Lab 4: Groundwater Arsenic

ENS-215

Winter 2022

Overview

The vast majority (>96%) of available freshwater on Earth exists as groundwater and as such it plays an important role in both environmental/ecological processes as well an important societal role[1]. Groundwater serves as a critical water source for much of humanity – with more than 50% of the global population relying on it for their drinking water. Furthermore, groundwater accounts for in excess of 40% of the water used for irrigation worldwide [2]. While groundwater can have many advantages over surface water (e.g. less likely to be affected by microbial contamination or human/industrial pollution) there are nonetheless water quality issues that can arise and threaten the health of those who consume it.

In South and Southeast Asia, exposure to naturally occurring arsenic (As) in the groundwater has led to what has been called *"the largest poisoning of a population in history"*, with over 100 million people exposed to groundwater with unsafe levels of arsenic[3,4]. Throughout S-SE Asia, wells often tap groundwater this is significantly in excess of the 10 ug/L level considered acceptable by the World Health Organization (WHO). One particular challenge faced throughout the region is that, while arsenic concentrations are frequently high, they nonetheless exhibit significant variability from well to well and between depths. Thus, one family may have a well that exceeds the WHO guidelines for arsenic, while their neighbor only 200 meters away has a well that is low in arsenic and safe to drink. This variability in space and depth into an aquifer creates challenges when trying to predict where low arsenic groundwater will occur.

While the ability to model/predict exactly where high vs. low arsenic will occur is still somewhat limited, the processes responsible for the release of arsenic into the groundwater are broadly understood. Arsenic is naturally find in the sediments of S-SE Asia (and generally on sediments throughout the world) and can be released into groundwater under the right set of geochemical conditions. Arsenic is generally associated with iron-oxide minerals and when microbes within an aquifer consume organic matter (*i.e.* their food) they can deplete oxygen. Once oxygen and other oxidizing agents (*e.g.* nitrate, sulfate, . . .) are depleted they can switch to "*breathing*" iron oxides. This dissolves the solid iron minerals and releases any associated arsenic into the groundwater.

In today's lab we will examine a dataset from the British Geological Survey (BGS) that was collected in the early 2000's, soon after the issue of widespread arsenic contamination came to light[5]. The BGS along with the government of Bangladesh, which is the most heavily affected country, collected and analyzed the chemistry of groundwater samples from several thousand wells throughout Bangladesh. This dataset was the first large dataset available to begin developing an understanding of the scope and scientific issues associated with the problem.

You will conduct some exploratory data analysis using the BGS data to help develop a better understanding of the variability in arsenic concentration and the factors associated with high levels of groundwater arsenic.

---

[1]USGS
[2]UN Water
[3]Fendorf et al. (2011)
[4]Smith et al. (2000)
[5]BGS Bangladesh data and reports

Background reading

Before writing up your report, you should read the following sources. These readings will give you helpful background info and context and will help guide your analysis. In addition to these sources, you may find it helpful to search out other sources to help you make sense of your analysis.

- Fendorf *et al.* (2010)

In addition to the article above you will also find the BGS report on arsenic in Bangladesh helpful. However, this report is hundreds of pages and while you may be interested in flipping through it, the following summary sections are only 1-2 pages each and will give you a nice overview of their report.

- BGS Report: Executive Summary
- BGS Report: Introduction

Below is a map of Bangladesh showing the regions (Divisions). This map may be helpful when interpreting your data.

Figure 1: **Map of Bangladesh** (source: Wikipedia)

Objectives

For today's lab, imagine that you are part of the initial reconnaissance team from the British Geological Survey and you've just finished the field and lab work and are now sitting down to examine the data. Your primary goals include:

- Characterizing the extent of As contamination
- Examining the regional and depth patterns of As contamination
- Exploring relationships between arsenic and other geochemical parameters, with an eye towards understanding the conditions associated with arsenic release to groundwater
- Synthesizing the information from your study to help make recommendations to policy makers regarding mitigating arsenic exposure

1. Examine the data

First you will need to download the data. The data can be downloaded from our class site using the following code.

A description of the variables in the dataset can be found in the readme file

```
library(tidyverse)

bangladesh_gw <- read_csv("https://stahlm.github.io/ENS_215/Data/NationalSurveyData_DPHE_BGS_LabData.csv
```

Take a look at the data frame and make sure everything looks good and that you understand the variables contained in the dataset.

2. Create additional variables

In this lab are going to examine how arsenic concentrations (as well as other geochemical parameters) vary with depth into the aquifer. Let's create a categorical variable that classifies each observation based on the depth fo the well the sample was collected from. This categorial variable will allow use to easily group our data into depth classes.

