, although the Unclassified/Multiple classification and the Orinda formation samples contain a slightly greater proportion of samples at higher concentrations than the Moraga/Mixed classification samples.

4.14 Selenium

Six selenium samples with concentrations greater than 10 mg/kg were excluded as outliers from the background analysis. Fifteen selenium samples with concentrations less than 0.1 mg/kg or which were non detect at very low detection limits were also excluded as outliers. Among these, eight non-detects had reporting limits of less than 0.0001 mg/kg.

After exclusion of the identified outliers, 84% of the sample data remaining were non-detects. Due to this large proportion of non-detects, no reliable statistical parameters could be calculated. Therefore, a value equal to the reporting limit was substituted for each non-detect result and the empirical 99th percentile of the resulting dataset was calculated. The resulting 99th percentile value was used as the LBNL maximum background value for selenium. There were 14 sample results with concentration values greater than the 99th percentile, only one of which was a non-detect sample, indicating a high degree of certainty in the selected 99th percentile value.

4.15 Silver

Four silver samples with concentrations greater than 10 mg/kg were excluded as outliers from the background analysis. The outlier samples were collected at depths between 0 and 15 feet bgs.

After exclusion of the outliers, 91% of the samples remaining in the background data set were non-detects. Due to the large proportion of non-detects, no reliable

statistical parameters could be calculated. Therefore, a value equal to the reporting limit was substituted for each non-detect result and the empirical 99th percentile of the resulting dataset was calculated. However, the data point corresponding to the 99th percentile was a reporting limit substituted for a non-detect result, and 20 of the 26 sample results with concentration values equal to or greater than the 99th percentile were also associated with non-detect results. Therefore, all non-detect results that were greater than or equal to this 99th percentile value were excluded from the data set and the 99th percentile recalculated. The resulting 99th percentile value was used as the LBNL maximum background value for silver.

4.16 Thallium

Three thallium samples with concentrations greater than 60 mg/kg and one non-detect sample with a method detection limit of 50 mg/kg were excluded as outliers from the background analysis. Each excluded sample came from a different location. The samples were collected at depths of 6, 25, and 126.6 feet bgs.

After exclusion of the outliers, 91% of the samples in the background data set were non-detects. Due to the large proportion of non-detects, no reliable statistical parameters could be calculated. Therefore, a value equal to the reporting limit was substituted for each non-detect result and the empirical 99th percentile of the resulting dataset was calculated. However, the data point corresponding to the 99th percentile was a reporting limit substituted for a non-detect result, and 23 of the 34 sample results with concentration values greater than or equal to the 99th percentile were also non-detect results. Therefore, all data related to non-detect results that were greater than or equal to this 99th percentile value were excluded from the data set and the 99th percentile recalculated. The resulting 99th percentile value was used as the LBNL maximum background value for thallium.

4.17 Vanadium

Two vanadium samples with concentrations greater than 120 mg/kg were excluded as outliers from the background analysis. The outliers were collected from a depth of about 15 feet bgs.

After removal of the outliers, the plots of the vanadium data show good agreement with a normal distribution. The histogram and the frequency distribution have a classic "bell" shape and the Q-Q plot is generally linear except for data above approximately 90 mg/kg. The data less than 90 mg/kg passed the Jarque-Bera test so parametric methods assuming a normal distribution were used to calculate the 95th and 99th percentiles for the data less than 90 mg/kg.

Box plots of sample data from the different geological classifications indicate that there was a large degree overlap between the data sets from the different geologic classifications and differences between the classifications were small, although higher concentrations appear to be associated with the Unclassified/Multiple and Moraga formations. These higher concentrations could constitute a second population, and may be the cause of the apparent deviation from a normal distribution observed for values greater than 90 mg/kg.

4.18 Zinc

Four zinc samples with concentrations greater than 200 mg/kg were excluded as outliers from the background analysis. Each was collected from a different location at depths of about 4 and 15 feet bgs.

After removal of the outliers, the plots of the zinc data suggest that the data may be lognormally distributed. However, the log-transformed data failed the Jarque-Bera test for normality; therefore the non-parametric bootstrap method was used to estimate the 95th and 99th percentiles.

Box plots of the different geologic classifications indicate that there is a large degree of overlap between the different classifications, with the exception of the Great Valley Group, which had higher concentrations than those measured in the other units.

5 CONCLUSIONS

A comparison of current 95th and 99th percentile estimates for each metal with the 95% upper tolerance limit (UTL), which was used for background estimation in the 1995 LBNL background data set, and the regional background concentration reported by the U.S. Geologic Survey is shown in Table 4. Except for nickel, the 95th percentile estimates (or the 99th percentile when 95th percentile was not estimated) are very similar to previous estimates of background at LBNL that were calculated using the USEPA's 95% UTL guidance. Both the 95th and 99th percentiles for nickel are greater than the 1995 LBNL background estimate. This difference may be explained in part because the 1995 estimate assumed that the data were normally distributed; however, this type of distribution is not supported by the histogram and probability density plot for the nickel data (Appendix B), which show a long upper tail.

DTSC approved the use of 99th percentiles as the upper estimates of background in their approval of the original (2002) version of this report. The rationale for the use of the 99th percentile was that when the background data sets are large and well defined it is generally considered to be acceptable practice to select an estimate of the 99th percentile as representative of the upper range of ambient conditions Cal-EPA (1997). The LBNL dataset is large and well defined. The upper estimates (99th percentile) of background concentrations of metals in soil at LBNL are listed in Table 5.

Table 4. Comparison of Background Values to Other Background Estimates

Metal	Upper Estimate Regional Background [1]	95% UTL[2]	Percentile Estimates 95th	Percentile Estimates 99th
	mg/kg			
Antimony	1.8	5.5	--	<6
Arsenic (all)	18	19.1	17	28
Arsenic (Great Valley Group)	NA	NA	28	42
Arsenic (other units)	NA	NA	14	24
Barium	1,500	323.6	280	410
Beryllium	3	1.0	--	1.0
Cadmium	1.1 [3]	2.7	--	5.6
Chromium	160	99.6	100	120
Cobalt	23	22.2	22	25
Copper	76	69.4	58	63
Lead (all)	48	16.1	17	43
Lead (excluding shallow data)	48	16.1	16	24
Mercury	0.2	0.4	--	0.42
Molybdenum	3.3	7.4	--	4.8
Nickel	55	119.8	164	272
Selenium	1.1	5.6	--	4.9
Silver	2.3 [3]	1.8	--	2.9
Thallium	1.0 [3]	7.6 [4]	--	10
Vanadium	230	74.3	77	90
Zinc	150	106.1	110	140

[1] Regional statistics are based on Shacklette and Boerngen 1984 for the western conterminous United States
[2] LBNL 1995
[3] Data from Bradford et al. 1996 are used for analytes not available from Shacklette and Boerngen 1984
[4] 95% UTL for thallium was revised subsequent to the value originally presented in LBNL 1995
NA = Not Available

Table 5. Upper Estimates of Background Metals Concentrations for LBNL

Metal	Concentration mg/kg
Antimony	<6
Arsenic (other units)	24
Arsenic (Great Valley Group)	42
Barium	410
Beryllium	1.0
Cadmium	5.6
Chromium	120
Cobalt	25
Copper	63
Lead (all)	43
Lead (>5 feet depth)	24
Mercury	0.42
Molybdenum	4.8
Nickel	272
Selenium	4.9
Silver	2.9
Thallium	10
Vanadium	90
Zinc	140

6 ACKNOWLEDGEMENTS

This work was part of the LBNL Environmental Restoration Program that was supported by the Office of Environmental Management of the U. S. Department of Energy. The authors would like to thank Drs. Jinsong Chen and Nicolas Spycher for reviewing this report and providing constructive suggestions.

7 REFERENCES

Blank, S., C. Seiter, and P. Bruce, 1999. Resampling Stats in Excel. Resampling Stats, Inc., Arlington, Virginia.

Bradford, G.R., A.C. Chang, A.L. Page, D. Bakhtark, J.A. Frampton, and H. Wright 1996. Background Concentrations of Trace and Major Elements in California Soils, Kearney Foundation Special Report, Kearney Foundation of Soil Science, Division of Agriculture and Natural Resources, University of California, Riverside, 52 p.

Cal-EPA 1997. Selecting Inorganic Constituents as Chemicals of Potential Concern at Risk Assessment at Hazardous Waste Sites and Permitted Facilities, Prepared by Human and Ecological risk Division, Department of Toxic Substances Control, February.

Comuzzi, C., P. Polese, A. Melchior, R. Portanova, and M. Tolazzi, 2004. SOLVERSTAT: A New Utility for Multipurpose Analysis. An Application to the Investigation of Deoxygenated (CoII) Complex Formation in Dimethlysulfoxide Solution. Talanta 59 p67-80, http://www.freewebs.com/solverstat/

Efron, B. 1981. Censored data and the bootstrap. Journal of the American Statistical Association. 76 (374). June.

Efron, B. and R. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman and Hall, San Francisco.

Gibbons, R. 1994. Statistical Methods for Groundwater Monitoring. John Wiley & Sons, New York.

Jarque, M. and A.K. Bera, 1980. Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, Economics Letters, v. 6, p255-289.

LBNL 1994a. Lawrence Berkeley Laboratory Environmental Restoration Program Quality Assurance Program Plan, Lawrence Berkeley Laboratory, Berkeley, California, August.

LBNL 1994b. Lawrence Berkeley Laboratory Environmental Restoration Program Standard Operating Procedures, Lawrence Berkeley Laboratory, Berkeley, California.

LBNL 1995. Protocol for Determining Background Concentrations of Metals in Soil at Lawrence Berkeley National Laboratory (LBNL), August.

LBNL 2000. Draft Final RCRA Facility Investigation Report for the Lawrence Berkeley National Laboratory, Environmental Restoration Program, September.

LBNL 2002. Analysis of Background Distributions of Inorganic Elements in the Groundwater at Lawrence Berkeley National Laboratory, Environmental Restoration Program, July.

Satman, H., 2008. Test for Normality (Jarque-Bera Test), Statistics with Java. http://www.mhsatman.com/applets.htm

Shacklette, H.T., and J.G. Boerngen 1984. Element Concentrations in Soils and Other Surficial Materials, Conterminous United States, U.S. Geological Survey Professional Paper 1270.

Shapiro, S.S. and R.S. Francia, 1972. An approximate analysis of variance test for normality. Journal of the American Statistical Association. 67:215-216.

USEPA 1989. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities, Interim Final Guidance, Office of Solid Waste and Emergency Response, Publication number PB89-151047, February.

USEPA 1992. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities. Addendum to Interim Final Guidance. July.

USEPA 1997. Technology Support Center Issue - The Lognormal Distribution in Environmental Applications. EPA/600/R-097/006. December.

USEPA 2000. Guidance for Data Quality Assessment, Practical Methods for Data Analysis. EPA QA/G-9.

APPENDIX A

THE EFFECT OF LARGE DATA SETS ON THE SHAPIRO-FRANCIA TEST

EXPLANATION OF GRAPHS IN APPENDIX A

Histogram

- Means and standard deviations of concentration data are shown on the X-axis
- The number of samples is shown on the Y-axis
- Each vertical bar represents a grouping of samples centered on the concentration at the bottom of the vertical bar

Density Estimate

- Means and standard deviations of concentration data are shown on the X-axis
- The probability of a sample having a particular concentration is shown on the Y-axis
- The probability density of the data was estimated by a smoothing operation that uses kernel estimators and are based on a Gaussian kernel
- This plot is essentially a "smoothed" histogram

Box Plot

- Means and standard deviations of concentration data are shown on the Y-axis
- The white line within the box is the median
- The ends of the box mark the interquartile range (IQR = difference between the 25th and 75th percentiles)
- Whiskers ([-- and --]) mark 1.5 IQR's on either side of the median (total width of 3 IQR)
- Horizontal lines outside the whiskers mark all data points that lie greater than 1.5 IQRs beyond the median.

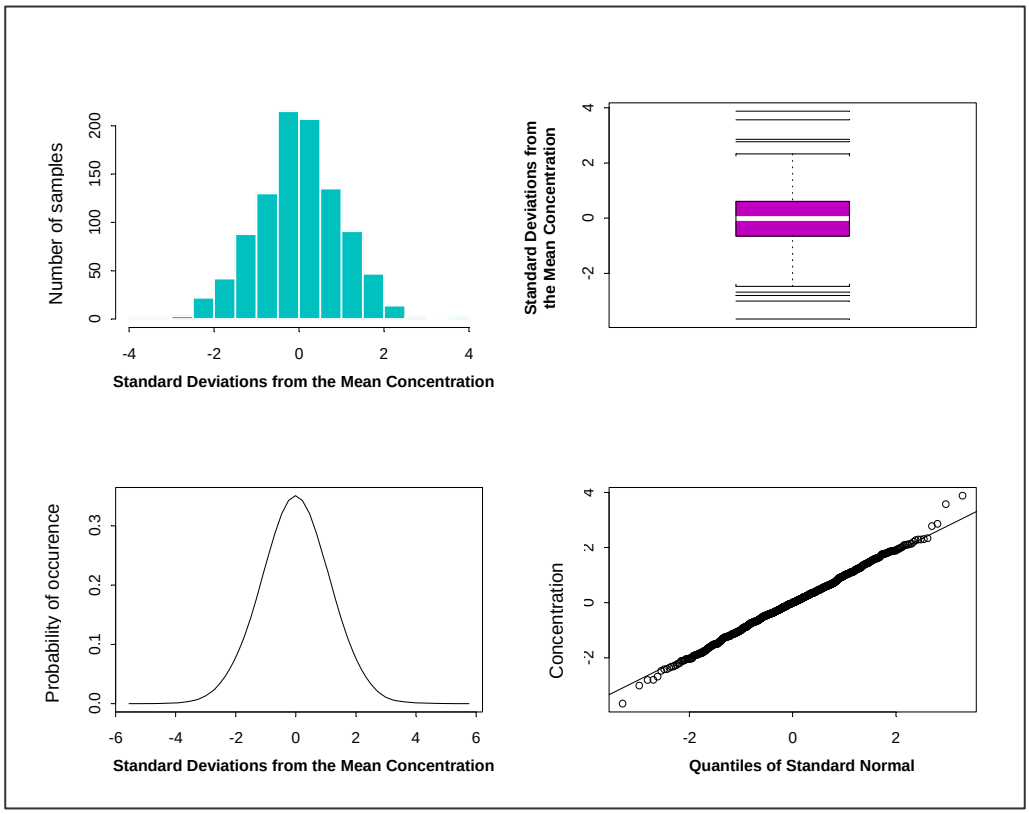
Q-Q Plot

- Quantiles (points dividing dataset into equal numbers of observations) of a standard normal distribution are shown on the X-axis
- Means and standard deviations of concentration data are shown on the Y-axis
- Ranked observed values of the sample distribution are plotted against quantiles

To demonstrate the effect of large data sets on the incorrect rejection of the null hypothesis (i.e., Type II error) by the Shapiro-Francia Test, a synthetic data set was generated. The data set consists of 1,000 random deviates from a standard normal distribution. The lower 1% critical value for the Shapiro-Francia test with a sample size of 1,000 is 0.9954 (determined empirically using the method described by Shapiro and Francia 1972). The Shapiro-Francia test gives a test statistic of W = 0.997 and a rejection of the null hypothesis. However, a visual inspection of the distribution of the synthetic data given on Figure A-1 indicates that the normality assumption may be justified because the histogram and box plot are symmetric and the quantile-quantile plot is linear. This graphical approach is useful when detection limits are low, so that there is little censoring of the data set. When a significant amount of censoring is present in a data set (i.e., due to a large proportion of non-detects), distribution shape is difficult to determine since the sample size is much smaller if an analysis is performed without non-detects. In addition, if the non-detects are retained in the analysis, only their maximal concentrations are known and upper percentiles may be overestimated, especially if most detected values are greater than the non-detects.

Figure A-1

Synthetic Data Used to Demonstrate Type II Error in the Shapiro-Francia Test on Large Data Sets



APPENDIX B

GRAPHICAL REPRESENTATIONS OF DATA BY METAL

EXPLANATION OF GRAPHS IN APPENDIX B

Histogram

- Concentrations in mg/kg are shown on the X-axis
- The number of samples is shown on the Y-axis
- Each vertical bar represents a grouping of samples centered on the concentration at the bottom of the vertical bar

Box Plot

- Concentrations in mg/kg are shown on the Y-axis
- The white line within the box is the median
- The ends of the box mark the interquartile range (IQR = difference between the 25th and 75th percentiles)
- Whiskers ([-- and --]) mark 1.5 IQR's on either side of the median (total width of 3 IQR)
- Horizontal lines outside the whiskers mark all data points that lie greater than 1.5 IQRs beyond the median.

Density Estimate

- Concentrations in mg/kg are shown on the X-axis
- The probability of a sample having a particular concentration is shown on the Y-axis
- The probability density of the data was estimated by a smoothing operation that uses kernel estimators and are based on a Gaussian kernel
- This plot is essentially a "smoothed" histogram

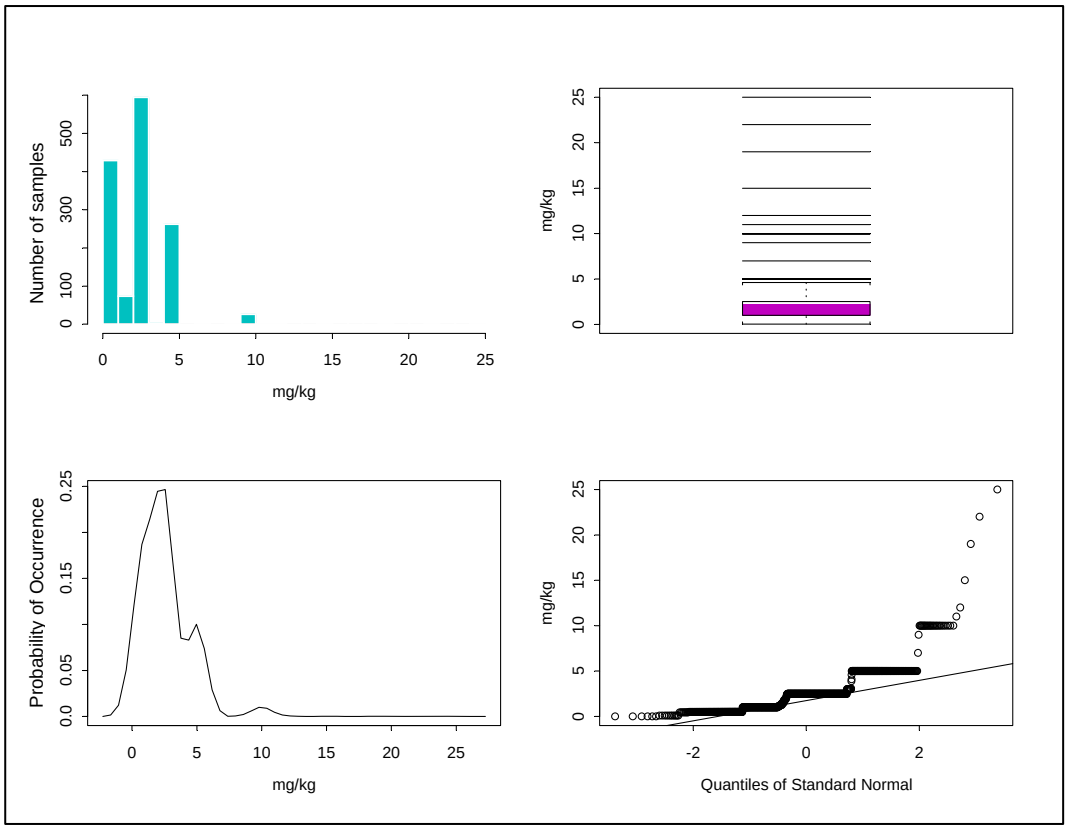
Q-Q Plot

- Quantiles (points dividing dataset into equal numbers of observations) of a standard normal distribution are shown on the X-axis
- Concentrations in mg/kg are shown on the Y-axis
- Ranked observed values of the sample distribution are plotted against quantiles of a standard normal distribution

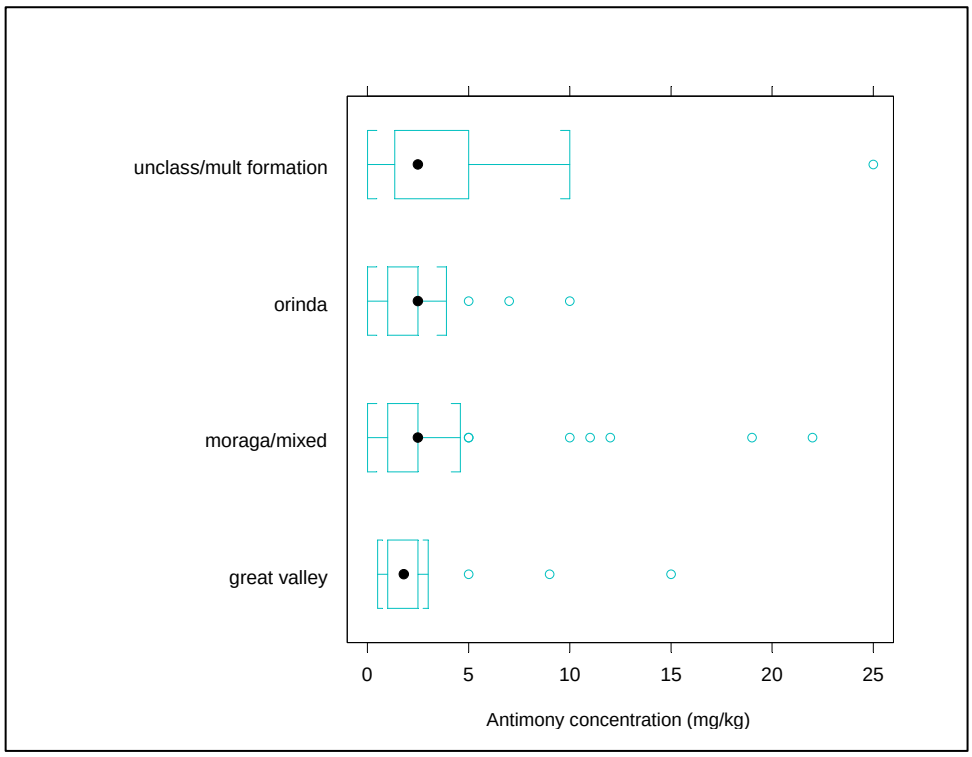
Box Plot

- Concentrations in mg/kg are shown on the X-axis
- Geologic classifications are shown on the Y-axis
- Solid dots are the median values
- The ends of the box mark the interquartile range (IQR = difference between the 25th and 75th percentiles_
- Whiskers ([-- and --]) mark 1.5 IQR's on either side of the median (total width of 3 IQR)
- Open circles outside the whiskers mark all data points that lie greater than 1.5 IQRs beyond the median.

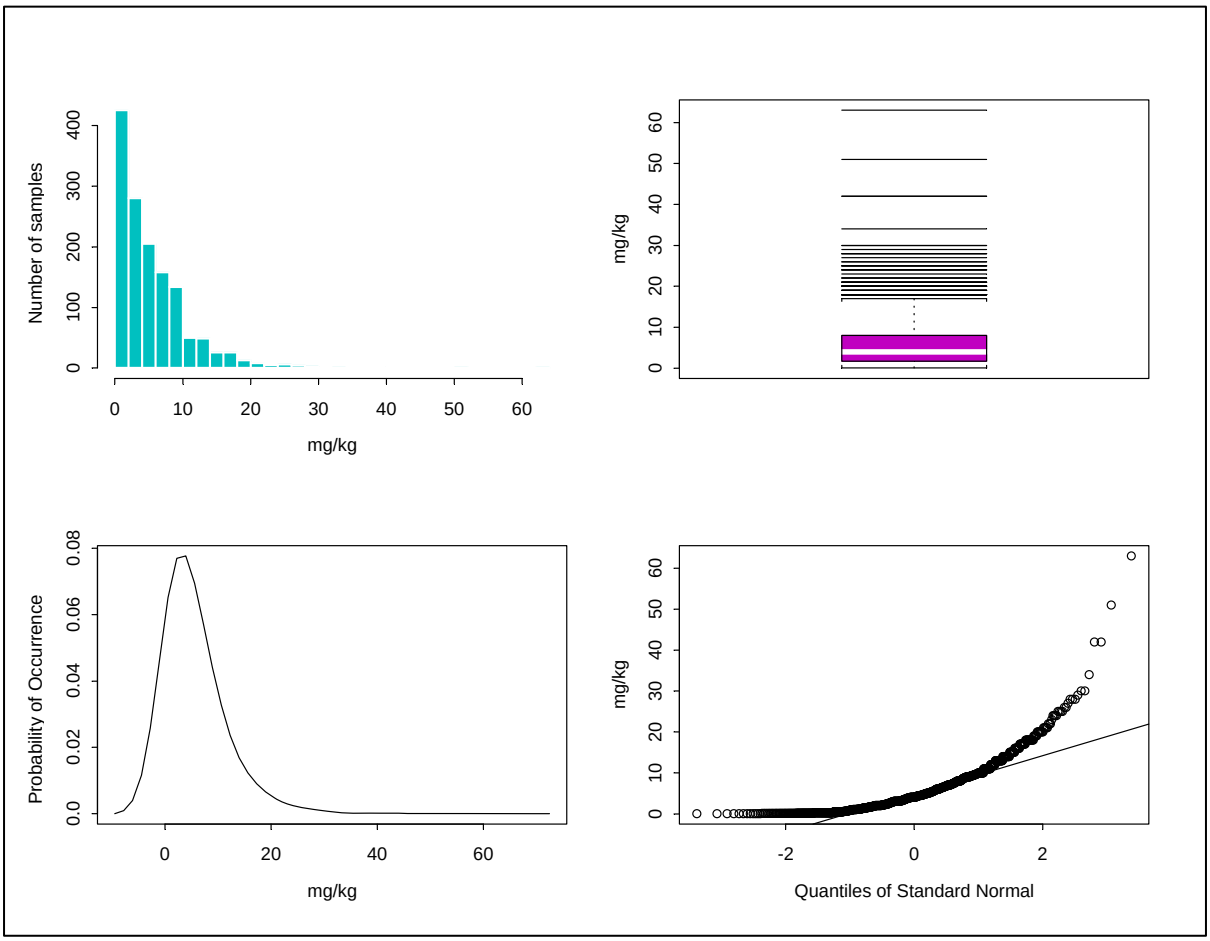
Four-Plot Presentation of Antimony Data after the Removal of Possible Outliers

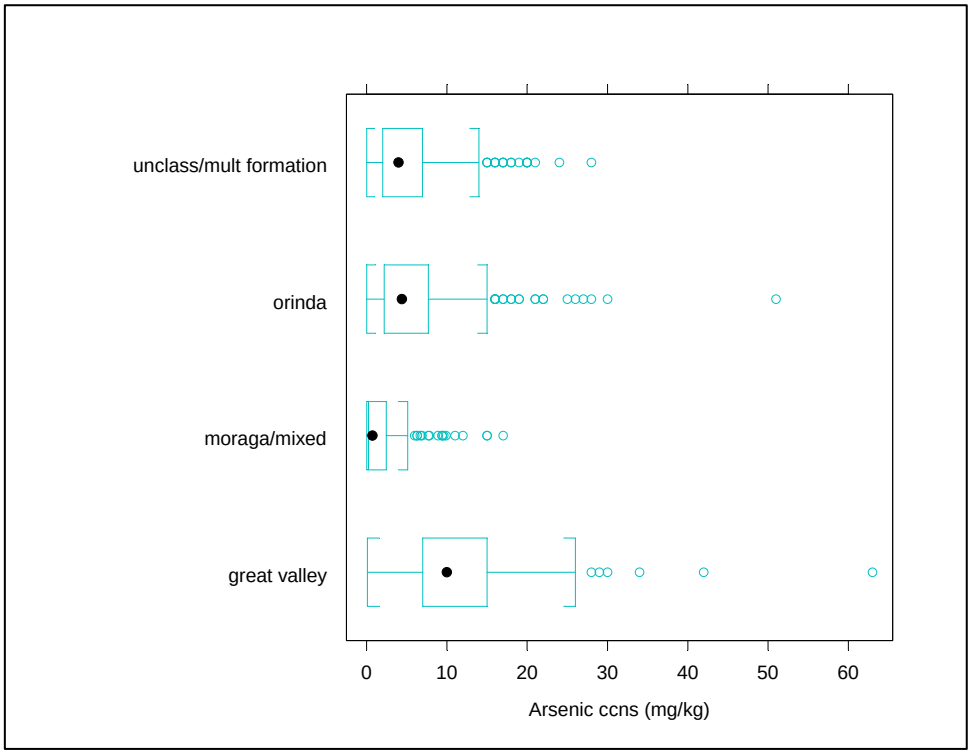


Box Plot of Antimony Soil Concentrations by Geologic Classification

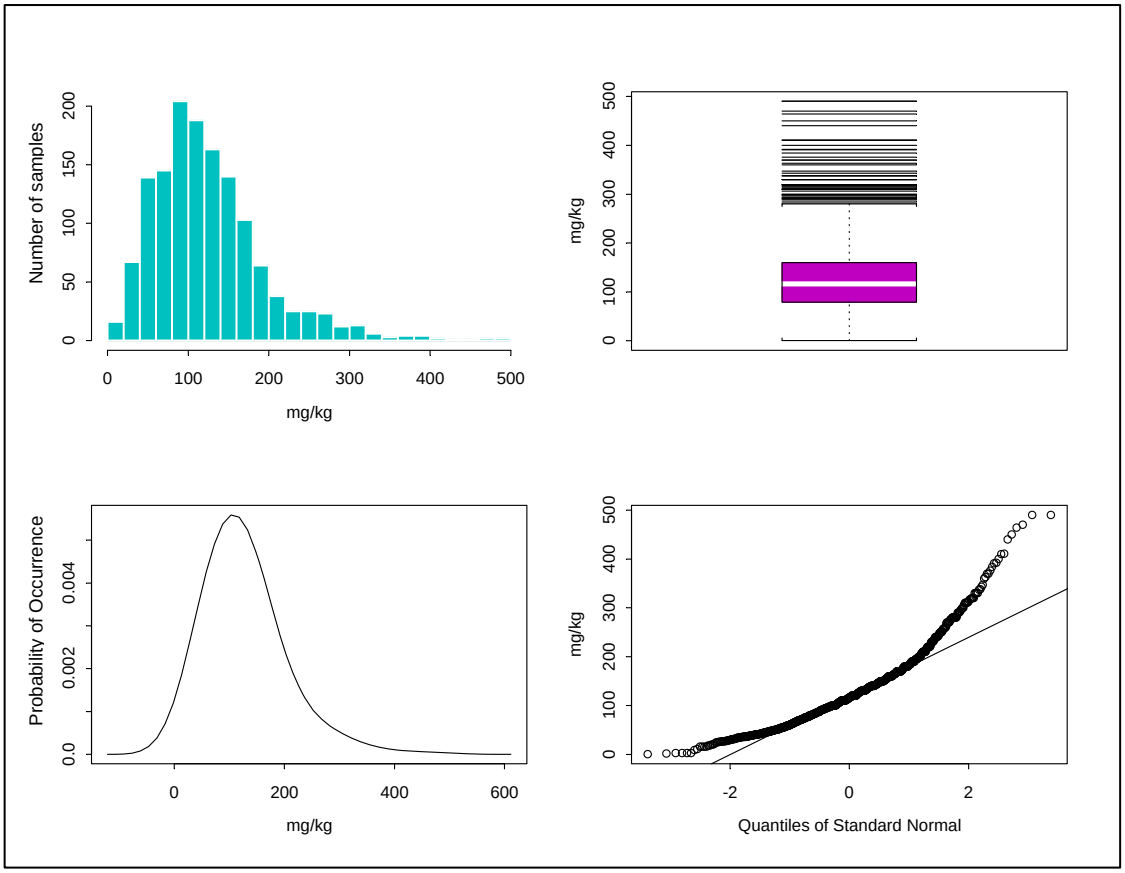


Four-Plot Presentation of Arsenic Data
after the Removal of Possible Outliers

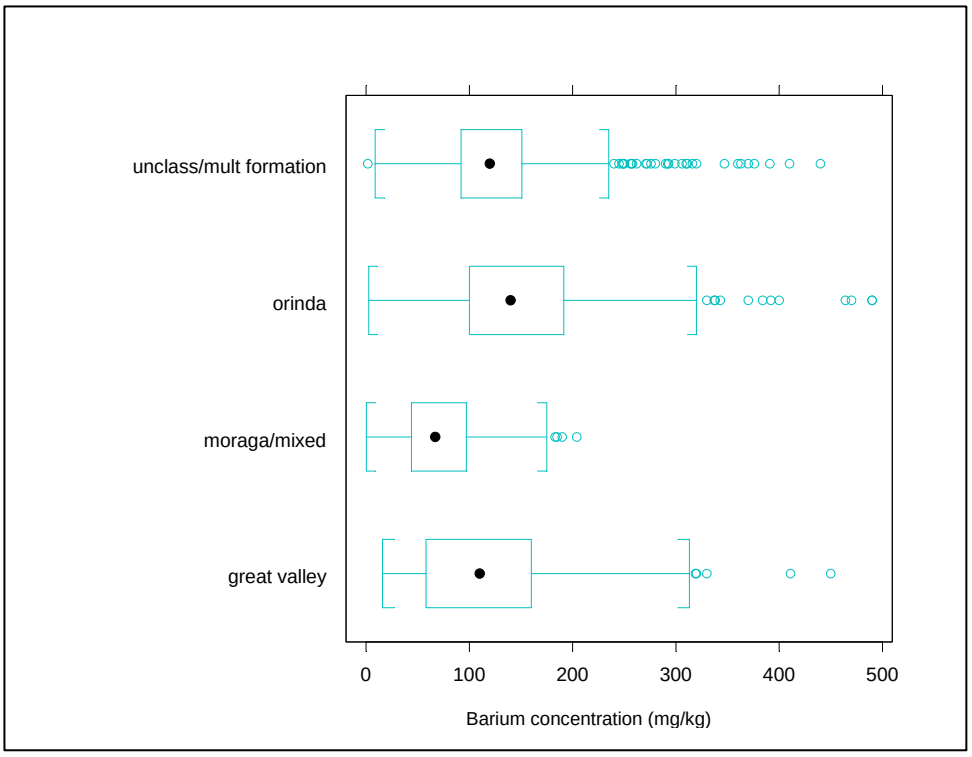




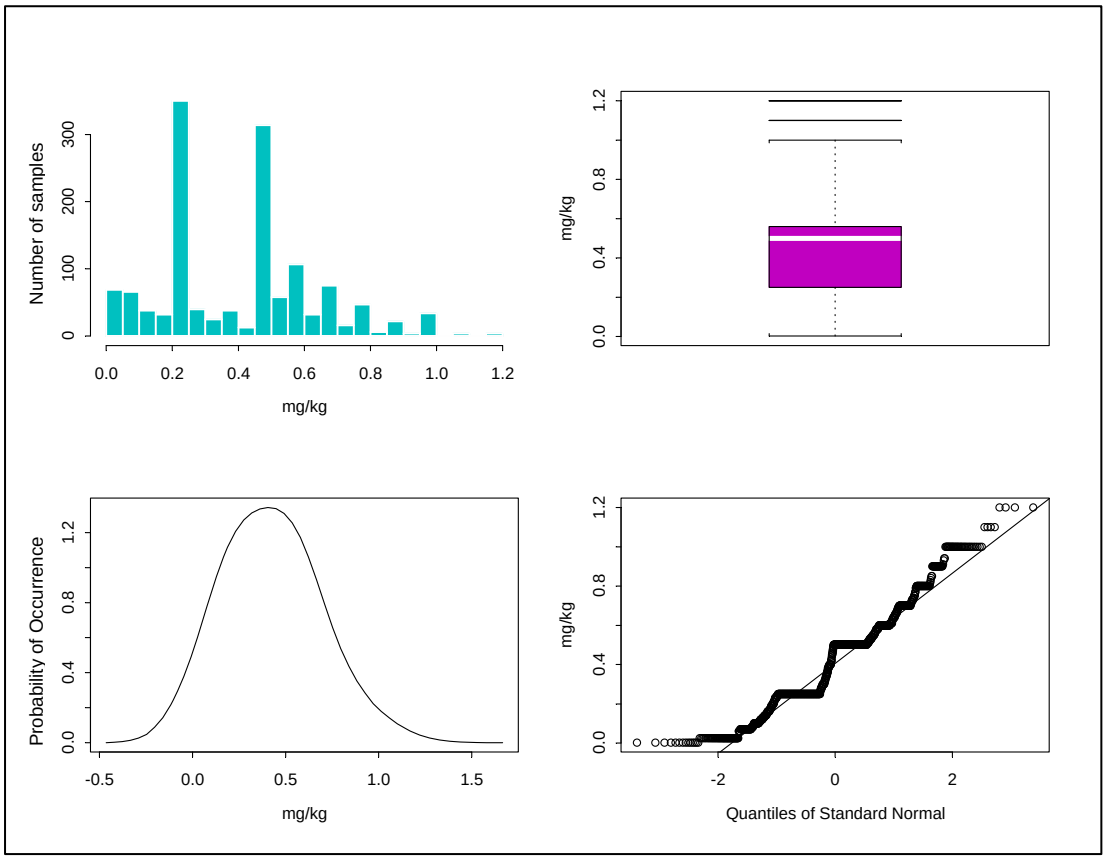
Four-Plot Presentation of Barium Data after the Removal of Possible Outliers



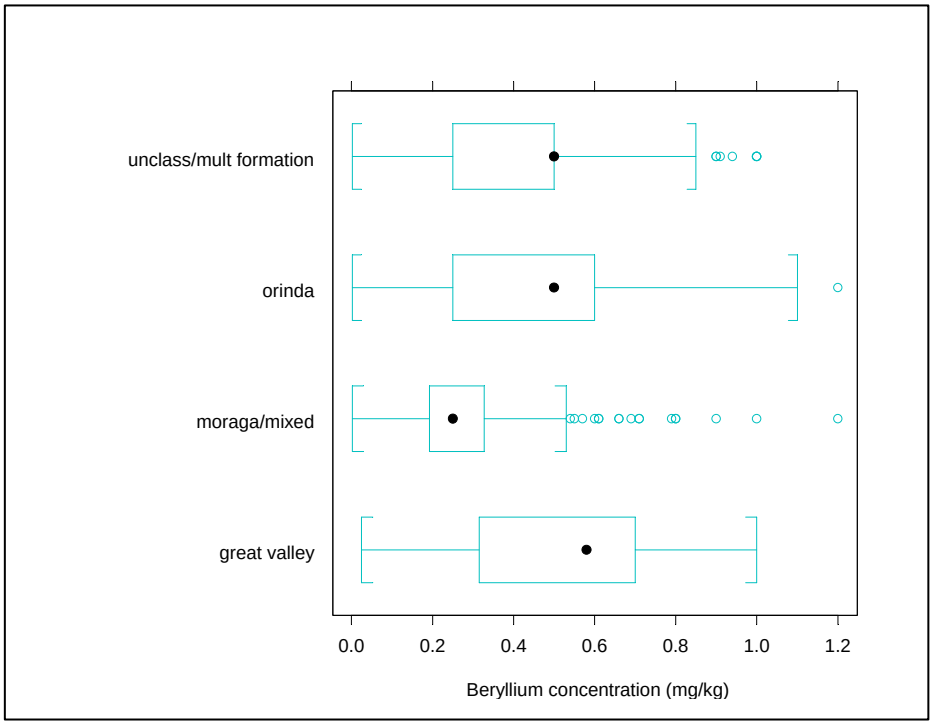
Box Plot of Barium Soil Concentrations by Geologic Classification



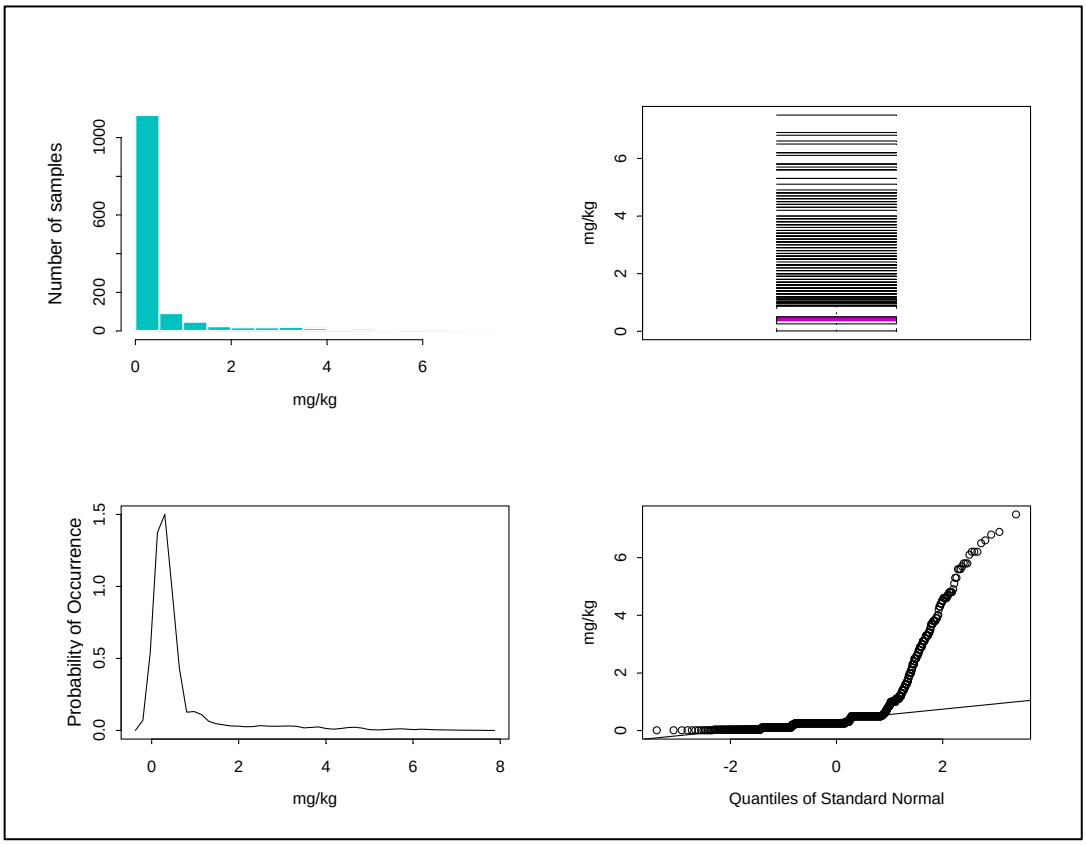
Four-Plot Presentation of Beryllium Data after the Removal of Possible Outliers



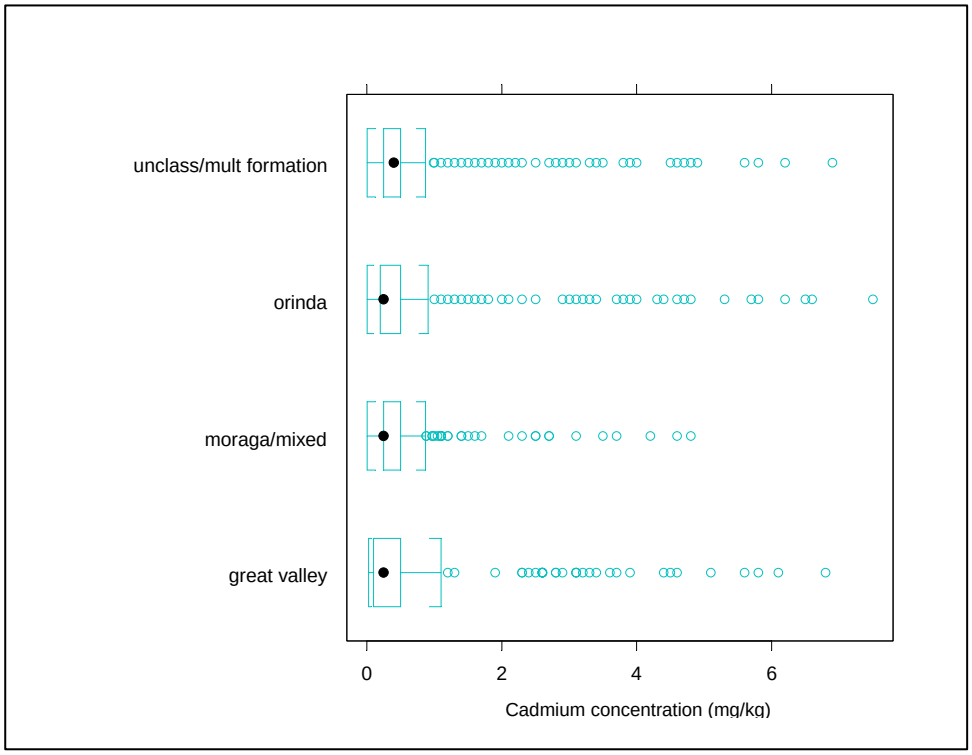
Box Plot of Beryllium Soil Concentrations by Geologic Classification



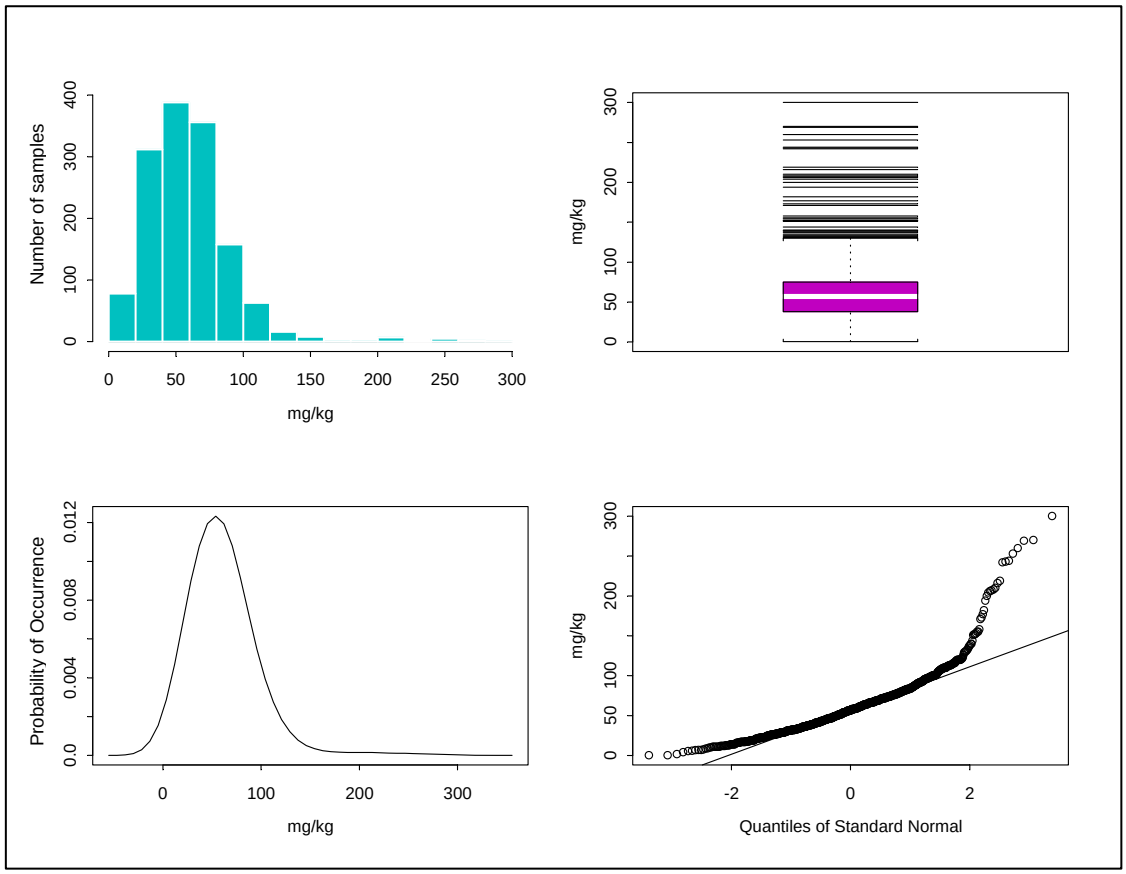
Four-Plot Presentation of Cadmium Data after the Removal of Possible Outliers