Categorical variable: depth Add a new variable `depth_cat` to your data frame. This variable will take the values *shallow*, *intermed*, or *deep* based on the sampled well's depth:

- shallow $\leq$ 50 m
- 50 m < intermed $\leq$ 100 m
- deep > 100 m

Once you've created `depth_cat` convert it to a **factor**. Make sure the factor is **ordered** with the following order:

- shallow $\rightarrow$ intermed $\rightarrow$ deep.

Categorical variable: Arsenic Also add a categorical variable `As_cat`. This variable will take the values *low As*, *med As*, or *high As* based on the sampled well's depth:

- low As $\leq 10 ug/L$
- $10 ug/L <$ med As $\leq 50 ug/L$
- high As $> 50 ug/L$

Once you've created `As_cat` convert it to a **factor**. Make sure the factor is **ordered** with the following order:

- low $\rightarrow$ med $\rightarrow$ high

3. Create summary tables

Now let's create summary tables that distill the information in the large dataset down to more easily digestable tables. These types of summary tables are great to make when generating papers/reports for others, where they won't necessarily have the time or expertise to go through all of the data. These tables are also a helpful tool for the person investigating the data (*i.e.* you) when trying to get a handle on a large dataset.

Summary tables: geochemical overview You plan on meeting with some fellow environmental scientists to discuss the scientific questions around arsenic contamination of groundwater in Bangladesh. Before you meet you want to have a few tables that provide a basic overview of the groundwater geochemistry.

1. Using all of the data for Bangladesh, create a summary table that has the mean and median for the following variables:
    i. Well depth, well contruction year, arsenic, iron, manganese, and sulfate

Note: Be smart about this and look at your `dplyr` cheat sheet. There is a very efficient way to do this.

2. Create the a single summary table that has the exact same statistics and variables above, but reported by DIVISION (*regions of Bangladesh*).

- Then sort the table in descending order by mean arsenic concentration.

Summary tables: risk/exposure assessment The summary geochemical summary tables above are great for discussions among fellow scientists however, we it is particularly important to convey information about arsenic risk exposure to government official and policy makers who may not have an understanding of groundwater chemistry. Let's create some tables that help to assess the exposure risk in Bangladesh.

1. Create a summary table describing the distribution of low, medium, and high arsenic wells by division (regions). In terms of the content contained in your table it should look like the example below (*with the rows and data filled in of course*).

Division	number of wells	% low As	% medium As	% high As
data here	*data here*	*data here*	*data here*	*data here*

- You should sort this table in a fashion that is sensible and makes it easy to read/interpret.

2. Create a summary table describing the distribution of low, medium, and high arsenic wells by depth category. In terms of the content contained in your table it should look like the example below (*with the rows and data filled in of course*).

Depth category	number of wells	% low As	% medium As	% high As
data here	*data here*	*data here*	*data here*	*data here*

- Think about what these tables imply with respect to supplying safe, low-As groundwater. Are there depths and regions that are particularly high risk or particularly low rish? You should discussion your observations/interpretation in your lab discussion.

4. Examine parameters related to groundwater arsenic levels.

From your background reading you have learned that arsenic mobilization (release) into groundwater is favored under anoxic geochemical conditions (*i.e.* no dissolved oxygen in the water). Once oxygen is depleted in the water, microbes within the aquifer will begin to "*breathe*" other oxidized substances such as dissolved sulfate (SO_4) iron-oxide minerals and manganese-oxide minerals. Typically microbes will favor using dissolved sulfate and manganese-oxides over iron-oxides – however, once these are low in supply they will switch over to "*breathing*" iron-oxides. When these iron-oxides are utilized by microbes, they dissolve releasing iron and and associated arsenic into the groundwater.

By looking at the concentrations of As, Fe, Mn, and SO_4 we can learn about the geochemical conditions within the groundwater and try to tease out the conditions that are connected to high levels of arsenic in the groundwater. In particular, we can try to understand the precise conditions/thresholds that lead to high or low arsenic, which can help to improve our understanding of why arsenic concentrations vary so much with space and depth. Furthermore, it is important to examine arsenic concentrations in relation other geochemical parameters (e.g. Fe, Mn, . . .) to corroborate our hypotheses regarding the mechanisms responsible for arsenic mobilization.

Let's start to examine the relationships between arsenic and other parameters (both geochemical and non-geochemical).

Depth profiles

As vs. depth   We made a table earlier that provided some summary information regarding arsenic levels versus depth into the aquifer. Now let's take a more detailed look by examining the data points themselves (_i.e. not summary statistics)

1. Create a scatter plot(s) of arsenic vs. depth, where depth is on the y-axis. It is helpful to reverse the y-axis so that shallow depths are shown near the top of the figure and deeper ones near the bottom (Hint: use `scale_y_reverse()` to do this).
    i. Focus on making a nice, easy to interpret figure. You may want to play around with additional aesthetic mappings or adjust things such as the symbol size, shapes, alphas,...
    ii. You may want to create multiple versions of this figure, with the different versions providing a different spin on the visualization. In particular, it may be helpful to "zoom in" by adjusting your y-axis to focus on a certain depth interval

You may want to create "multiples" of these figures by DIVISION. You can use the `facet_wrap` function to easily do this.