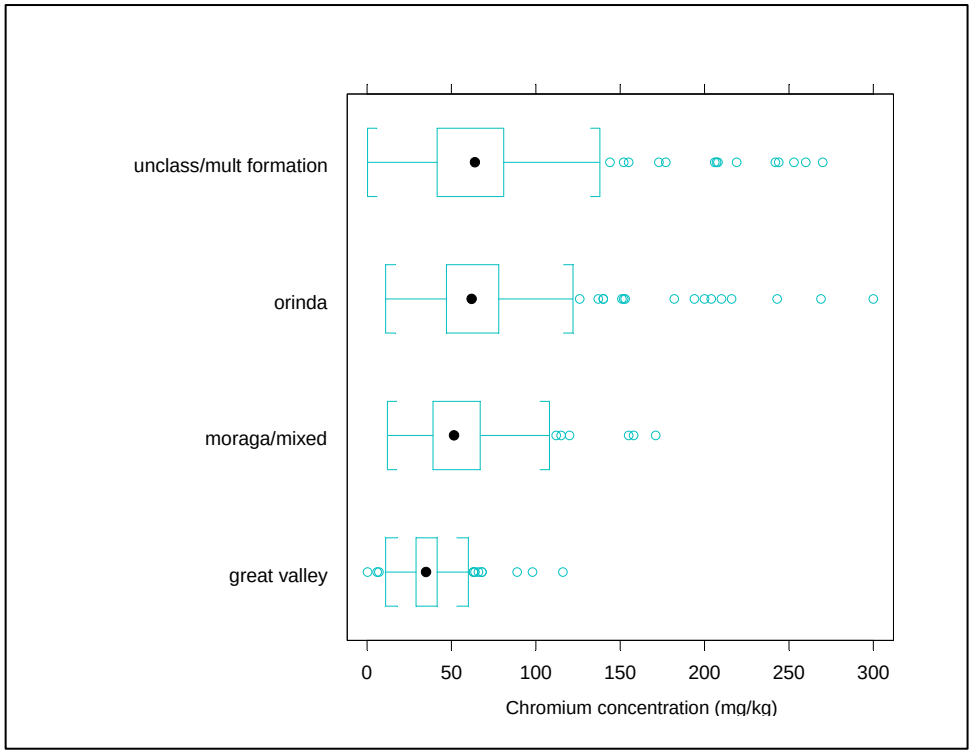
Box Plot of Cadmium Soil Concentrations by Geologic Classification



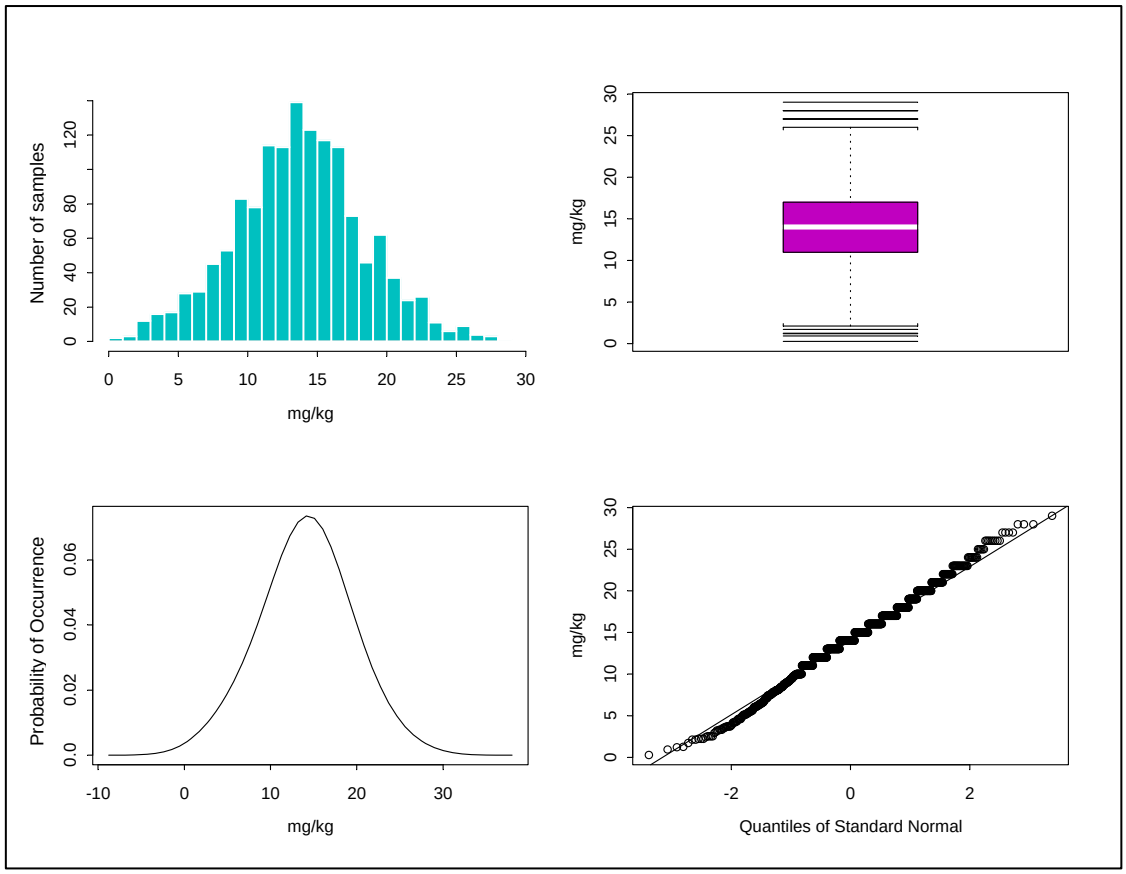
Four-Plot Presentation of Chromium Data after the Removal of Possible Outliers



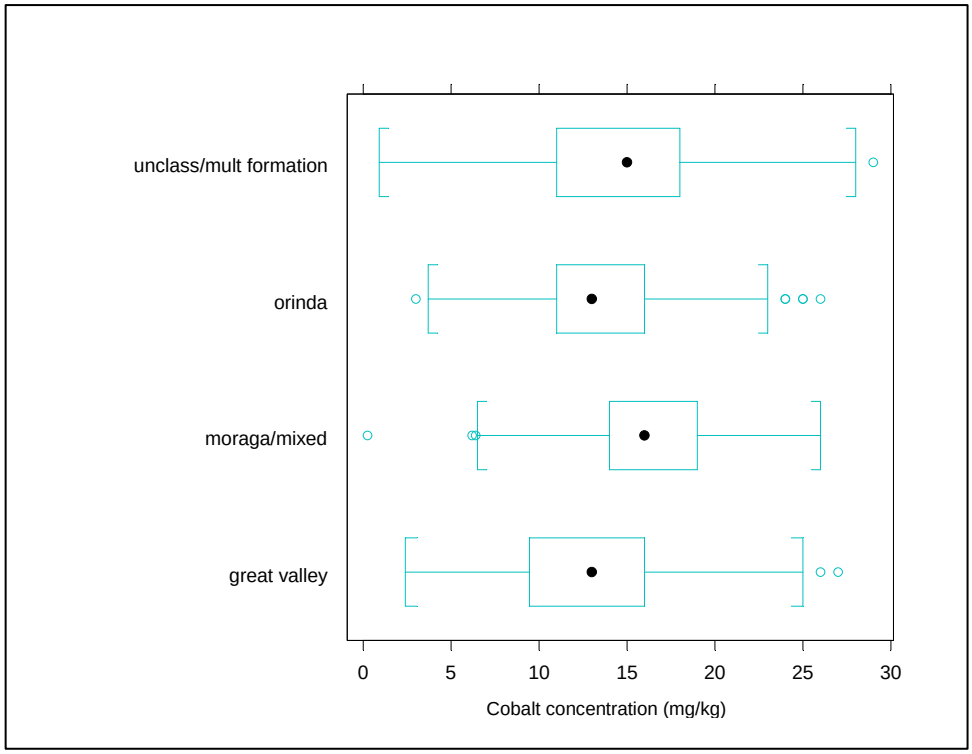
Box Plot of Chromium Soil Concentrations by Geologic Classification



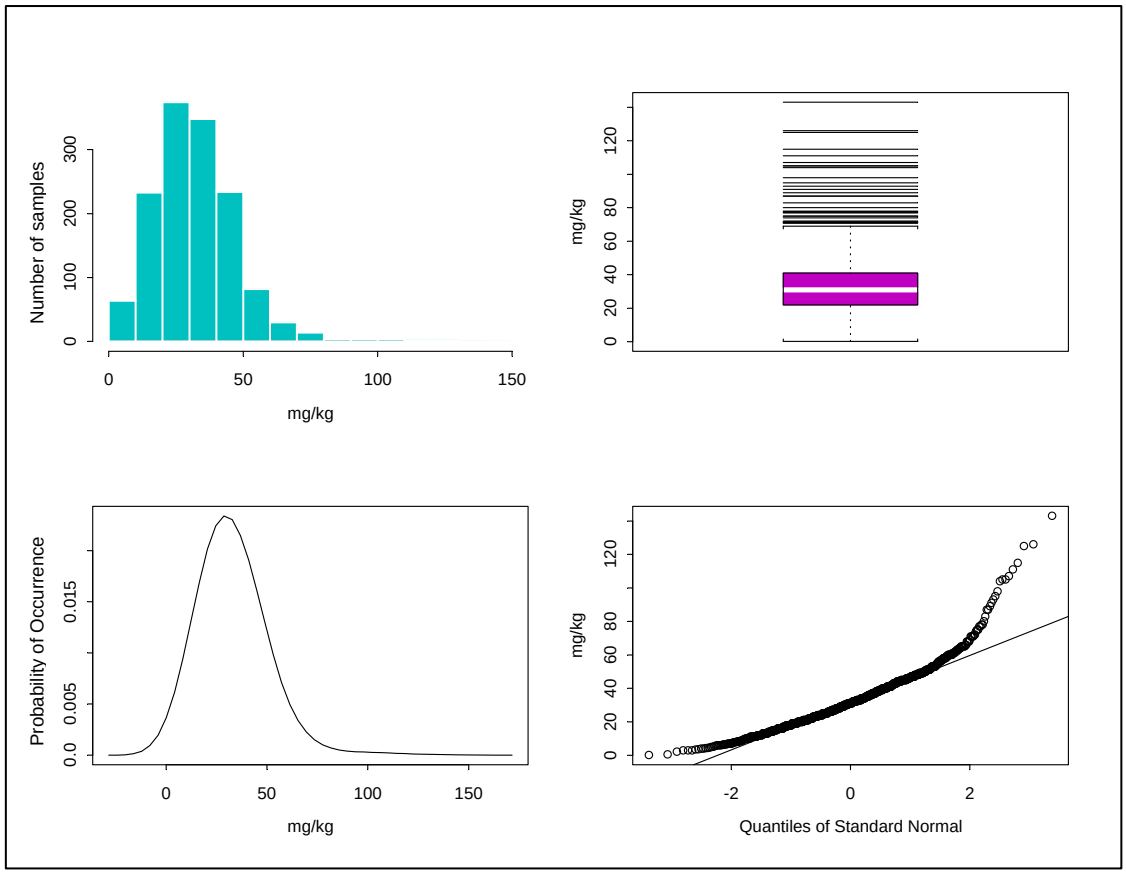
Four-Plot Presentation of Cobalt Data after the Removal of Possible Outliers



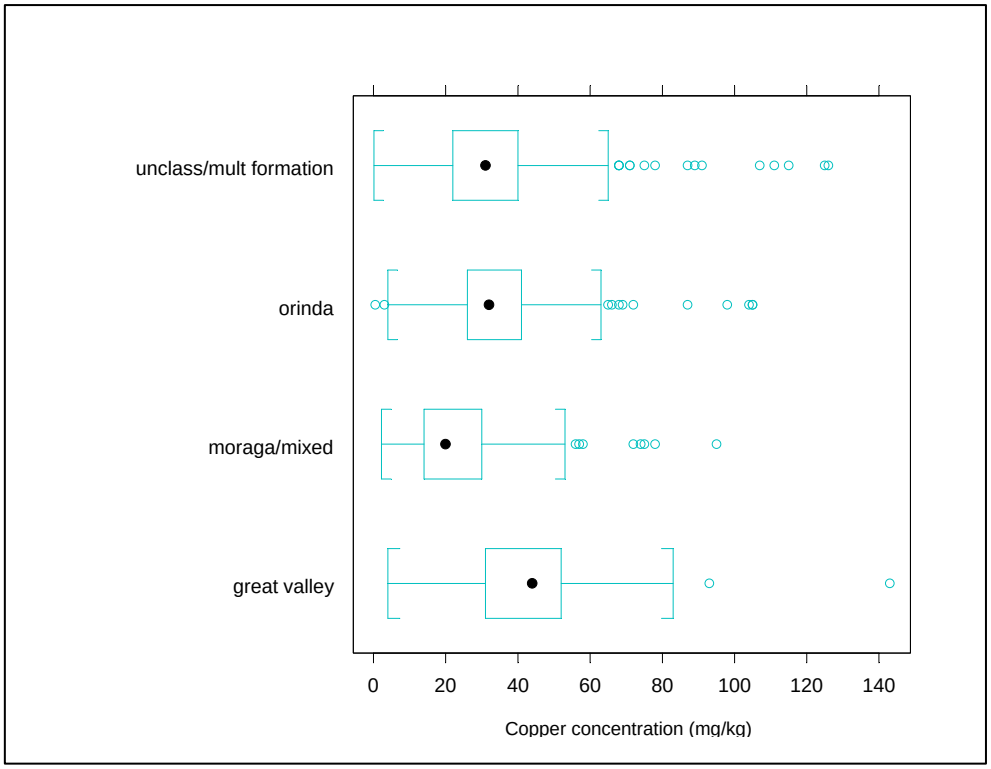
Box Plot of Cobalt Soil Concentrations by Geologic Classification



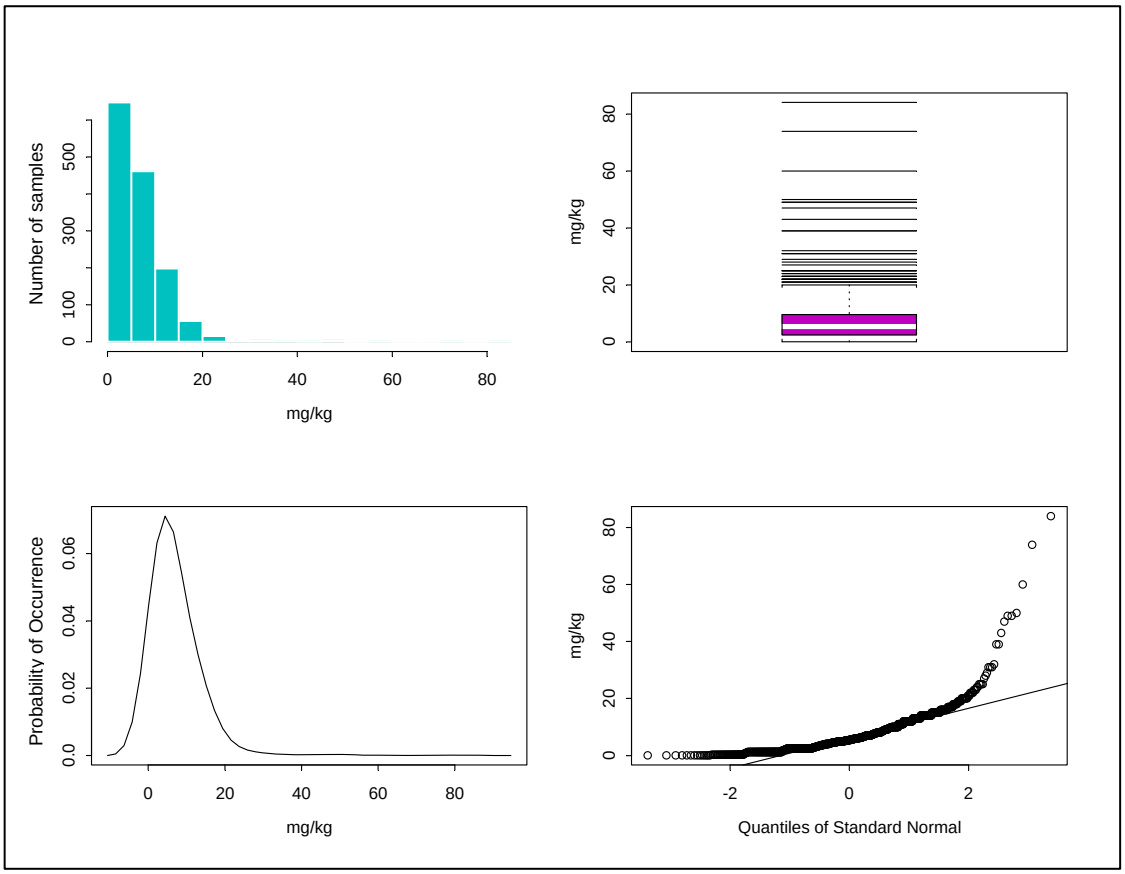
Four-Plot Presentation of Copper Data after the Removal of Possible Outliers



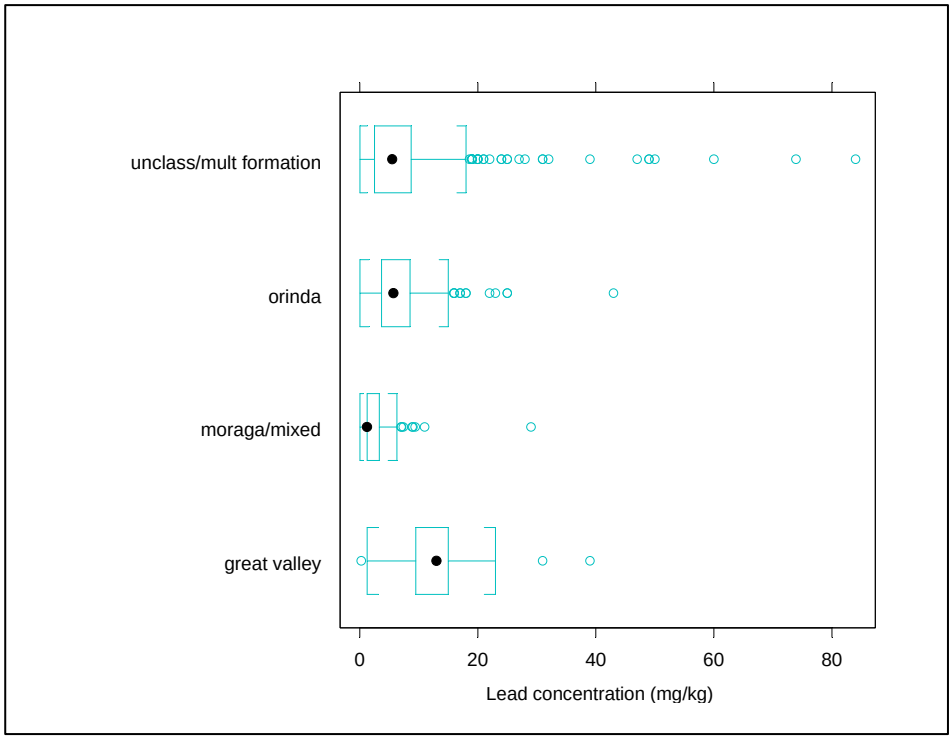
Box Plot of Copper Soil Concentrations by Geologic Classification



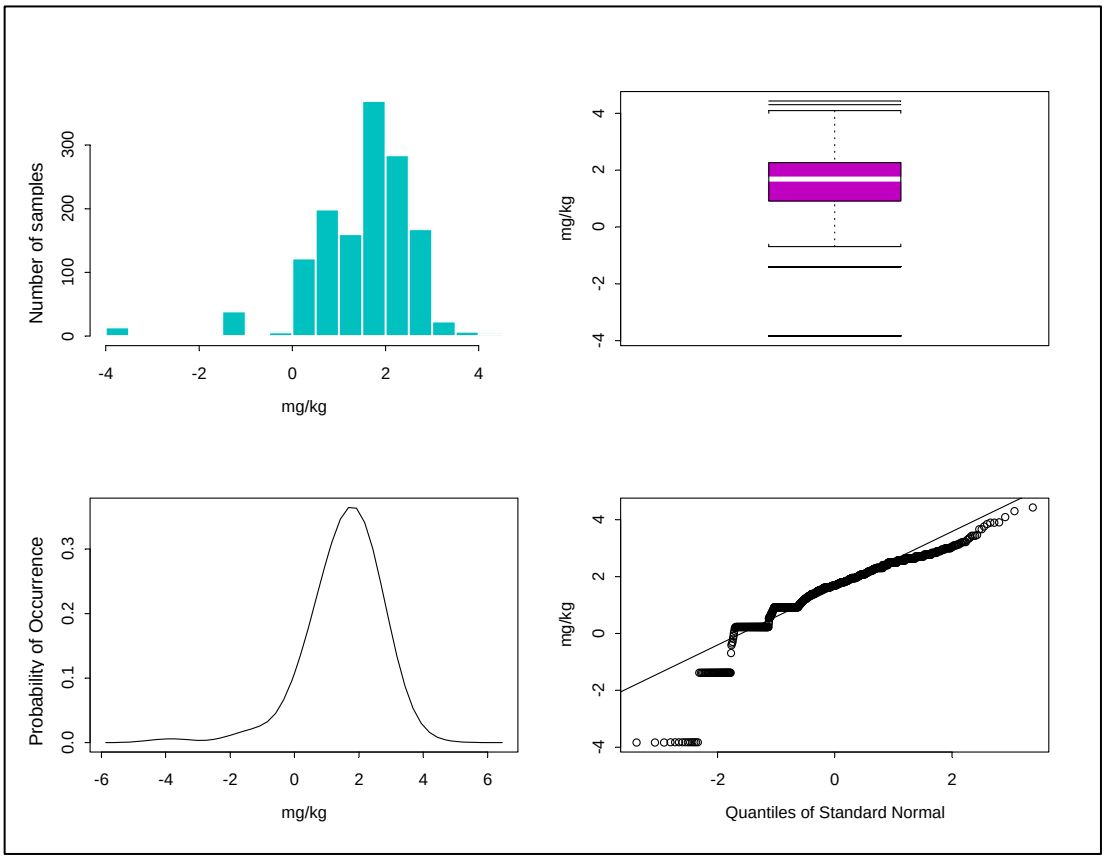
Four-Plot Presentation of Lead Data after the Removal of Possible Outliers



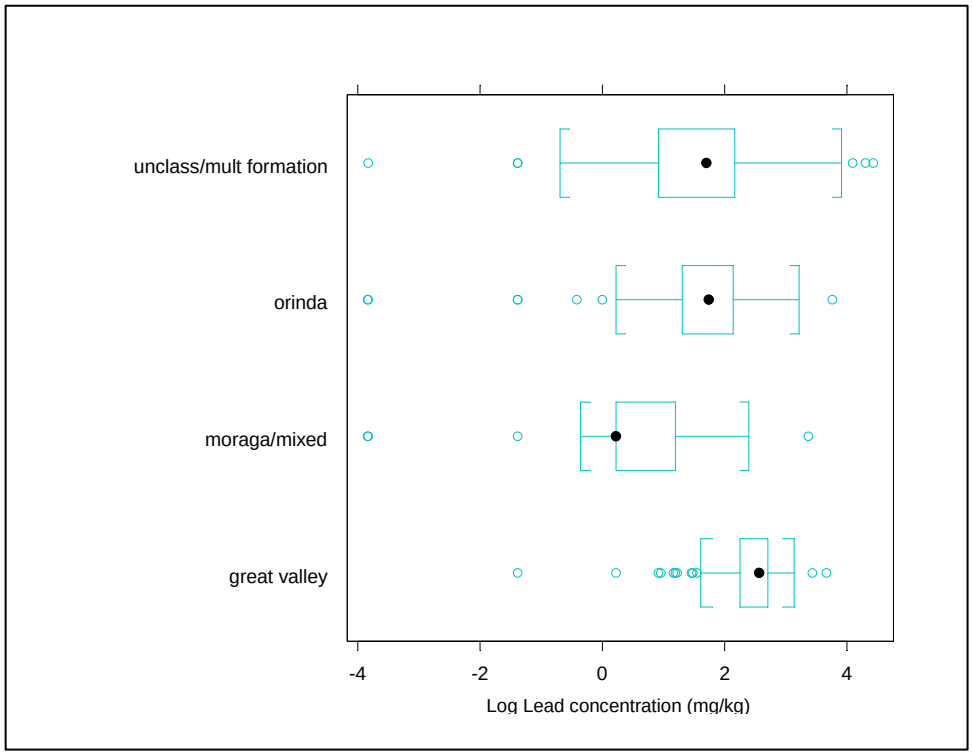
Box Plot of Lead Soil Concentrations by Geologic Classification



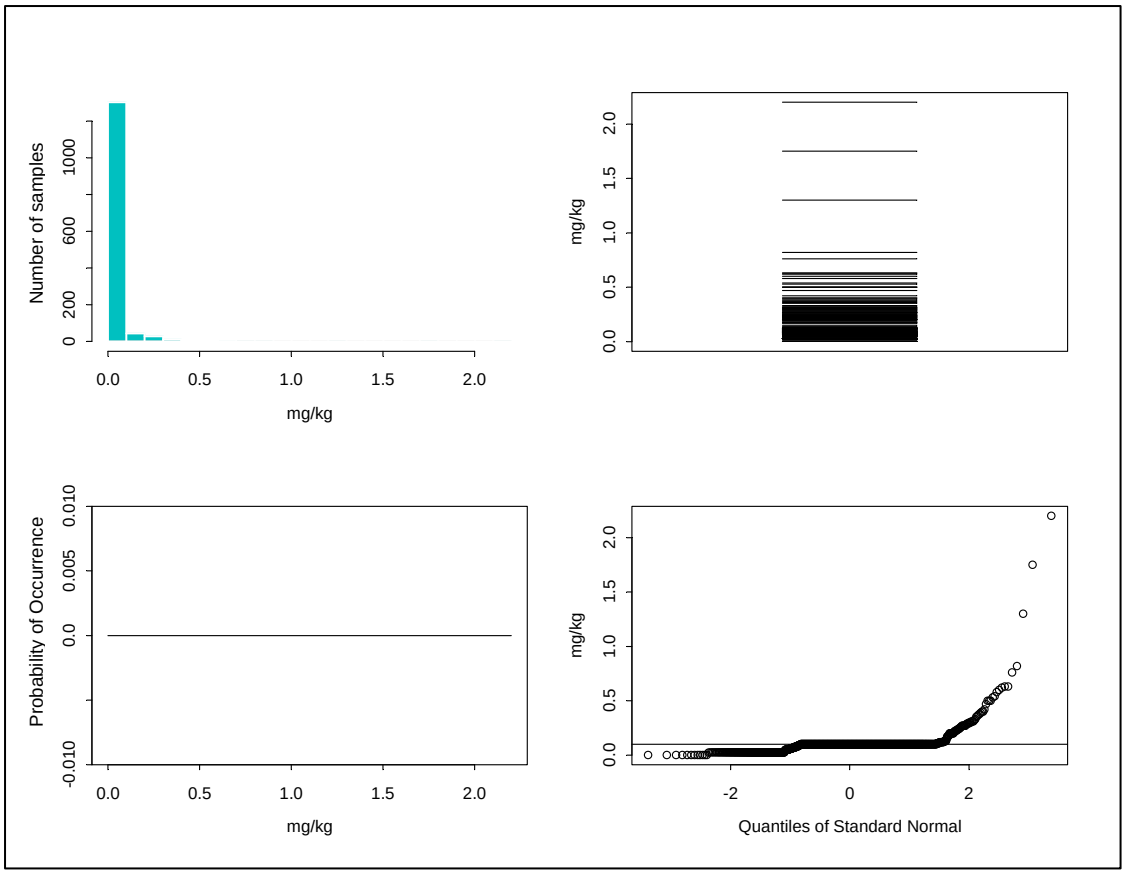
Four-Plot Presentation of Lead Data after the Removal of Possible Outliers Log-Transformed Data



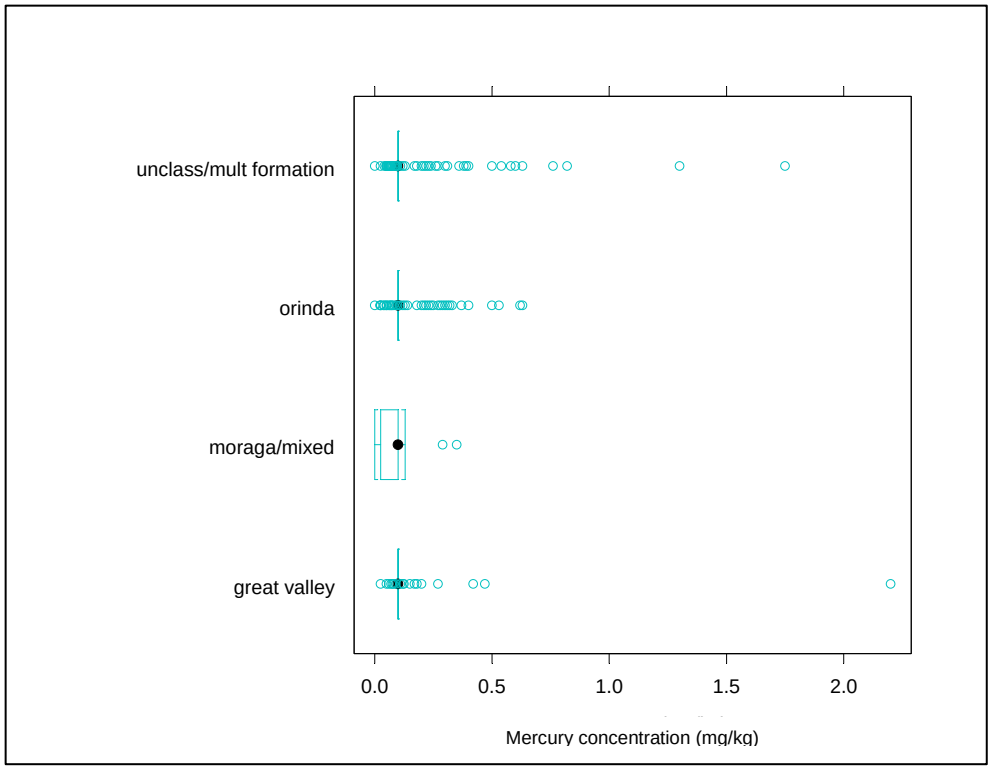
Box Plot of Lead Soil Concentrations by Geologic Classification Log-Transformed Data



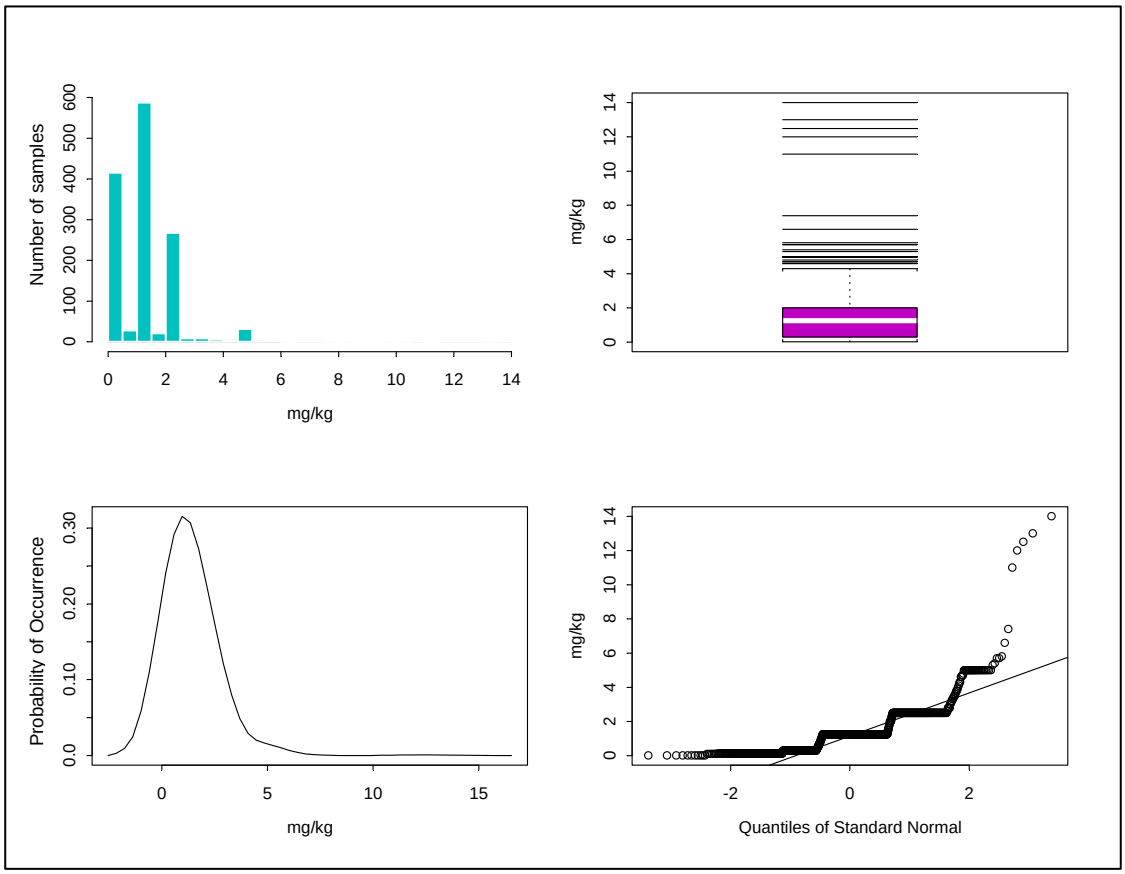
Four-Plot Presentation of Mercury Data after the Removal of Possible Outliers



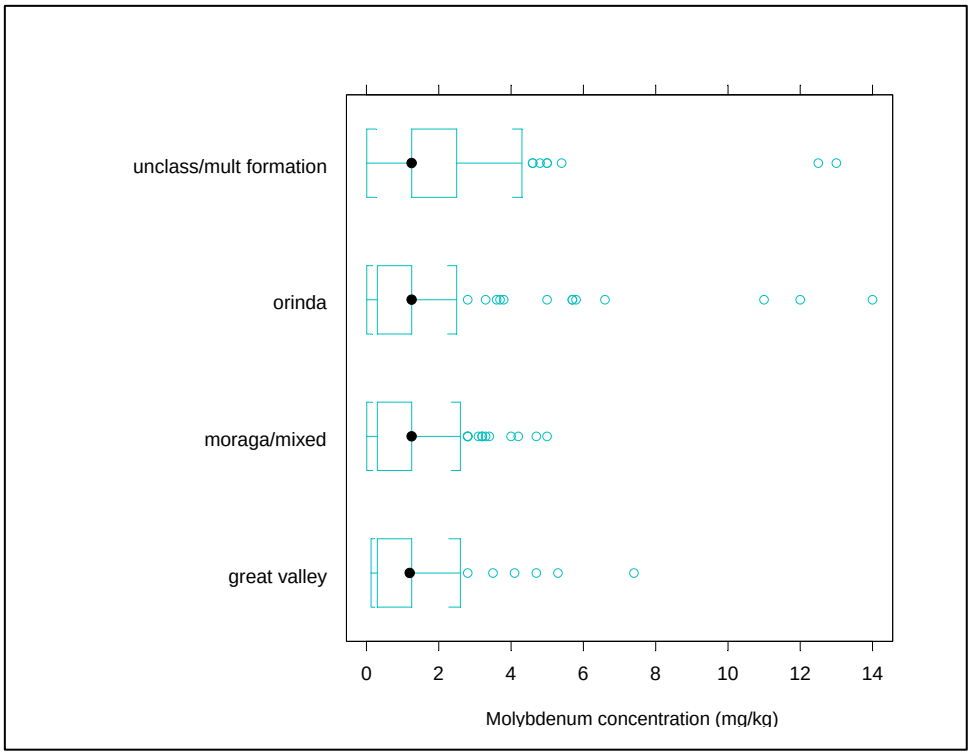
Box Plot of Mercury Soil Concentrations by Geologic Classification



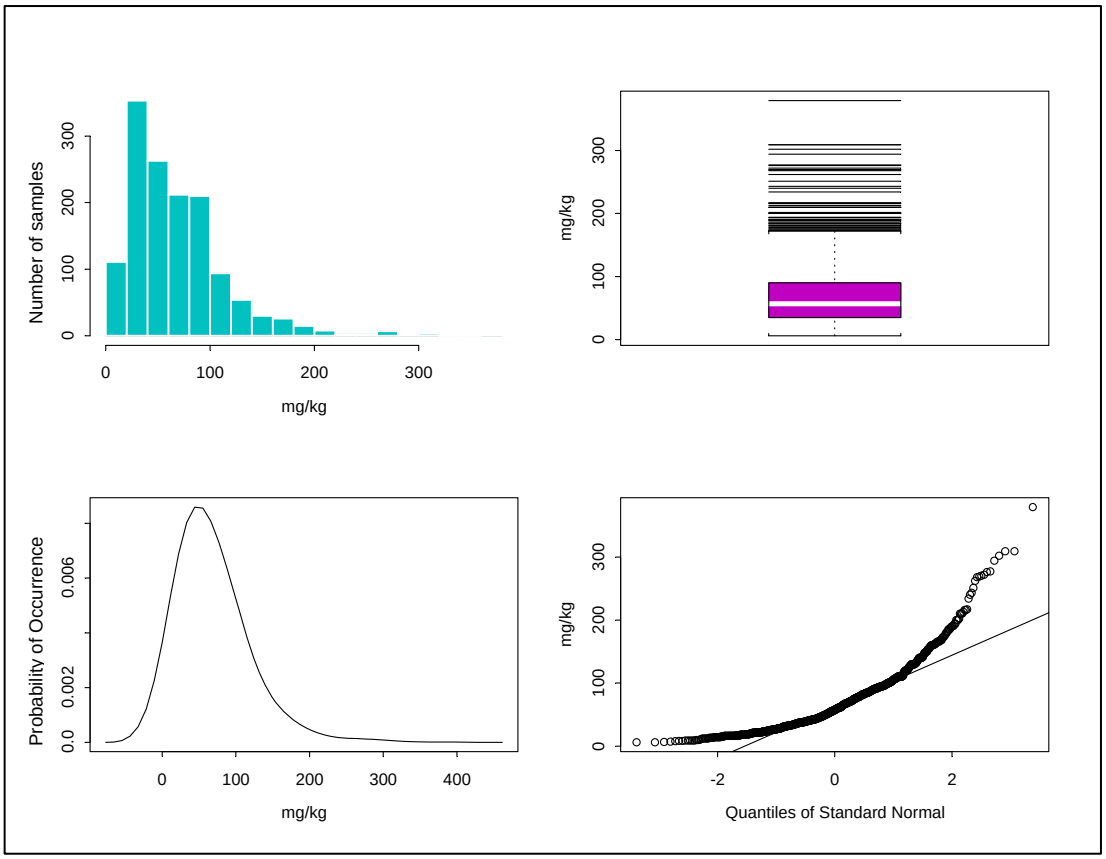
Four-Plot Presentation of Molybdenum Data after the Removal of Possible Outliers



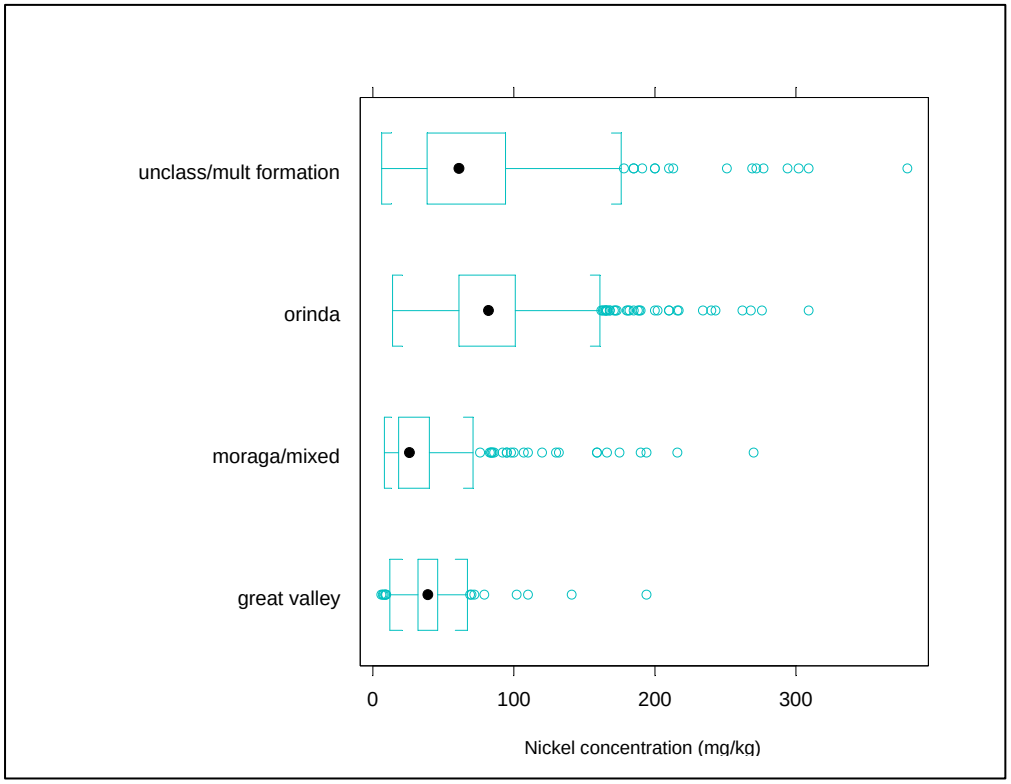
Box Plot of Molybdenum Soil Concentrations by Geologic Classification



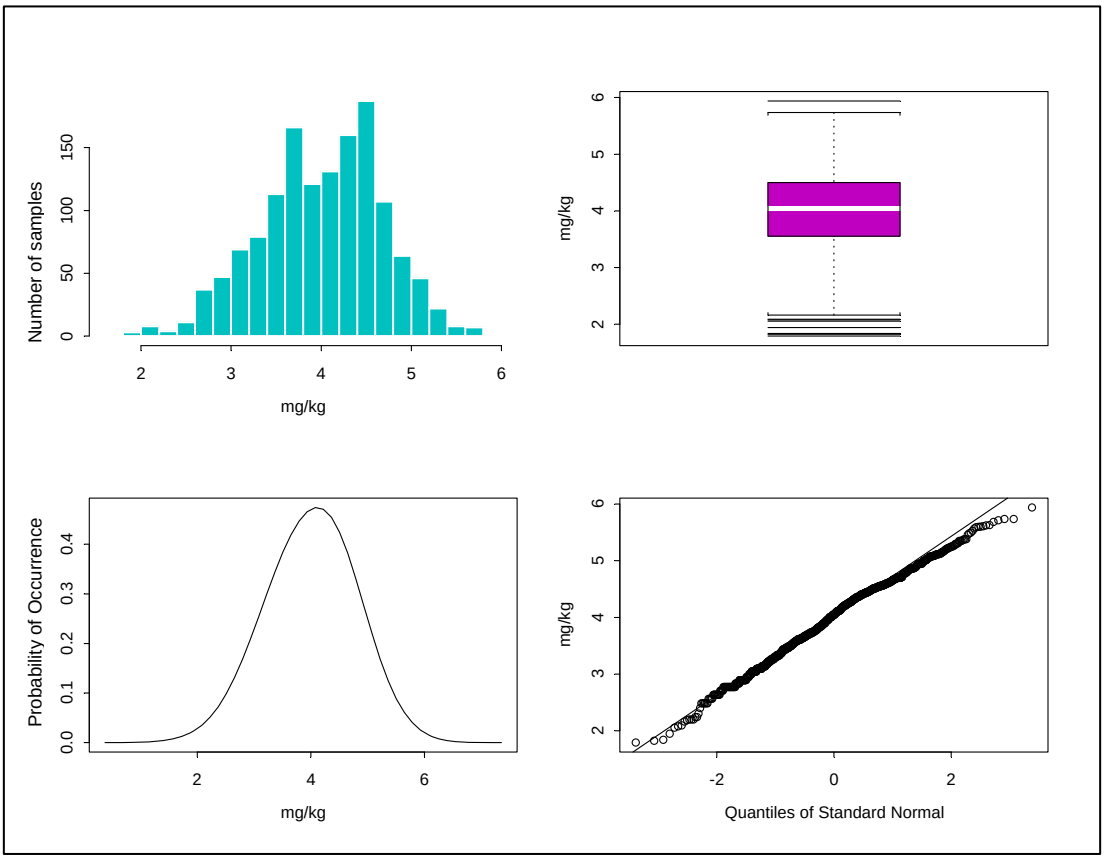
Four-Plot Presentation of Nickel Data after the Removal of Possible Outliers



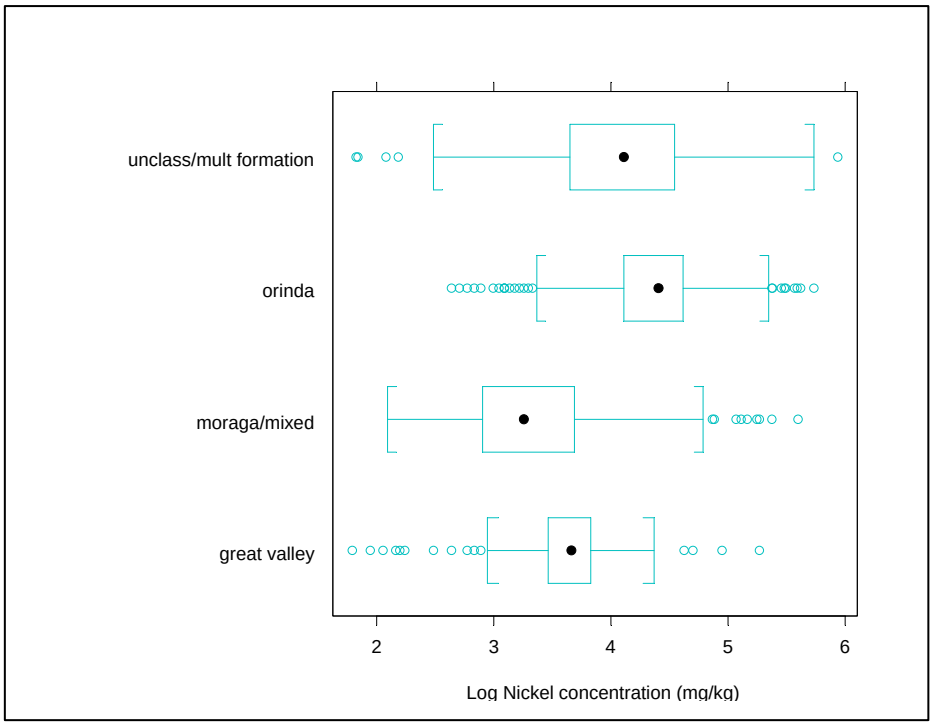
Box Plot of Nickel Soil Concentrations by Geologic Classification



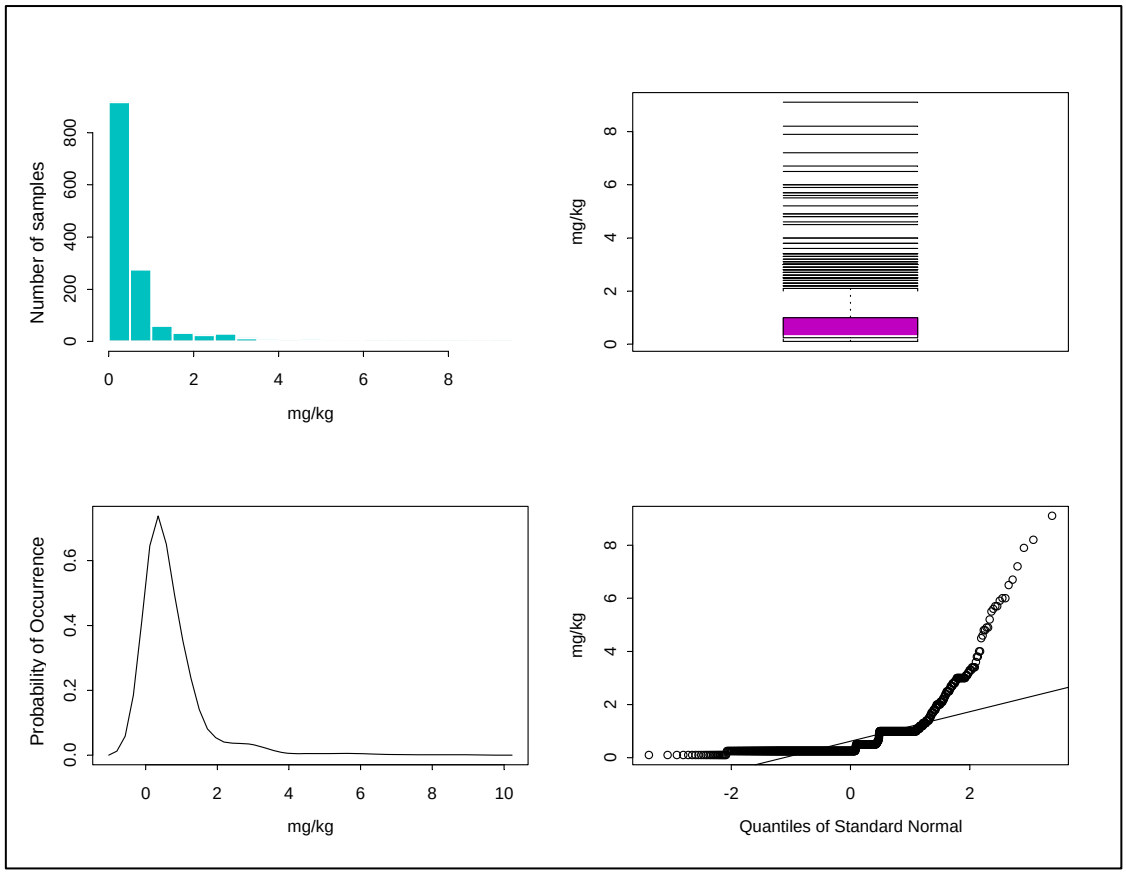
Four-Plot Presentation of Nickel Data after the Removal of Possible Outliers Log-transformed Data