- Does arsenic display any interesting depth patterns?

- Think about depths that might be ideal and not-ideal for installing safe wells

Fe, Mn, SO$_4$ vs. depth   Create similar plots to the one you made above for As, however this time create versions for Fe, Mn, and SO$_4$

Think about how the depth profiles of these parameters compare to arsenic. For instance, does high As coincide with high values for some of these parameter? Does high As coincide with low values for some of these parameters? Look back to the paper by Fendorf *e.t. al* (2010) and/or discuss with me if you have any questions.

- Does these parameters display any interesting depth patterns?

- Are their depth patterns similiar? Are they similar to the arsenic depth pattern?
- Do these depth patterns provide any insight into the processes/factors driving As mobilization?

Relationships between parameters   Now that you've investigated how the geochemistry varies with depth, let's take a look at how arsenic from a given water sample relates to other geochemical parameters in that same sample. This can help us understand the relationship between the different chemicals and make inferences into the processes driving arsenic release to the groundwater.

As vs. Fe and As vs. Mn scatter plots   One of the primary mechanisms for arsenic release is the dissolution of arsenic-bearing, iron-oxides which release both As and Fe into the groundwater. However, the relationship between dissolved Fe and As can be complex since, both of these elements can later become associated with solid-phases.

Let's examine the relationship between these parameters by creating a scatter plot of As and Fe.

You should make this a nice and easy to read figure. Think about potentially mapping other variables to the aesthetics to add more information to your figure.

- How do As and Fe relate to one another?
- Does the relationship conform with your expectations, based on what you've learned about As mobilization?

Now, create a similar figure for As and Mn and examine/assess the relationship between these two elements.

As vs. SO$_4$ scatter plot   High levels of SO$_4$ can help to inhibit microbial reduction of iron-oxides, thereby limiting the release of As to the groundwater. Let's examine the relationship between these parameters by creating a scatter plot of the two elements.

You should make this a nice and easy to read figure. Think about potentially mapping other variables to the aesthetics to add more information to your figure.

- How do As and SO4 relate to one another?
- Does the relationship conform with your expectations, based on what you've learned about As mobilization?

Additional Analysis

For your additional analysis you should continue exploring the dataset. You may decide to expand upon the work above and/or develop completely new lines of inquiry. When conducting your additional analysis, think back to your overall goals that were outlined in the **Objectives** section. Some ideas for additonal analysis include, but are not limited to:

- Exploring differences in arsenic levels between the regions (DIVISIONS) of Bangladesh. Among other things, this might involve box and whisker plots or other plots helping to show the distribution of arsenic.

- Exploring differences in arsenic levels between depth categories.
- Examining the relationship between As and other parameters (e.g. Ca, Sr, K, . . .)
- Creating additional categorical variables (e.g. low, med, high Fe) which may be helpful in grouping or color coding data points in your graphics.
- Examining differences in other geochemical (other than As) between regions of Bangladesh.

- Examining some of the attributes of wells in Bangladesh to get a better idea of construction patterns (e.g. ages of the well, depth distribution of wells) across the country and by region.
- Create additional summary tables that help to illuminate important features of the data.

You are free to pursue these and any other ideas that you may think of.

Prepare your final lab report

Advice/guidance   Remember that your lab should be a nicely formatted and organized report that includes your code, output, and discussion. You should have an introduction and conclusion section and you should clearly delineate different sections with headers. Also remember to cite any sources that your rely/reference.

At this point in the term you should start to think about the organizational logic of your final lab report (i.e. order your analysis sections and discussion in a sensible manner). You don't need to get bogged down on this point when working in lab, but it is good to consider when you are putting together your final report for submission.

The "additional analysis" section of your lab is a critical component and should be given careful thought and effort.

It is important to realize that my expectations for each of the graded criteria increase as the term progresses and you gain more skills and are better able to develop your own analyses.

Make sure you are familiar with the expectations and evaluation criteria presented in the links below:

- Evaluation rubric
- Template for lab and homework assignments: (html) (Rmd)
- Lab report guidelines

Submitting the lab   Your lab is due prior to the start of next week's lab. Once you are finished and satisfied with your work you should Knit it to an `.html` file and submit both your `html` and `Rmd` file to Nexus.

Please make sure that the `.html` file you submit is `.html` and not `nb.html`.