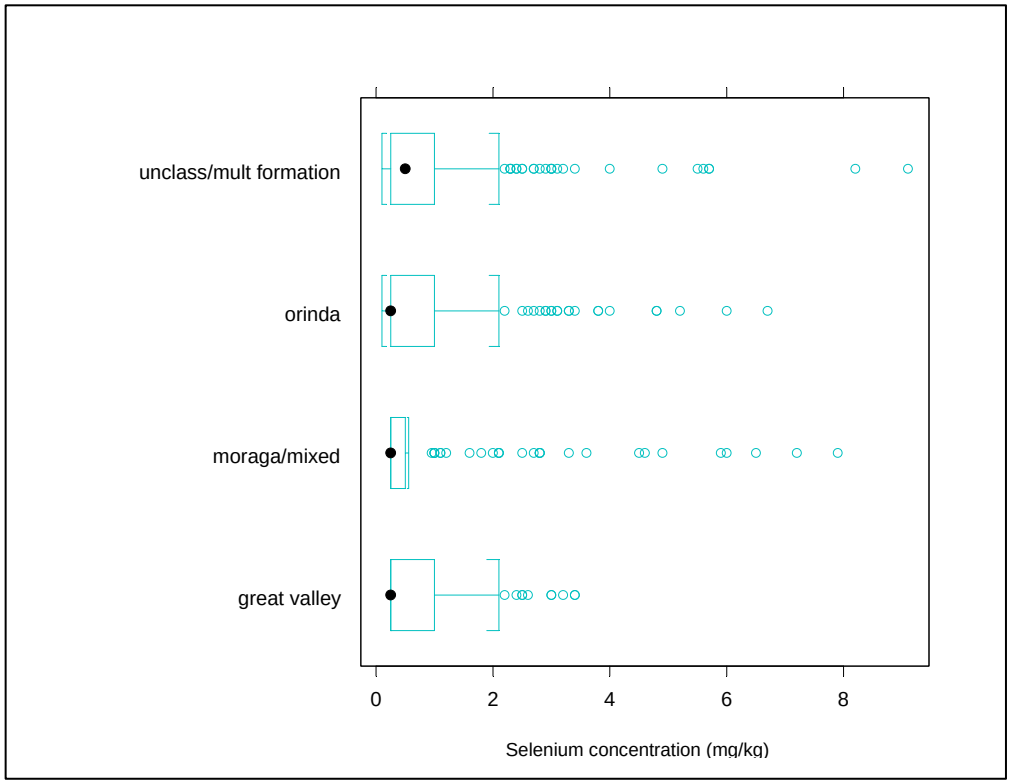
Box Plot of Nickel Soil Concentrations by Geologic Classification Log-Transformed Data



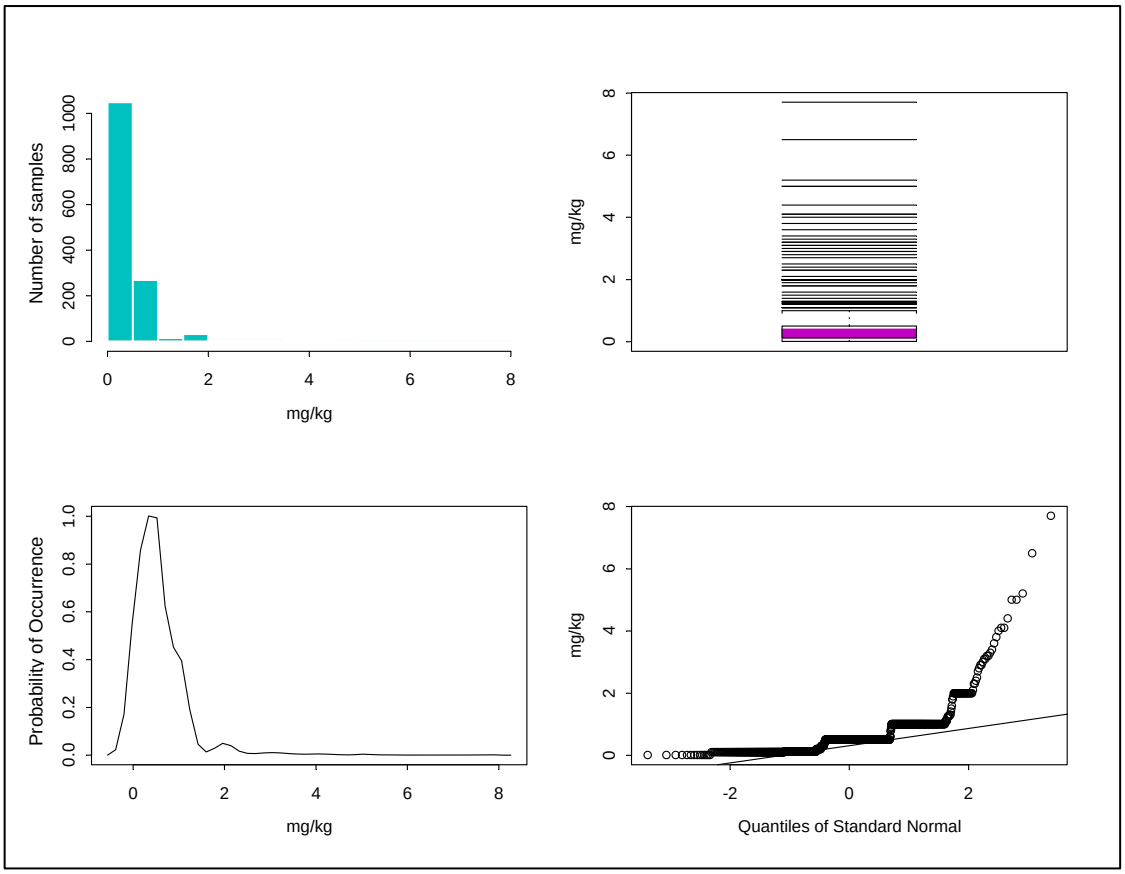
Four-Plot Presentation of Selenium Data after the Removal of Possible Outliers



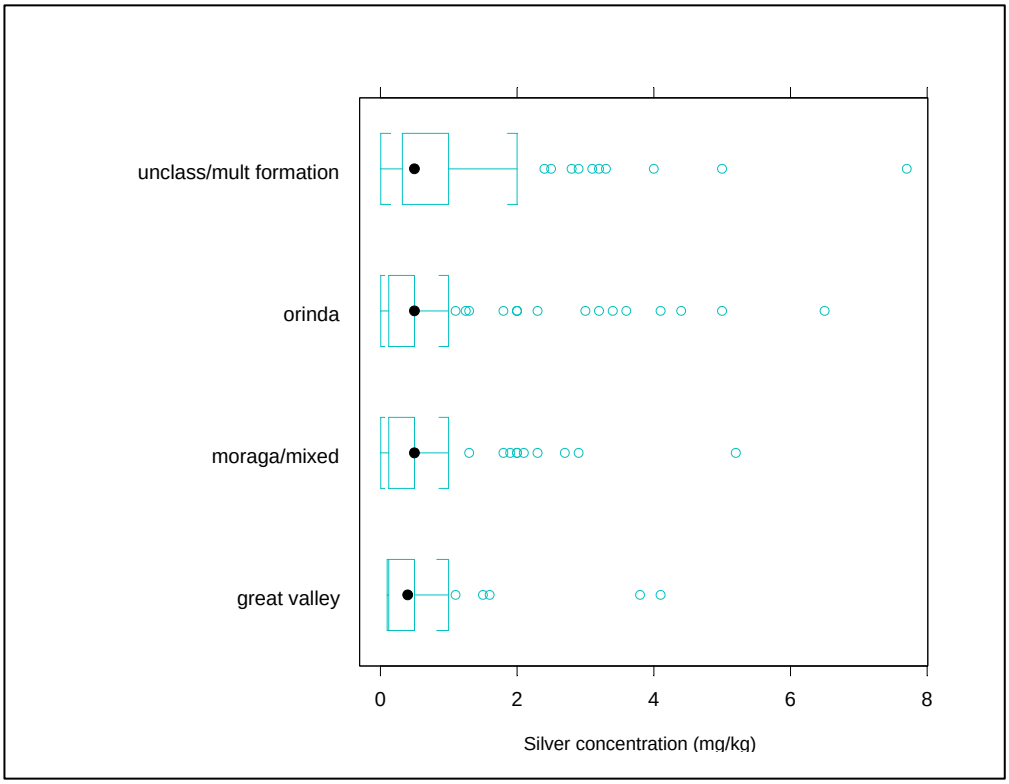
Box Plot of Selenium Soil Concentrations by Geologic Classification



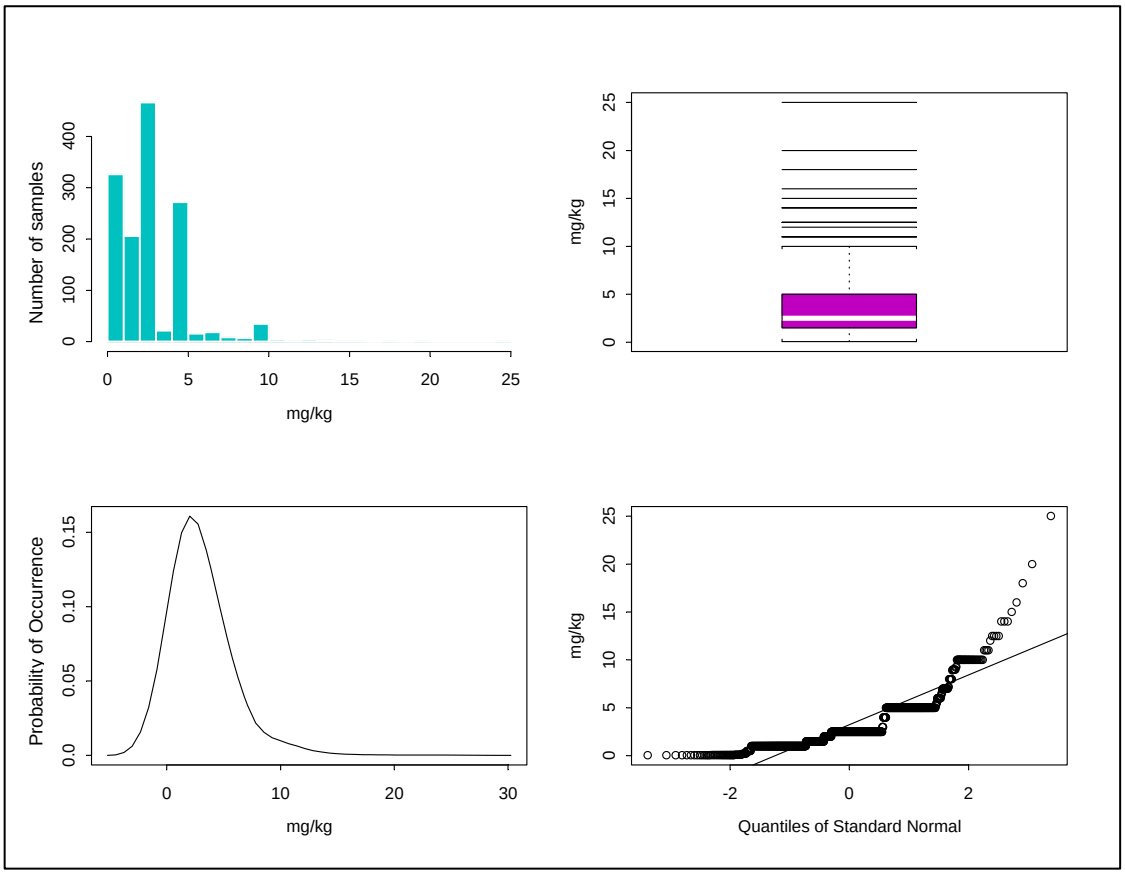
Four-Plot Presentation of Silver Data after the Removal of Possible Outliers



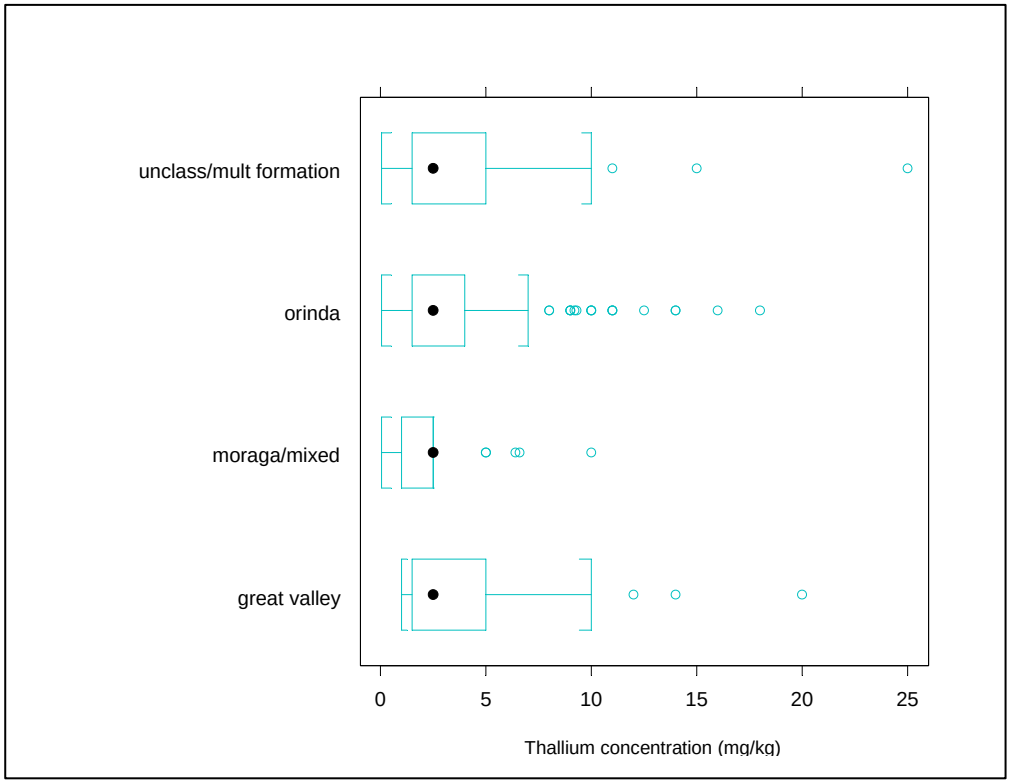
Box Plot of Silver Soil Concentrations by Geologic Classification



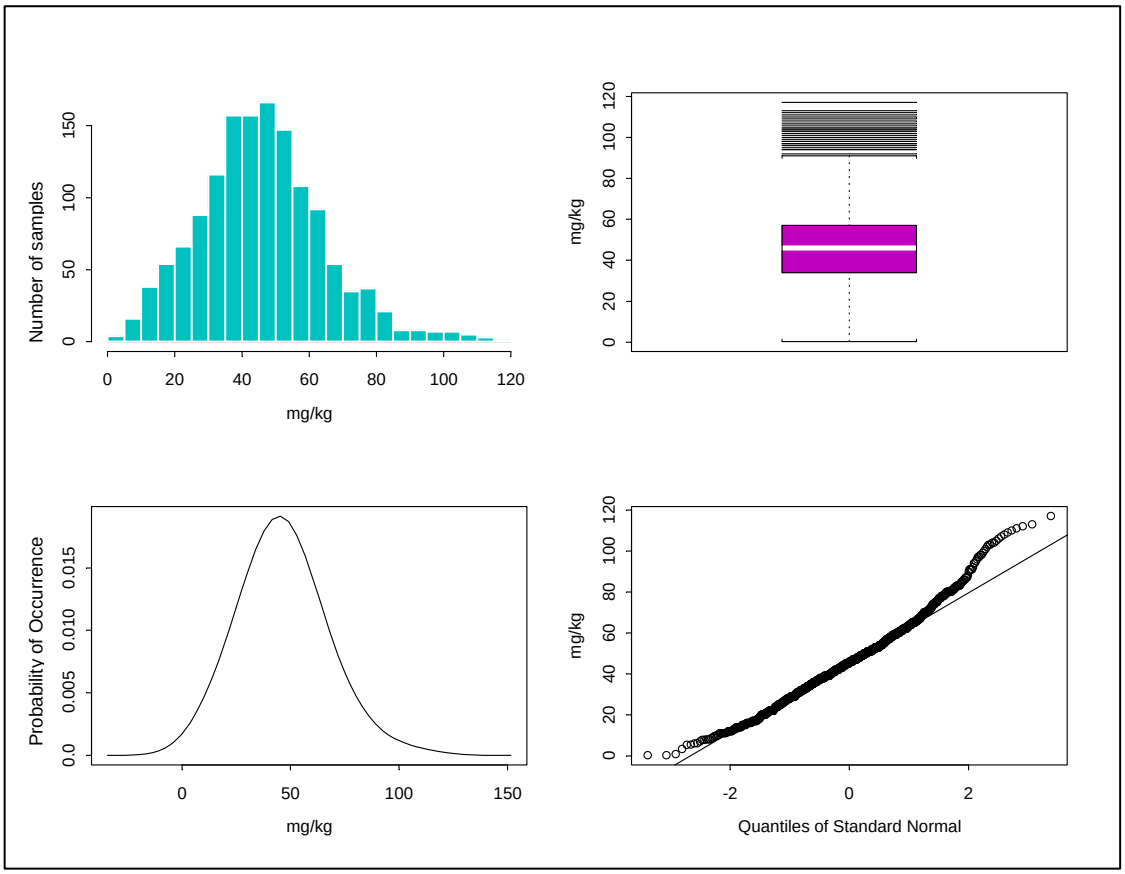
Four-Plot Presentation of Thallium Data after the Removal of Possible Outliers



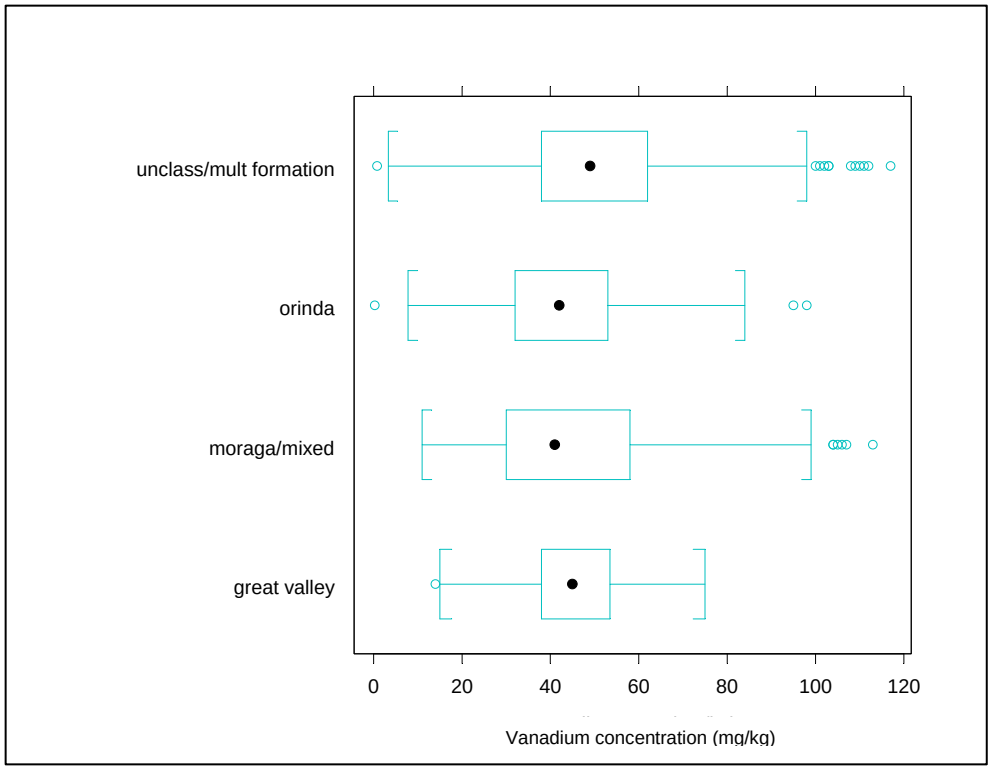
Box Plot of Thallium Soil Concentrations by Geologic Classification



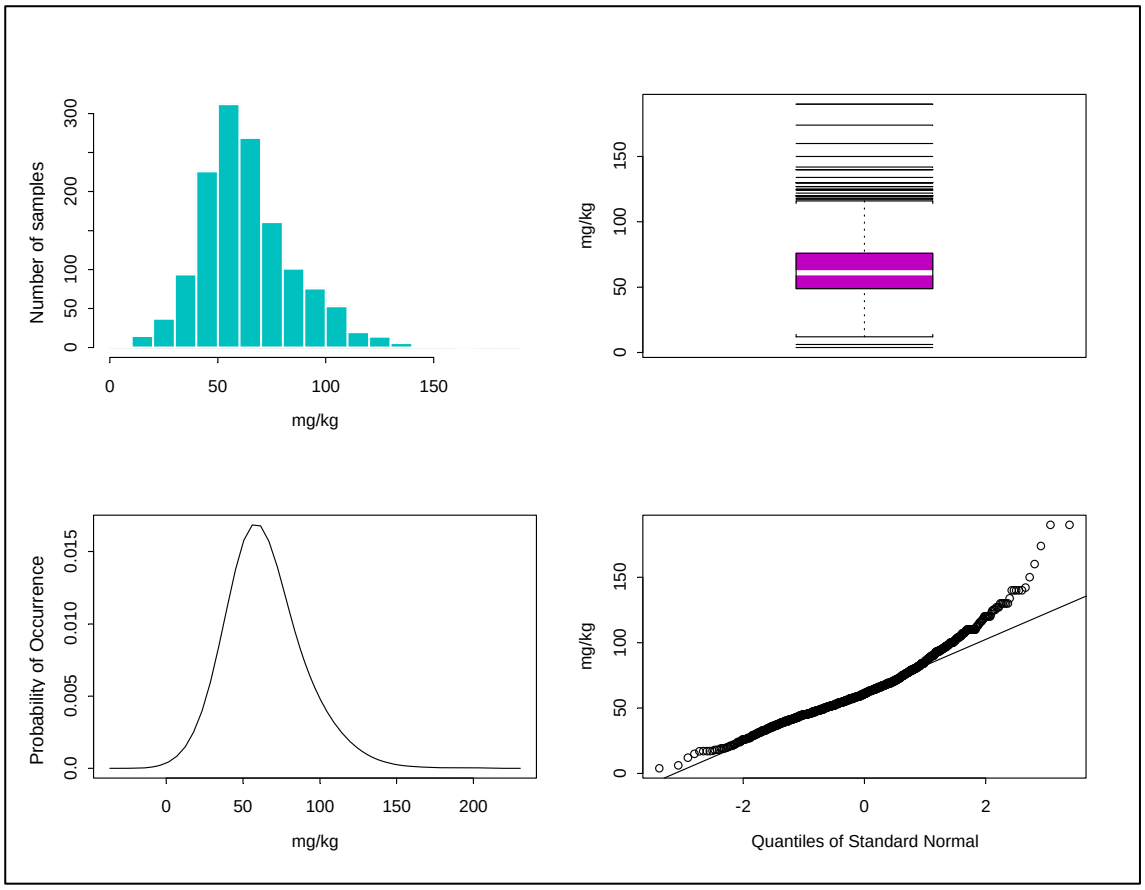
Four-Plot Presentation of Vanadium Data after the Removal of Possible Outliers



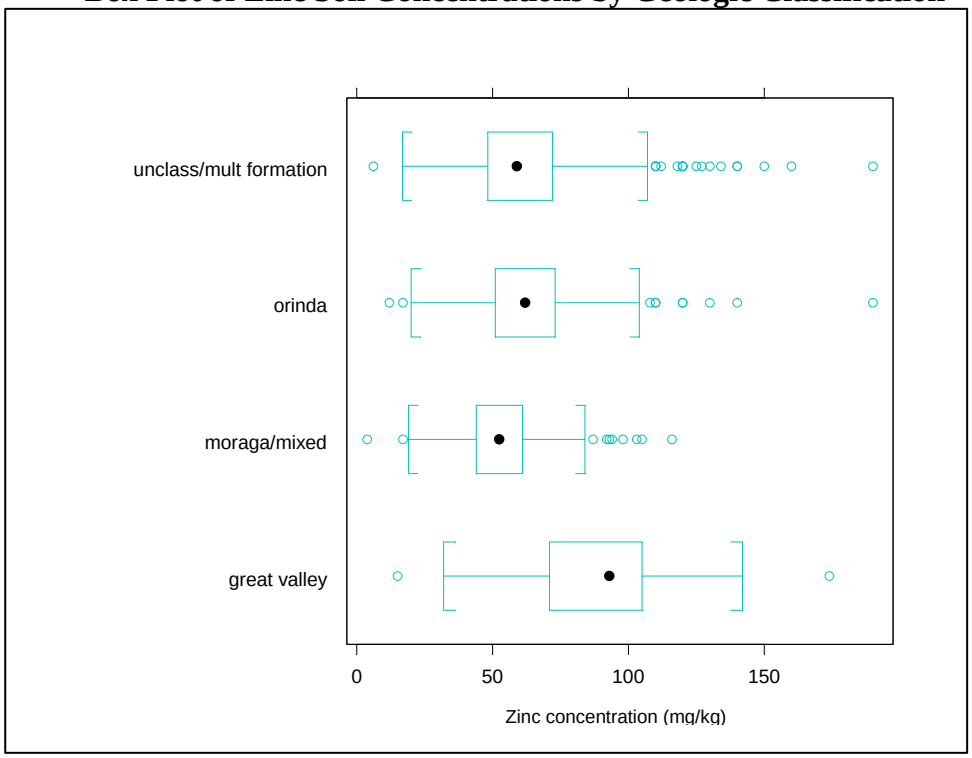
Box Plot of Vanadium Soil Concentrations by Geologic Classification



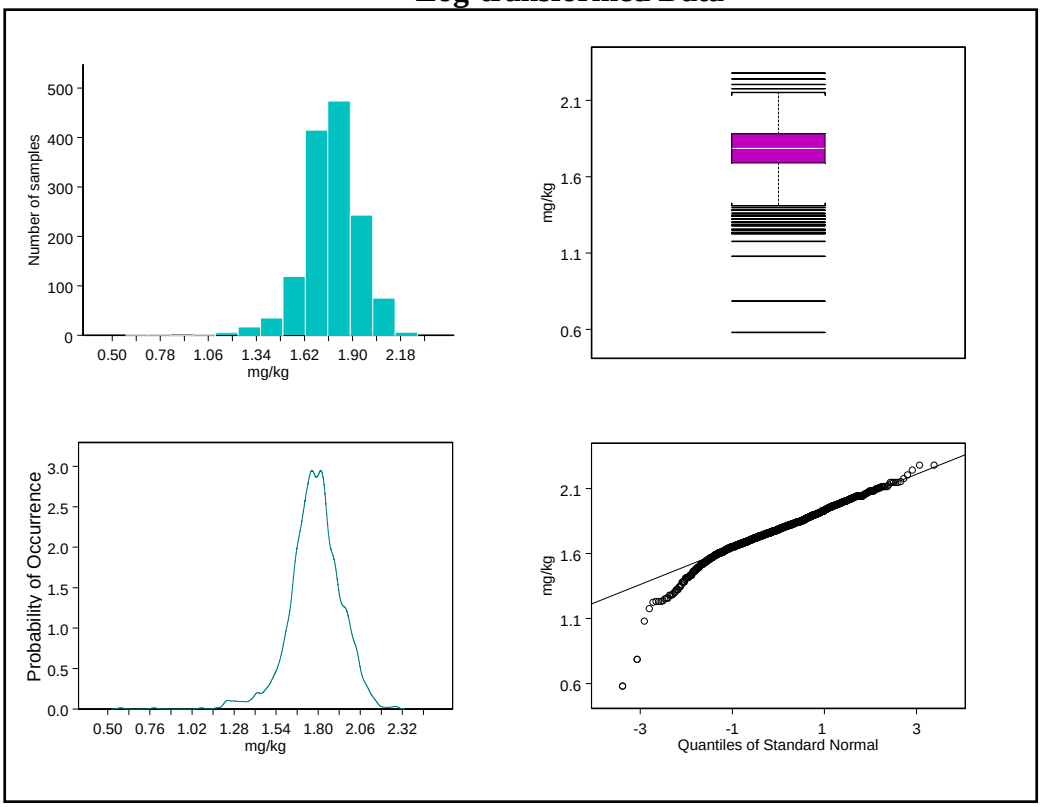
Four-Plot Presentation of Zinc Data after the Removal of Possible Outliers



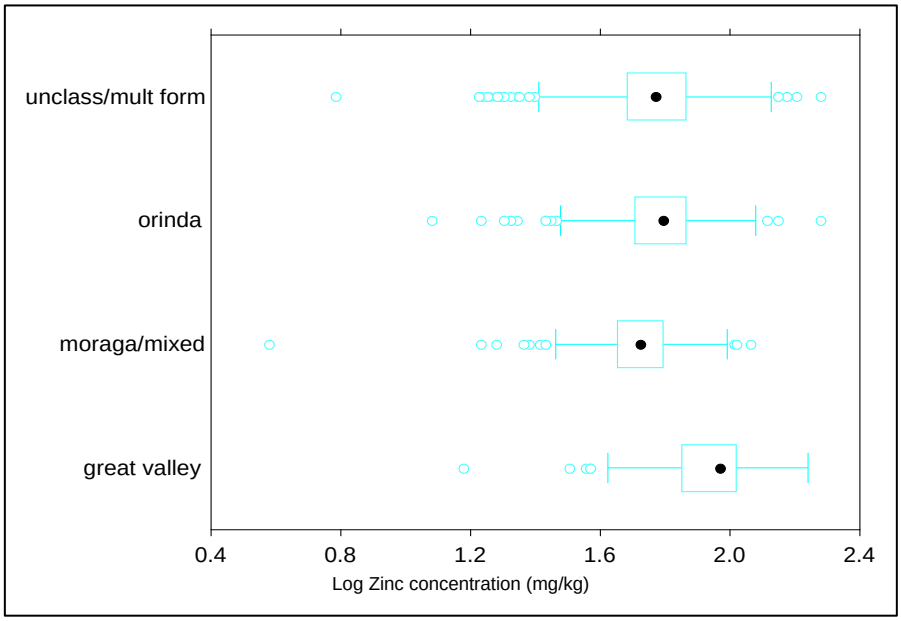
Box Plot of Zinc Soil Concentrations by Geologic Classification



Four-Plot Presentation of Zinc Data after the Removal of Possible Outliers Log-transformed Data



Box Plot of Zinc Soil Concentrations by Geologic Classification Log-transformed Data



APPENDIX C

EXPLANATION OF STATISTICAL METHODS

Appendix C
Description of Statistical Methods

Per USEPA (1989, 1992, and 2000) guidance, statistical methods often are dependent on the distribution of the data. A summary of the statistical methods used to estimate the 95th and 99th percentile background metal concentrations in soil at LBNL is provided in the following sections for data assumed to be normally, lognormally, or nonparametrically distributed. Example calculations based on background metal concentrations measured at LBNL also are provided for each of these distribution types.

PERCENTILE ESTIMATES FOR NORMALLY DISTRIBUTED DATA

Estimates of the 95th and 99th percentiles for normally distributed data were calculated based on USEPA (1989) guidance and the following equations:

$$95^{th}\ Percentile = \overline{X} + 1.645\,s \qquad (1)$$

$$99^{th}\ Percentile = \overline{X} + 2.33\,s \qquad (2)$$

Where:

$\overline{X}$ = arithmetic average metal concentration of the untransformed data

s = standard deviation of the untransformed data

<u>Example Calculation.</u> As discussed in the text, background concentrations of cobalt fit a normal distribution. The 95th and 99th percentile concentrations of cobalt were calculated as shown in equations 3 and 4 (refer to Table 3 for cobalt arithmetic mean [$\overline{X}$] and standard deviation [s]).

$$Cobalt\ 95^{th}\ Percentile = \overline{X} + 1.645\,s = 14.1 + 1.645(4.79) = 22\,mg/kg \qquad (3)$$

$$Cobalt\ 99^{th}\ Percentile = \overline{X} + 2.33\,s = 14.1 + 2.33(4.79) = 25\,mg/kg \qquad (4)$$

PERCENTILE ESTIMATES FOR LOGNORMALLY DISTRIBUTED DATA

Estimates of the 95th and 99th percentiles for lognormally distributed data were calculated based on USEPA (1997) guidance using the Maximum Likelihood Estimate (MLE) method. In order to use the MLE method, the data were log-transformed (i.e., the natural logarithm for each concentration was calculated). The average and standard deviation of the log-transformed data were calculated and designated as $\hat{X}$ and $\hat{s}$, respectively. Percentile estimates of the untransformed data were calculated by taking the inverse logarithms of the 95th and 99th percentile estimates of the log-transformed data as shown in equations 5 and 6.

$$95^{th}\ Percentile = \exp(\hat{X} + 1.645\,\hat{s}) \qquad (5)$$

$$99^{th}\ Percentile = \exp(\hat{X} + 2.33\,\hat{s}) \qquad (6)$$

Where:

$\hat{X}$ = arithmetic average metal concentration of the natural logarithm-transformed data

$\hat{s}$ = standard deviation of the natural logarithm-transformed data

Example Calculation. As discussed in the text, background concentrations of lead fit a lognormal distribution. The 95th and 99th percentile concentrations of lead were calculated as shown in equations 7 and 8. The arithmetic average and standard deviation background metal concentrations for lead were 7.0 mg/kg and 6.5 mg/kg, respectively (Table 3). The natural logarithm-transformed arithmetic average ($\hat{X}$) and standard deviation ($\hat{s}$) concentrations for lead were 1.54 ln(mg/kg) and 1.07 ln(mg/kg), respectively. Percentile estimates of the untransformed data were calculated by taking the inverse logarithms of the 95th and 99th percentile estimates of the log-transformed data as shown in equations 7 and 8.

$$Lead\ 95^{th}\ Percentile = \exp(1.54 + 1.645(1.07)) = 27\, mg/kg \qquad (7)$$

$$Lead\ 99^{th}\ Percentile = \exp(1.54 + 2.33(1.07)) = 57\, mg/kg \qquad (8)$$

PERCENTILE ESTIMATES FOR NONPARAMETRICALLY DISTRIBUTED DATA

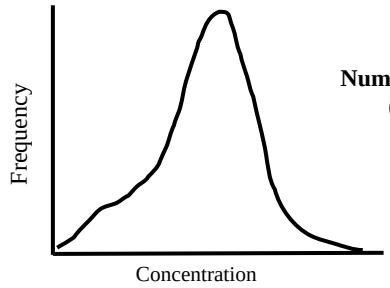
The "bootstrap" method (Efron, 1981), a nonparametric method recommended by USEPA (1997), was used to estimate 95th percentiles for nonparametrically distributed data. The bootstrap method is a computer-based resampling method for assigning measures of accuracy to statistical estimates (Efron and Tibshirani, 1993). Resampling consists of randomly selecting results from the original data set to create new, "resampled" data sets. The Excel 97 add-in software, *Resampling Stats* (Blank, *et al.*, 1999) was used to perform the bootstrap method. It does not require assumptions regarding the statistical distribution of the underlying population (Efron, 1981). Bootstrap 95th or 99th percentile estimates were calculated as follows:

1. Randomly resample the data set with replacement;
2. Estimate the 95th or 99th percentiles of the resampled data set;
3. Perform steps one and two 3,000 times and create new data sets of 3,000 resampled 95th or 99th percentile estimates, respectively; and
4. Estimate the 95th or 99th percentiles of the new data sets.
5. Repeat steps 1 through 4 using 10,000 instead of 3,000 iterations to ensure that the bootstrapped percentiles are stable.

The percentiles estimated in step 4 are the 95th or 99th percentile estimates of the original data set. The bootstrap method described above is summarized graphically in the following figure.

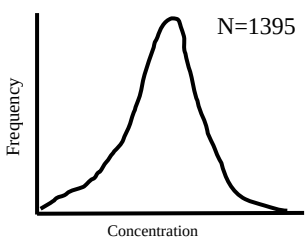
FLOW CHART OF BOOTSTRAP METHODOLOGY

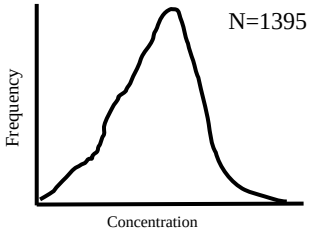
Original Background Data Set (e.g., lead)

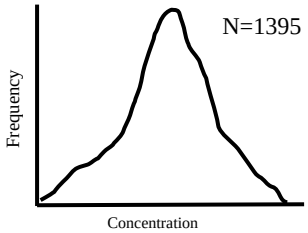


Number of Samples (N) = 1395

Generate 3,000 new (i.e., "resampled) data sets by randomly selecting sample results from the original background data set.

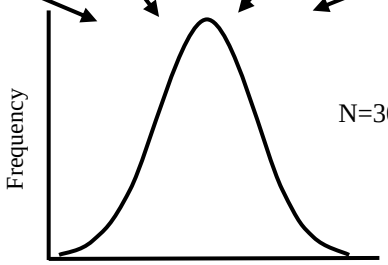






. . . continue until there are 3,000 "resampled" data sets

Select the 95th (or 99th) percentiles from the "resampled data sets and plot in a single graph.

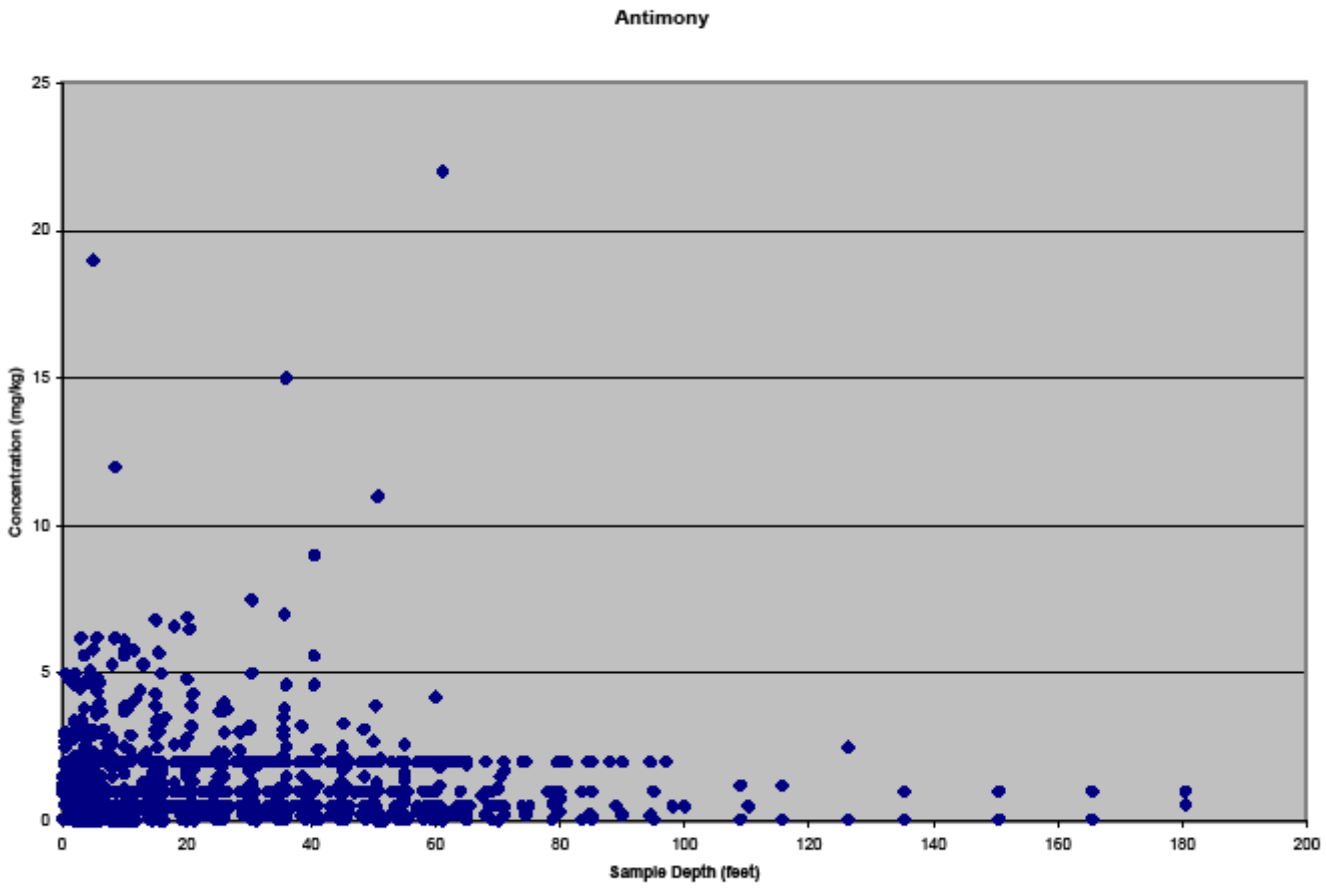


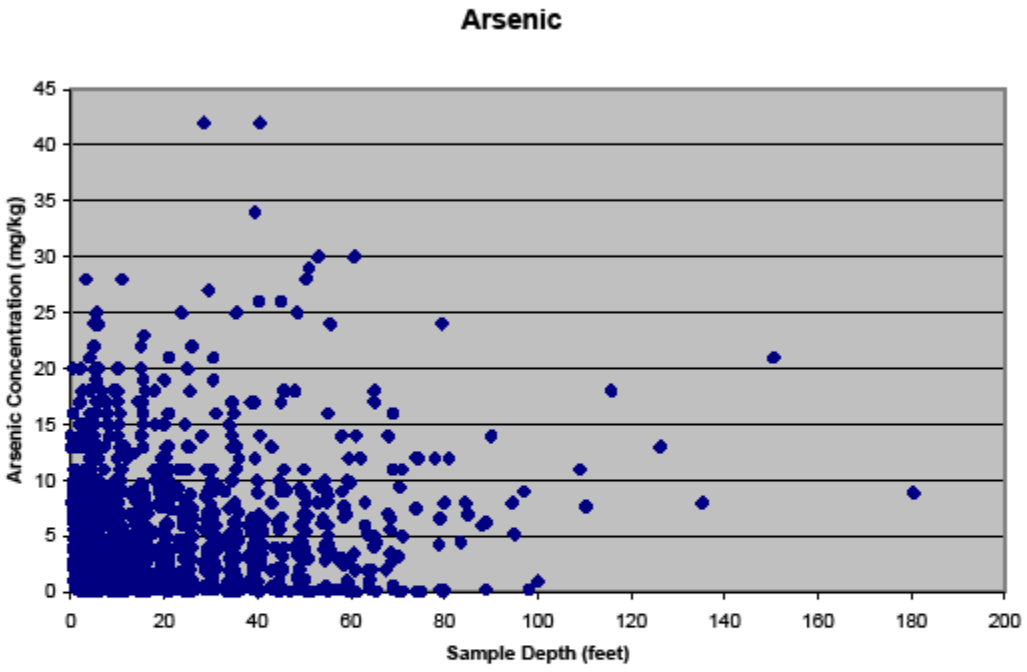
Select the 95th (or 99th) percentile from the plot of "resampled" 95th (or 99th) percentiles, to yield the 95th (or 99th) percentile estimate of the original data set.

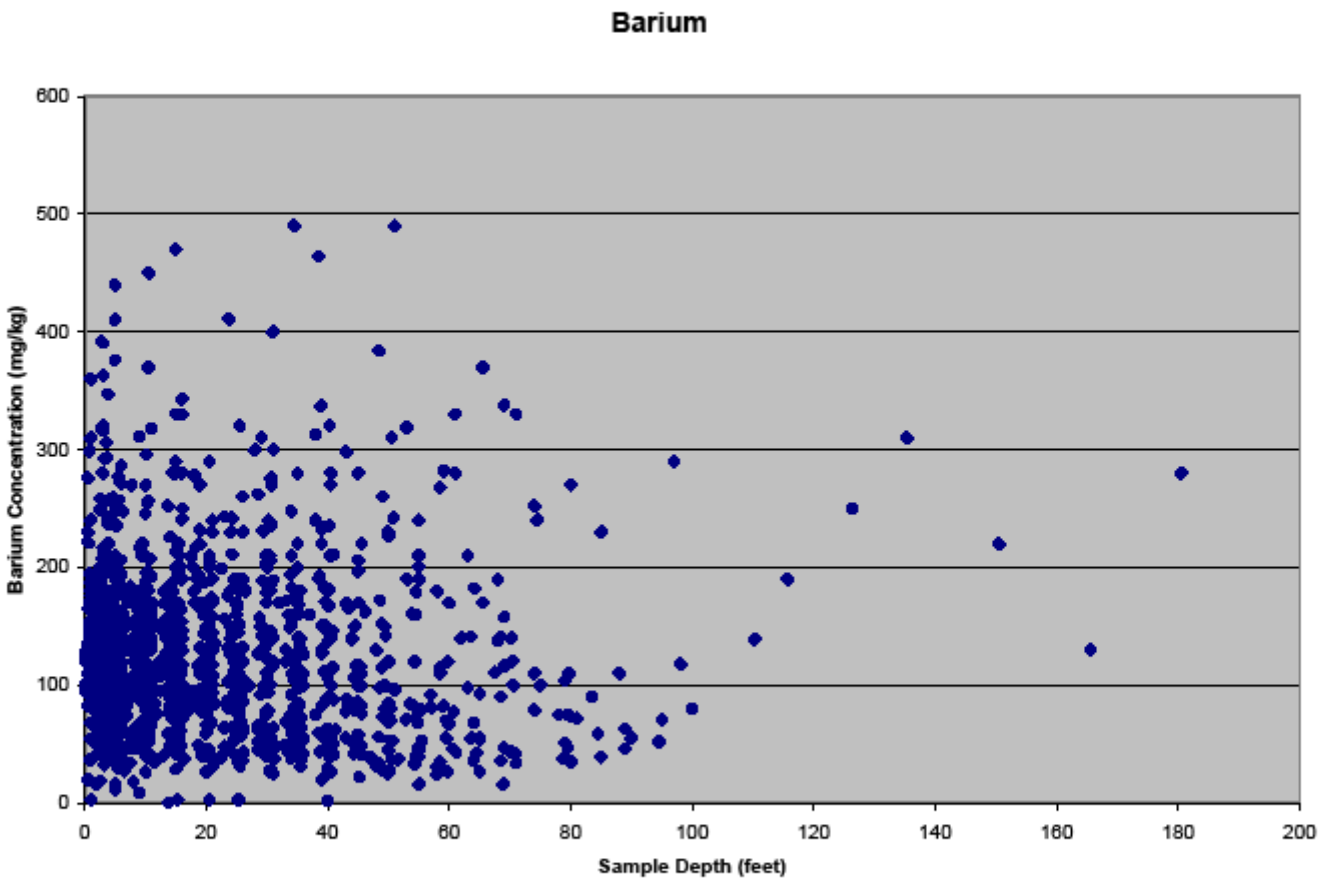
APPENDIX D

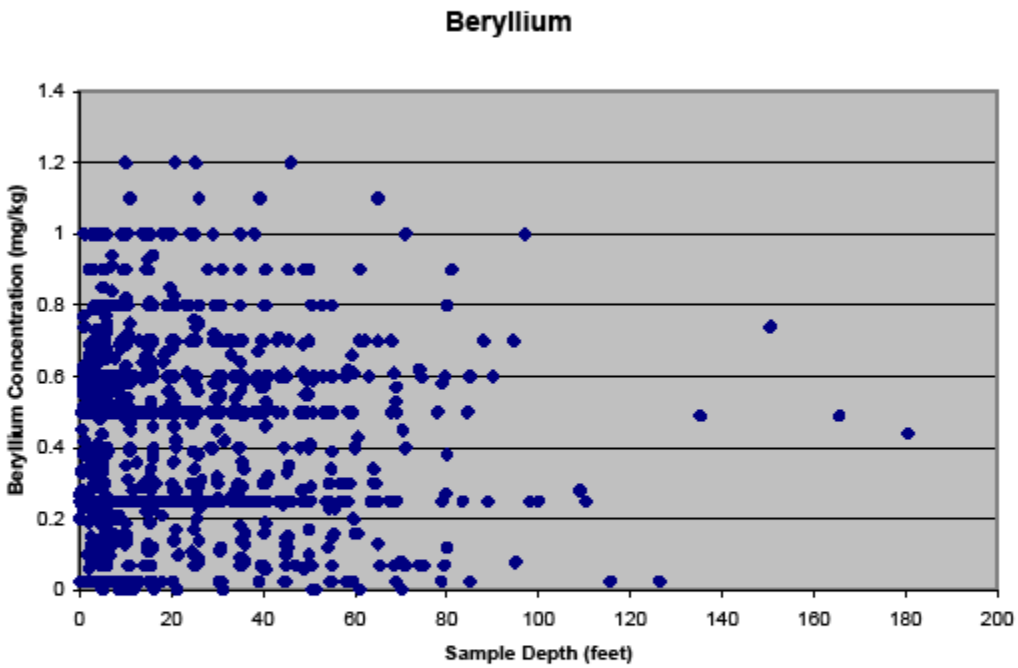
DEPTH VERSUS CONCENTRATION CORRELATION COEFFICIENTS AND SCATTERPLOTS

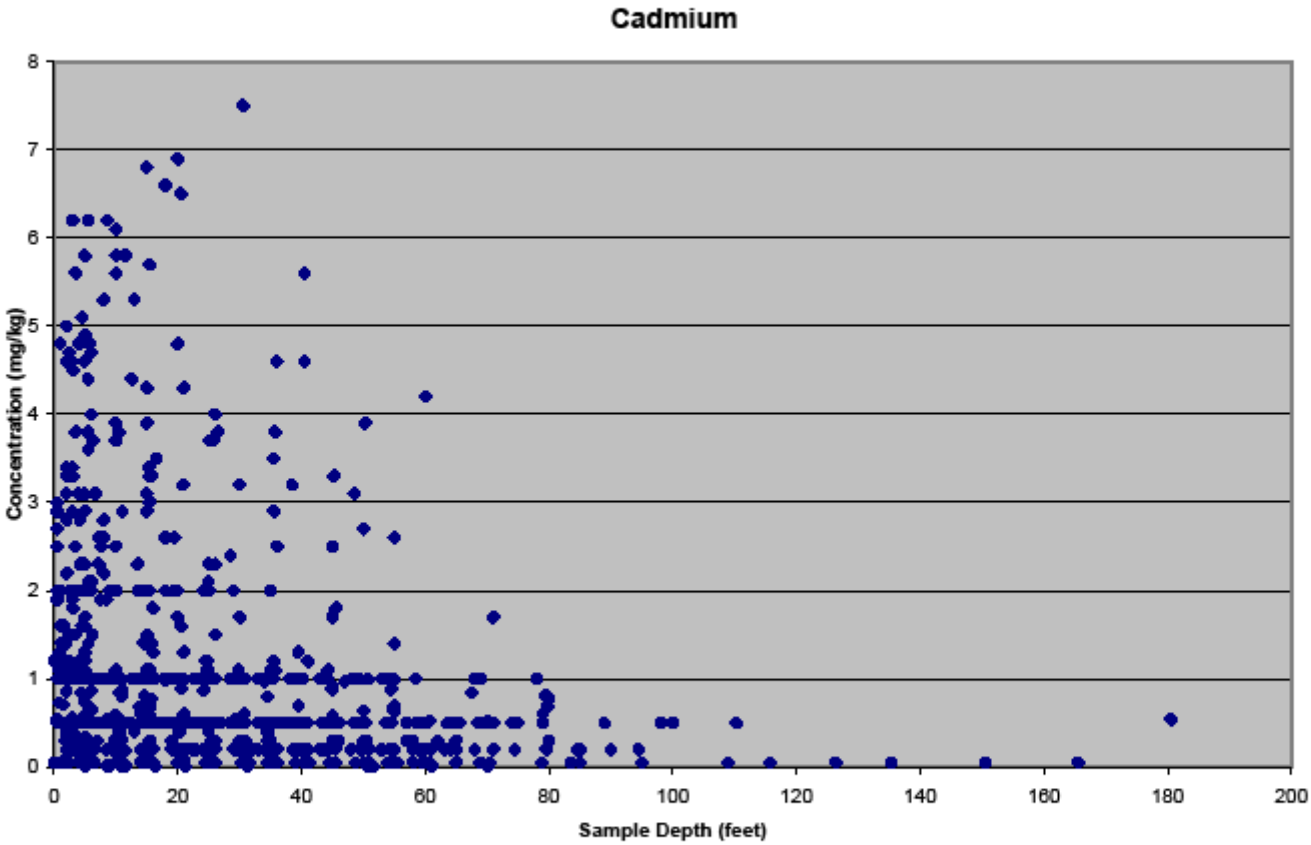
Correlation Coefficients for Depth / Concentration Scatterplots	
Metal	**Correlation Coefficient**
Antimony	0.094
Arsenic	0.061
Barium	-0.020
Beryllium	-0.042
Cadmium	-0.159
Chromium	-0.123
Cobalt	-0.061
Copper	-0.028
Lead	-0.078
Mercury	-0.052
Molybdenum	-0.055
Nickel	-0.058
Selenium	0.191
Silver	-0.117
Thallium	0.014
Vanadium	-0.126
Zinc	0.016

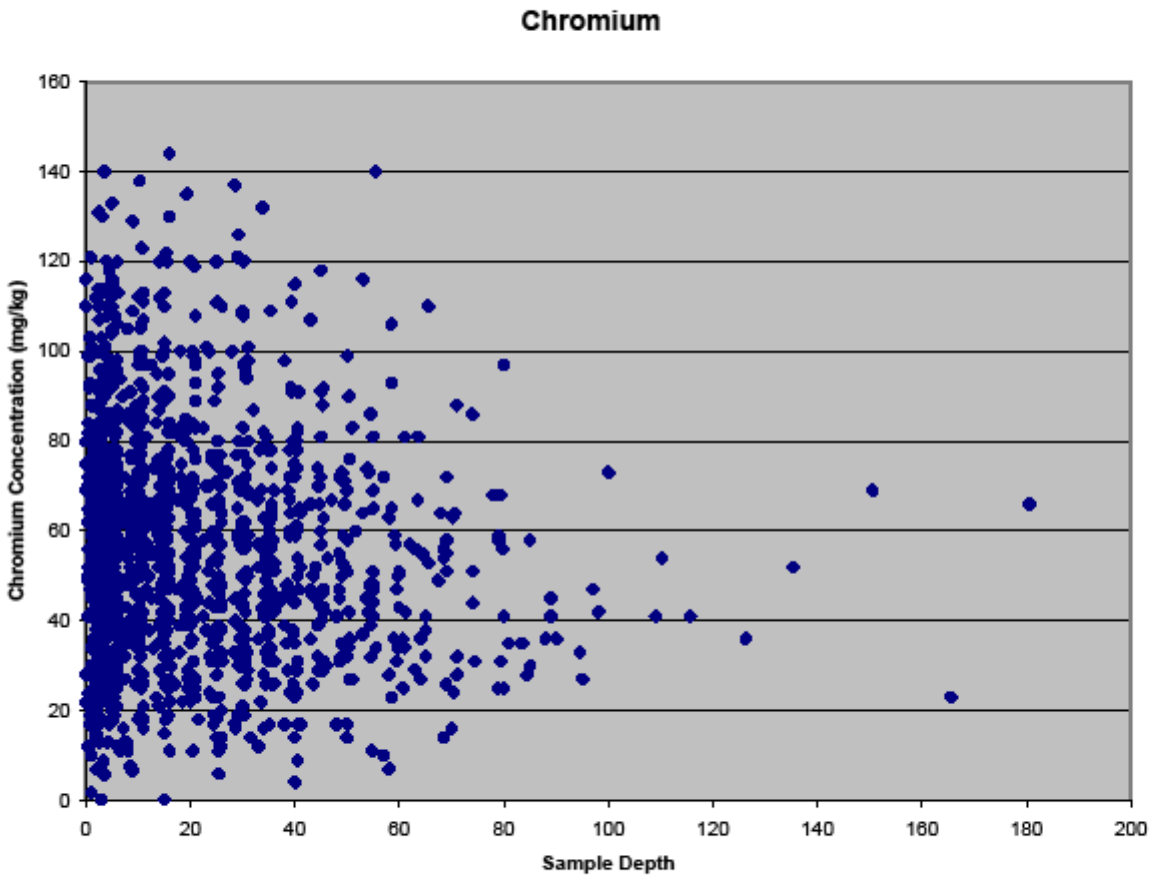

Antimony
Concentration (mg/kg)
25
20
15
10
5
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)

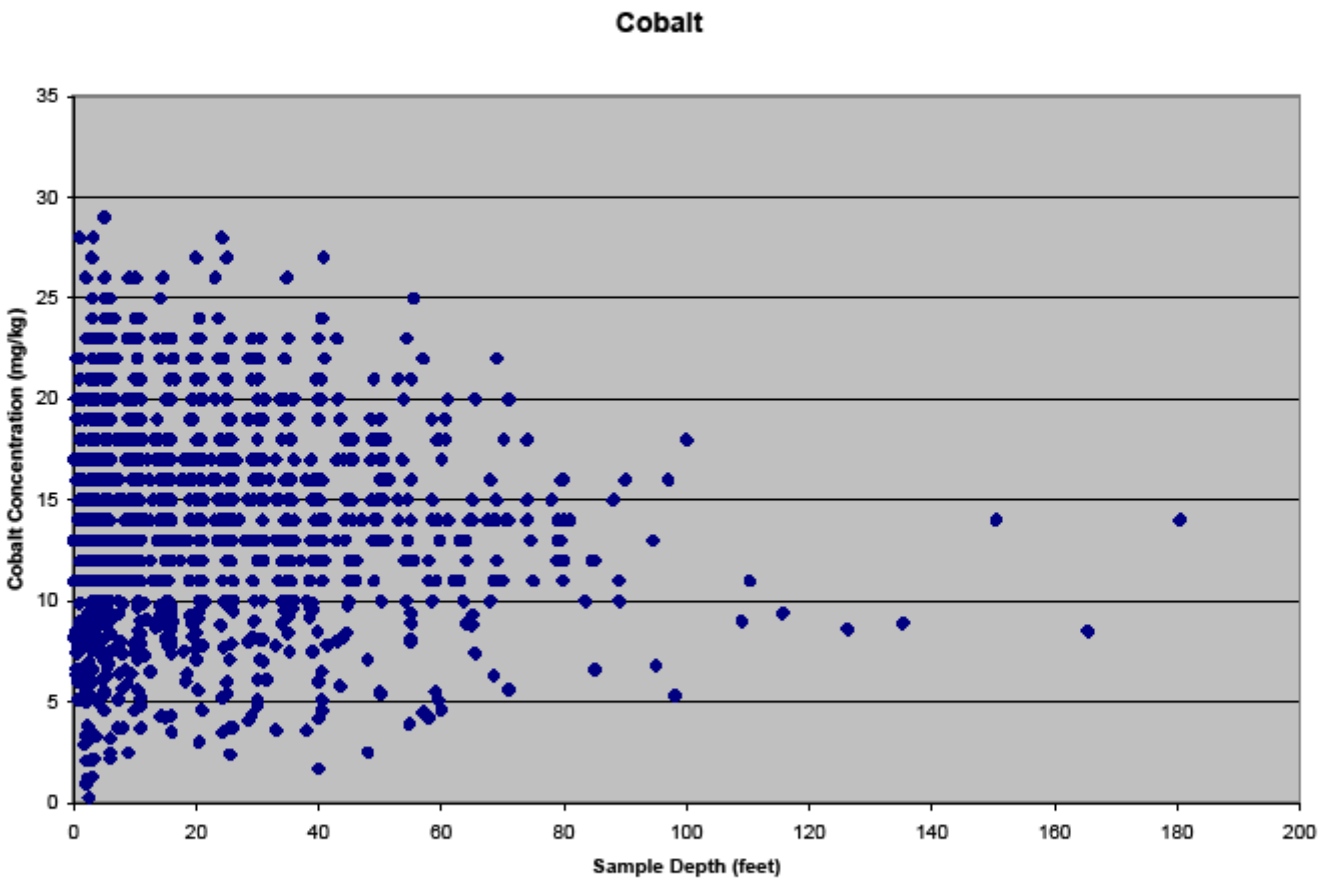

Arsenic
Arsenic Concentration (mg/kg)
45
40
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)

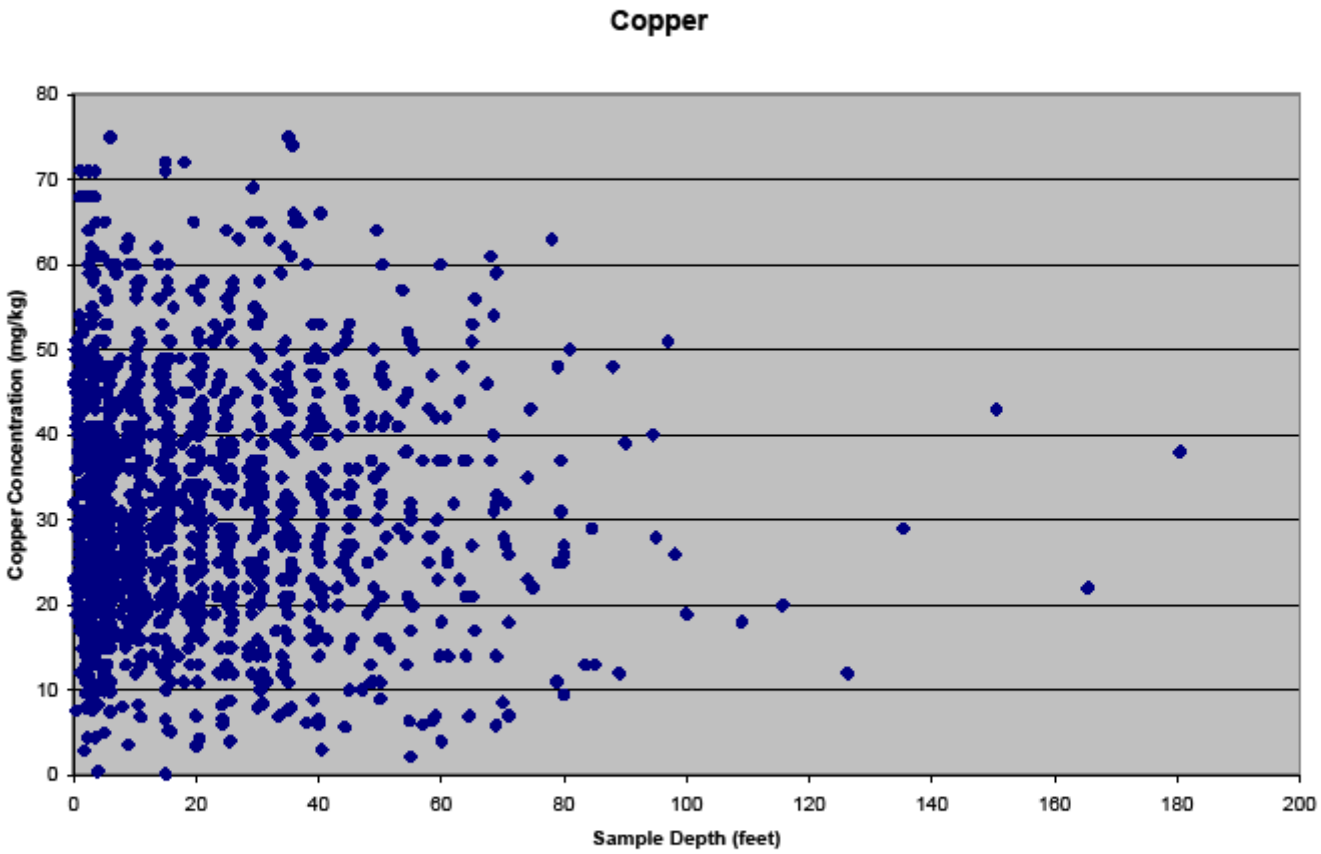

Barium
Barium Concentration (mg/kg)
600
500
400
300
200
100
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)

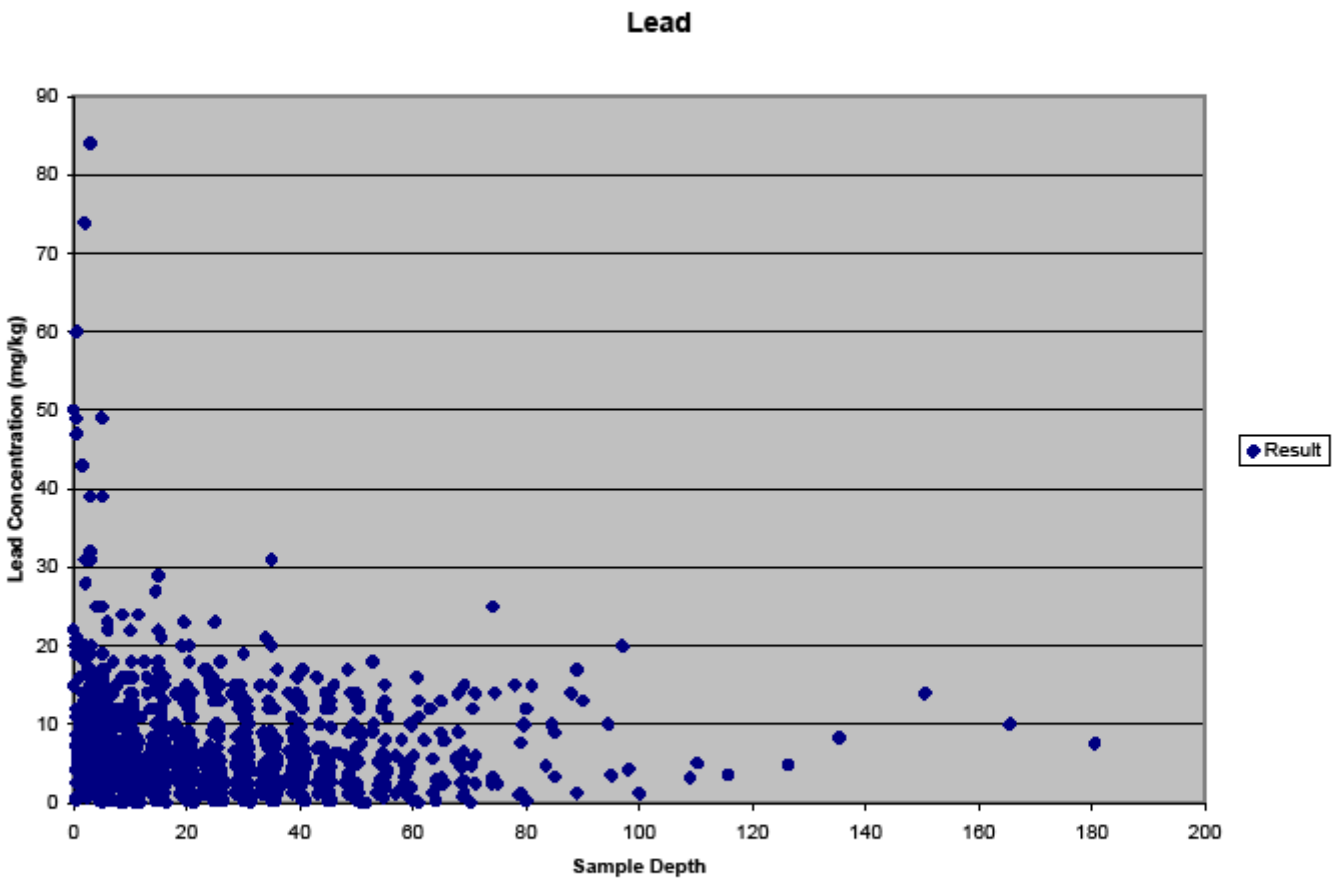

Beryllium
Beryllium Concentration (mg/kg)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)

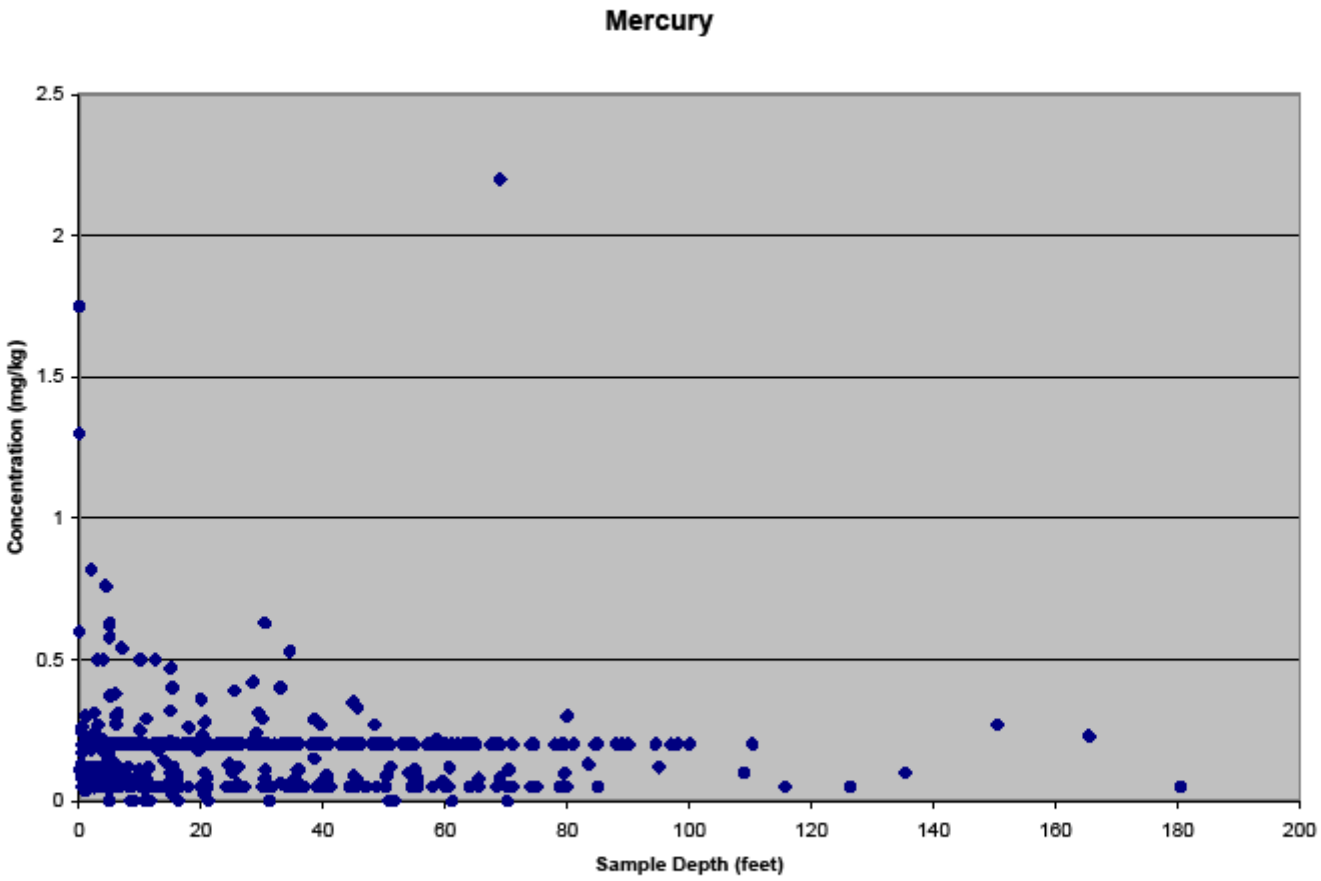

Cadmium
Concentration (mg/kg)
Sample Depth (feet)
0
1
2
3
4
5
6
7
8
0
20
40
60
80
100
120
140
160
180
200

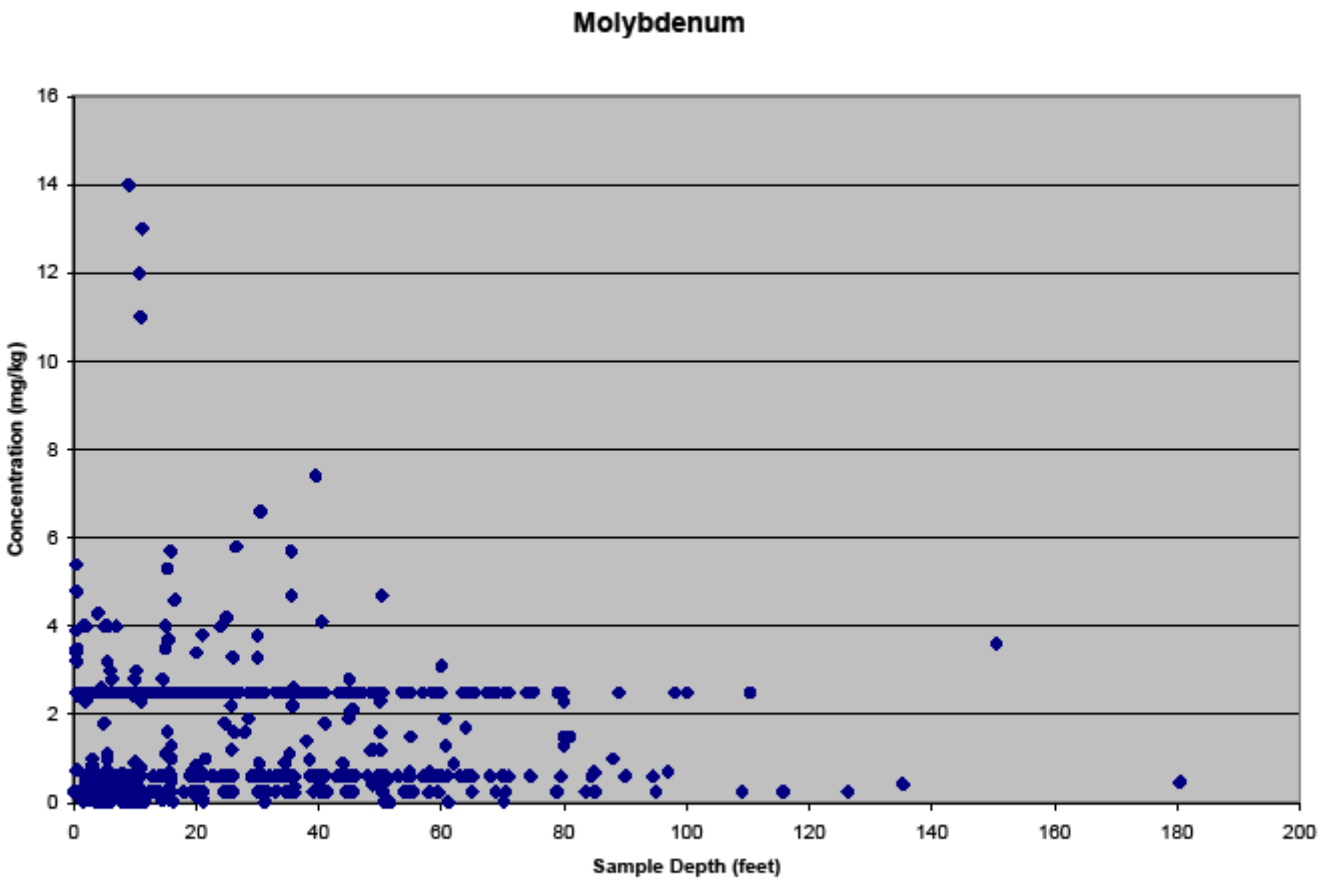

Chromium
Chromium Concentration (mg/kg)
Sample Depth
0
20
40
60
80
100
120
140
160
180
200

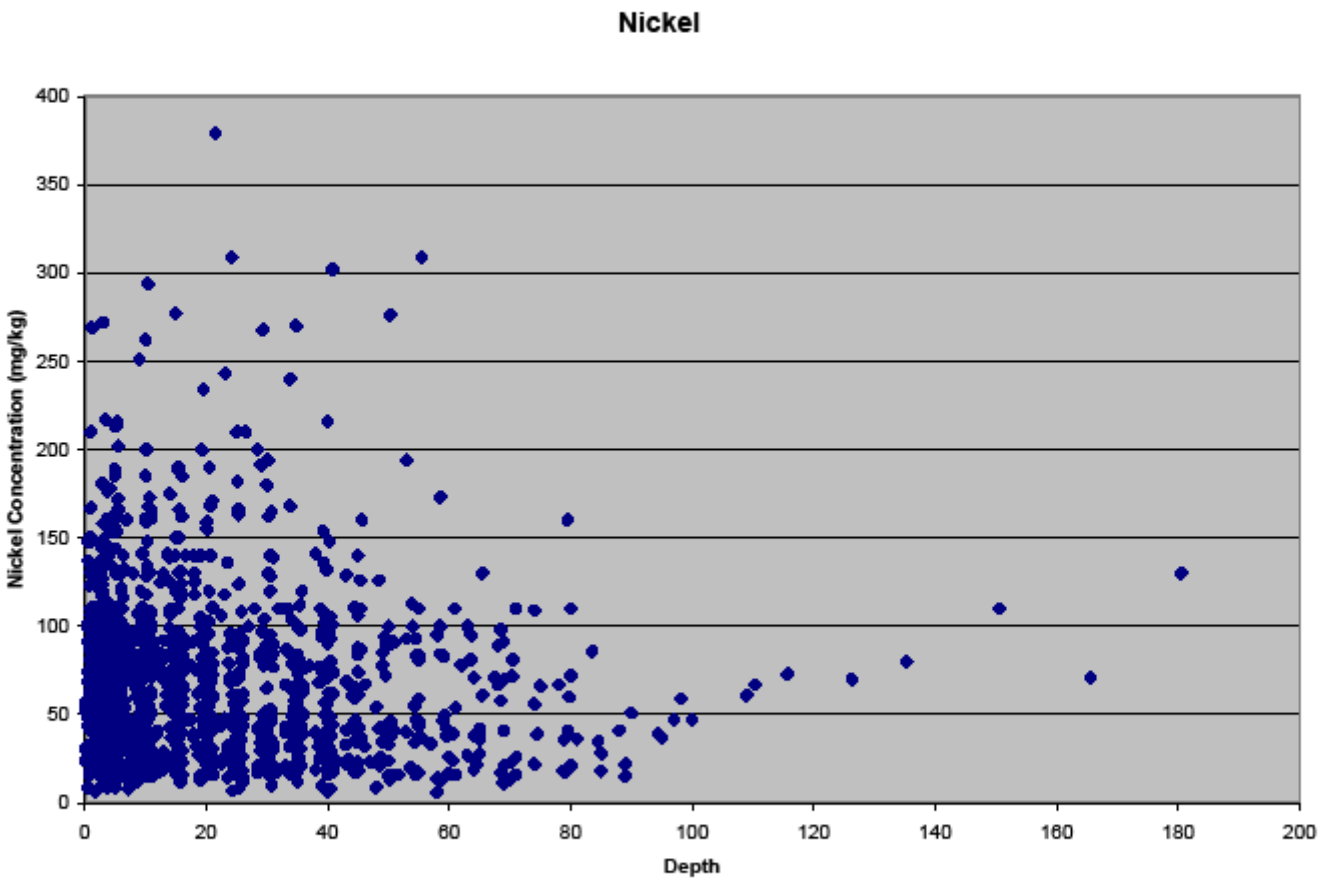

Cobalt
Cobalt Concentration (mg/kg)
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)

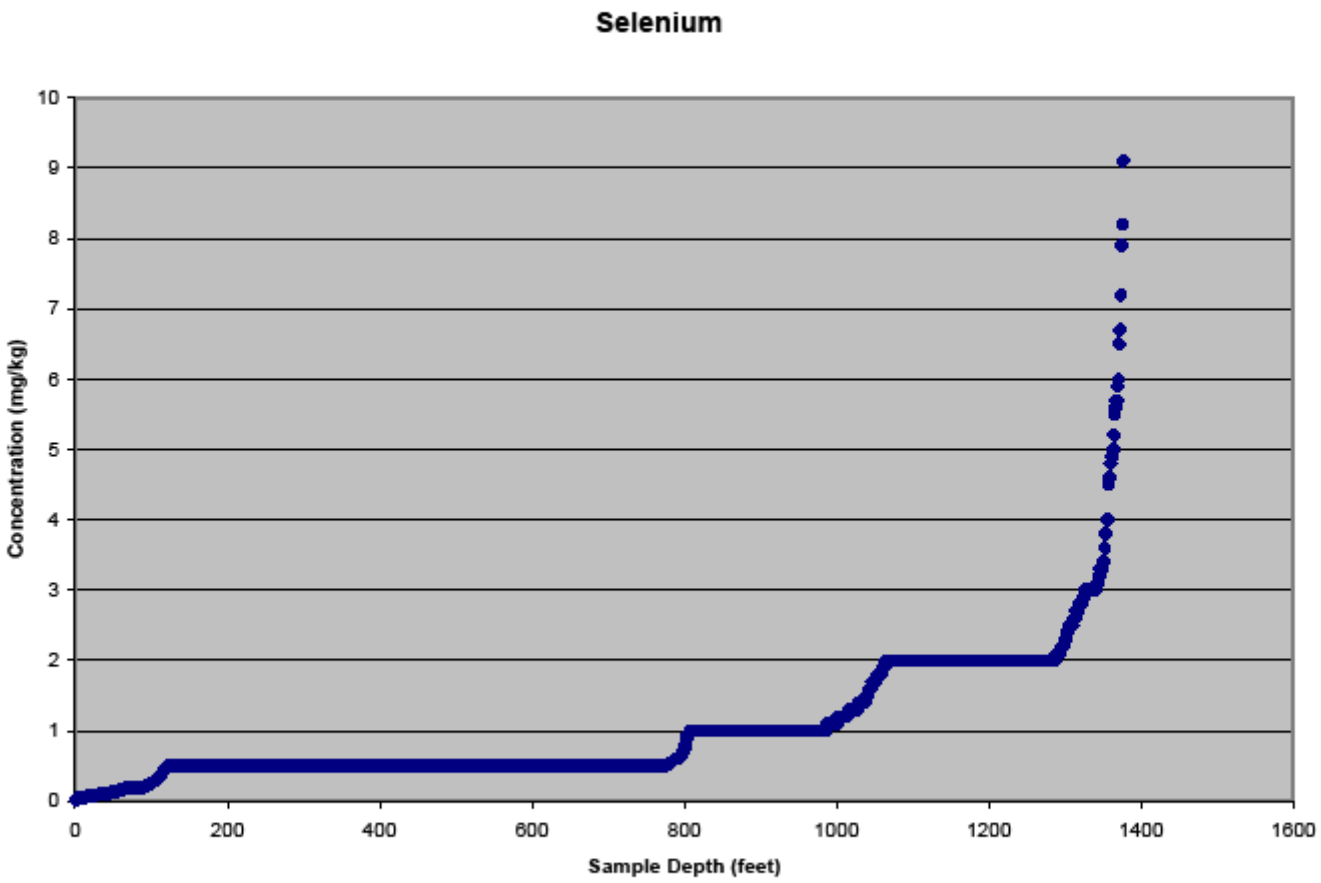

Copper
Copper Concentration (mg/kg)
Sample Depth (feet)
0
10
20
30
40
50
60
70
80
0
20
40
60
80
100
120
140
160
180
200

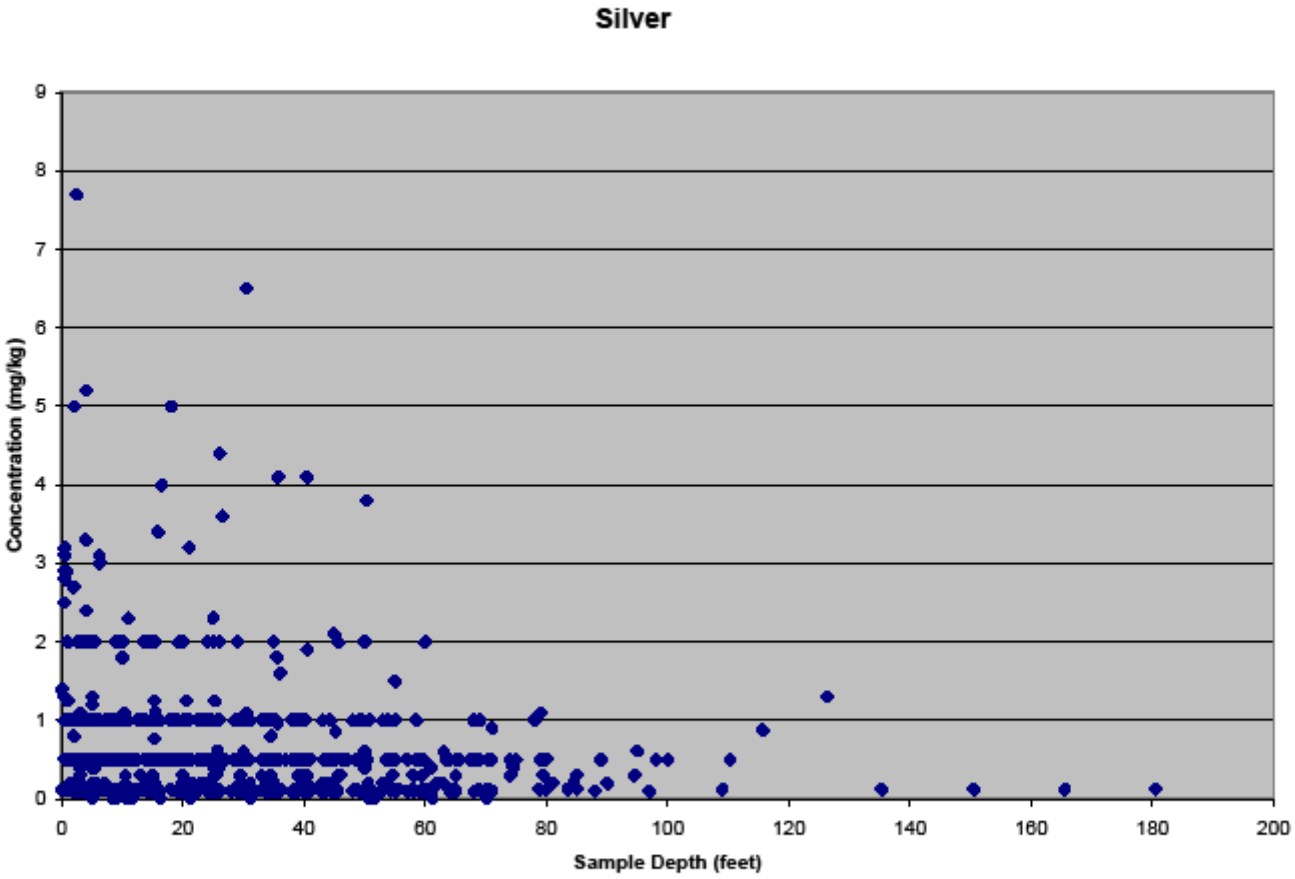

Lead
Lead Concentration (mg/kg)
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth
Result

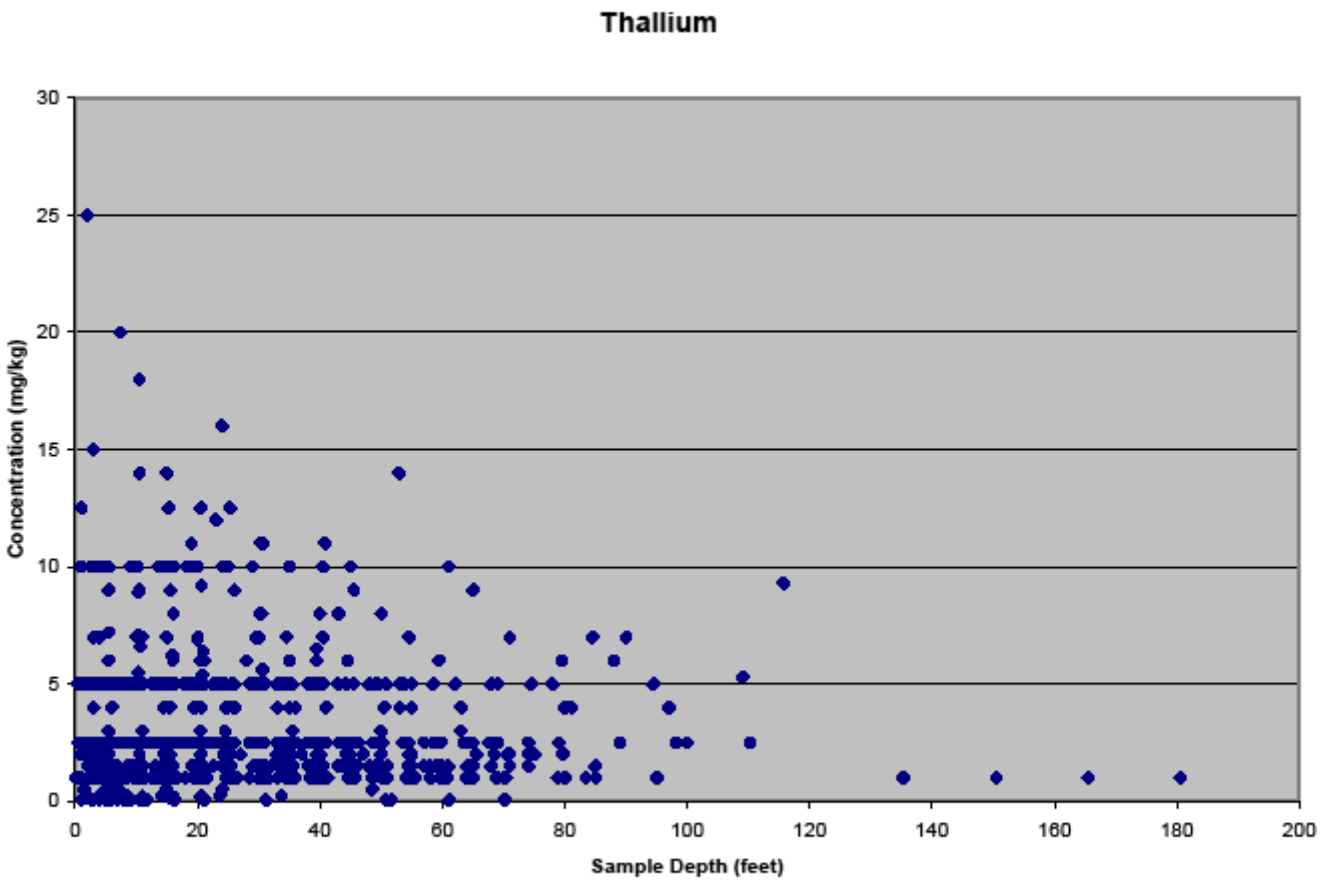

Mercury
Concentration (mg/kg)
2.5
2
1.5
1
0.5
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)

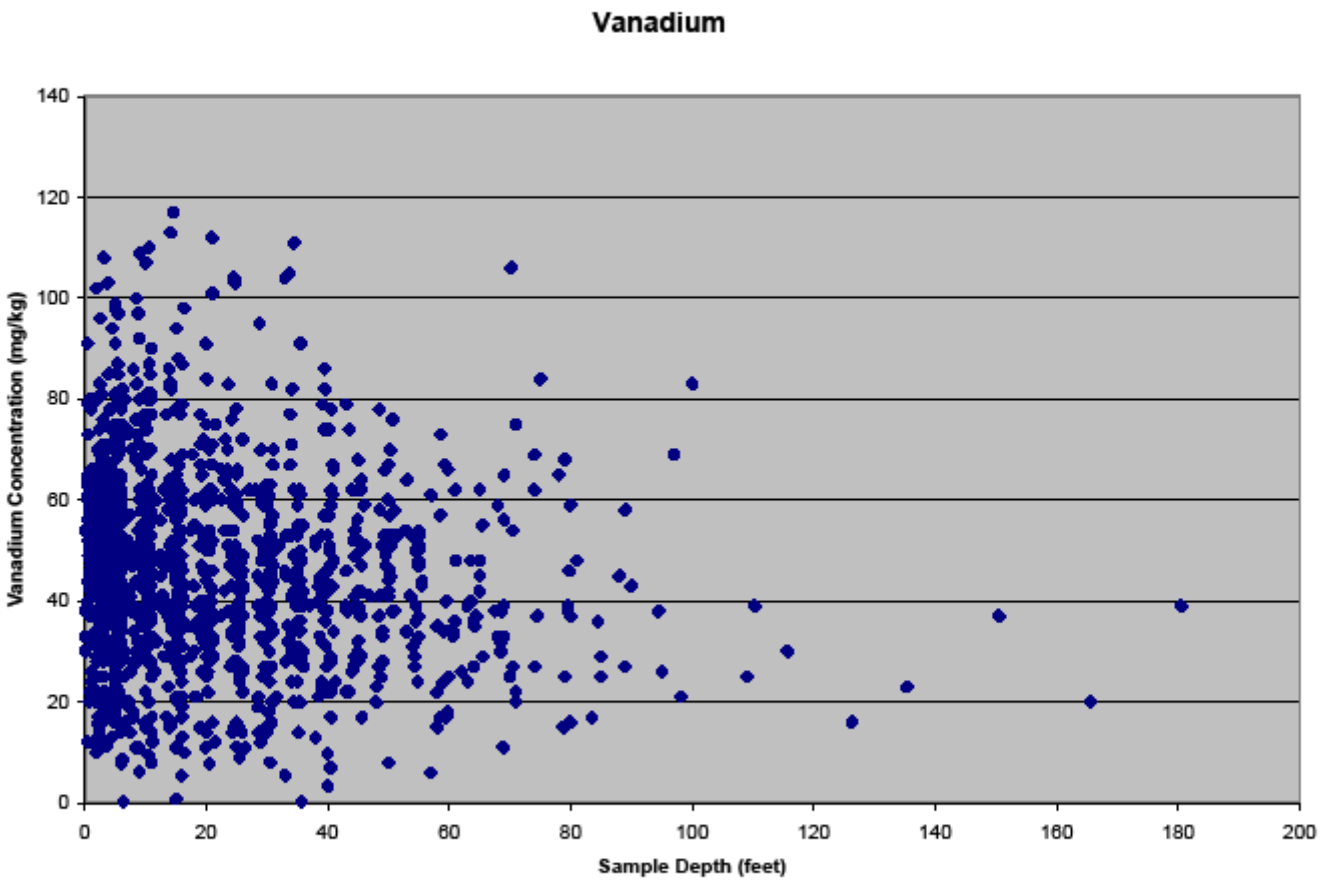

Molybdenum
Concentration (mg/kg)
Sample Depth (feet)
0
2
4
6
8
10
12
14
16
0
20
40
60
80
100
120
140
160
180
200

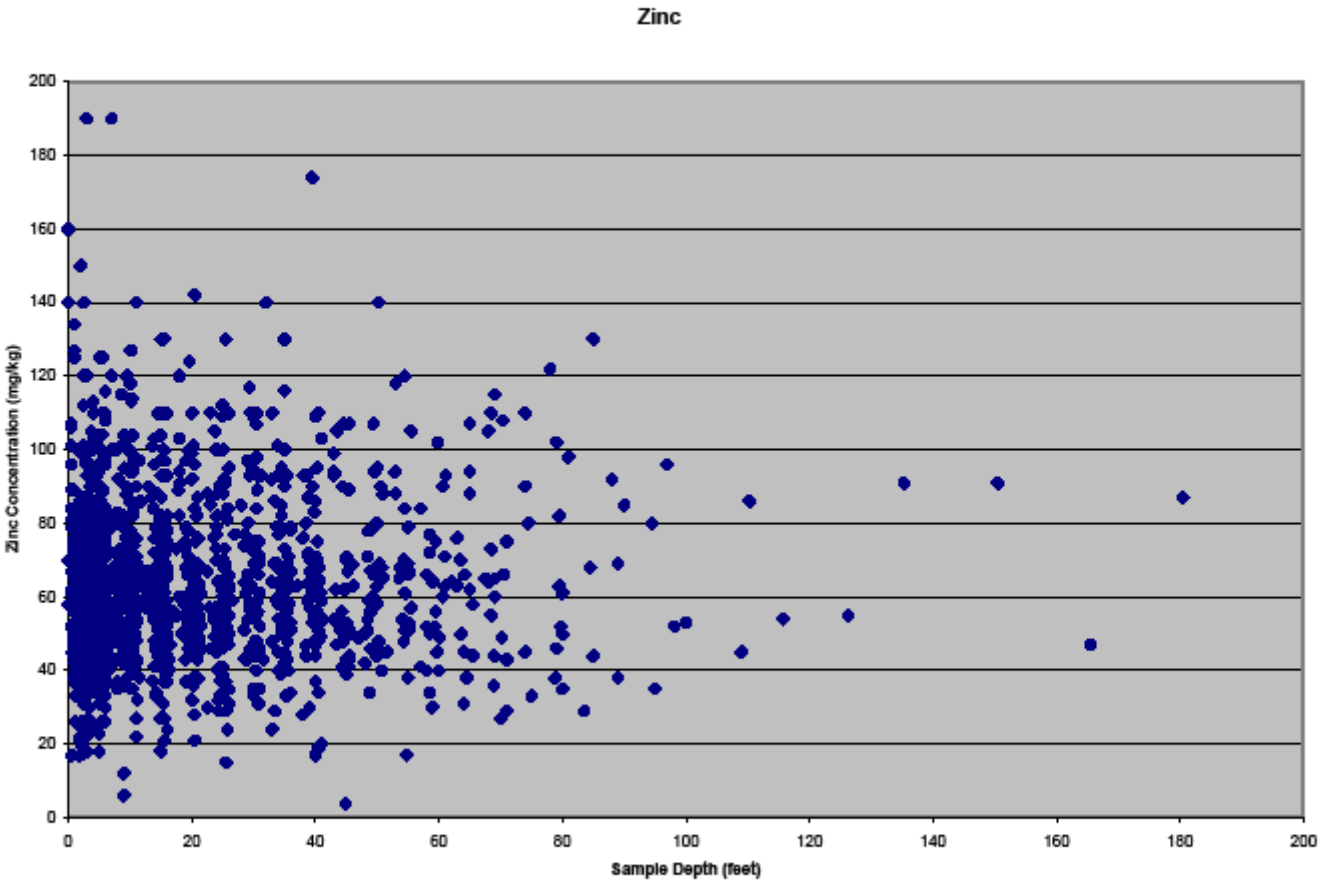

Nickel
Nickel Concentration (mg/kg)
400
350
300
250
200
150
100
50
0
0
20
40
60
80
100
120
140
160
180
200
Depth


Selenium
Concentration (mg/kg)
Sample Depth (feet)
0
1
2
3
4
5
6
7
8
9
10
0
200
400
600
800
1000
1200
1400
1600


Silver
Concentration (mg/kg)
Sample Depth (feet)
0
1
2
3
4
5
6
7
8
9
0
20
40
60
80
100
120
140
160
180
200


Thallium
Concentration (mg/kg)
30
25
20
15
10
5
0
0
20
40
60
80
100
120
140
160
180
200
Sample Depth (feet)


Vanadium
Vanadium Concentration (mg/kg)
Sample Depth (feet)
0
20
40
60
80
100
120
140
160
180
200


Zinc
Zinc Concentration (mg/kg)
Sample Depth (feet)
0
20
40
60
80
100
120
140
160
180
200