Case Study

```
## -- Attaching packages ---------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

Background

Coal is one of the most prevalent combustible fuels being burned all across the world, as it is one of the easiest methods of obtaining energy due to the abundance of the substance. Generally, coal plants produce electricity by burning coal, which produces coal ash as a byproduct. Over 100 million tons of coal ash are produced every year at these plants, which are then disposed through landfills and waste ponds at these plants. The main concern of ecologists regarding this matter is that the coal ash produced by these plants can often contaminate the local groundwater, leading to toxic contaminants being found in local water sources. Coal ash is dangerous due to its composition, which contains a long list of dangerous chemicals including – but not limited to: arsenic, radium, boron, and a large list of other contaminants toxic to humans and animals alike [@Kelderman2019].

Complaints and concerns regarding the disposing practices of coal plants have only increased after the 2010 Kingston Fossil Plant coal ash incident in Tennessee. This area has become an attractive location in which many sites of ecological studies have been conducted the years following the incident. Leaching experiments conducted by @Ruhl2010 has revealed significant levels of dissolved Arsenic, Boron, Strontium, and Barium in the water which has been in contact with the coal ash, which they note to be threat to infaunal species in the water. Prompted by environmental organizations, groups, and individuals alike, an onslaught of pressure was put on the Environmental Protection Agency, which resulted in the Coal Ash Rule being put into effect in 2015 [@Kelderman2019].

This rule has forced over 265 coal power plants – about 3/4 of all coal power plants in the US - to make data regarding chemical concentrations publicly available to the general population. In their analysis using this data, the @EIP2020 – a non-profit organization dedicated to issues involving environmental justice, has discussed the prevalence of groundwater contamination for wells located near coal related facilities.

Typically in a coal ash plant, there exists two types of wells: upgradient wells and downgradient wells. These wells are essential to measure the amount of contamination being caused by coal ash. Upgradient wells, also known as background wells, measures the concentrations of chemicals in groundwater before it passes through an coal ash dump. Conversely, downgradient wells measure the concentrations of chemicals in groundwater after it passes through a coal ash dump.

While both types of well are susceptible to contamination through coal ash related means, it is more frequently the case that we focus on the downgradient wells, as they are more closely linked to the water accessed by the general public.

The goal of the study conducted by @Kelderman2019 was to identify the percentage of coal plants which have unsafe levels of contamination. Determining whether if a well was contaminated or not was largely influenced by the average mean concentrations of the contaminant in question for that well. If the average

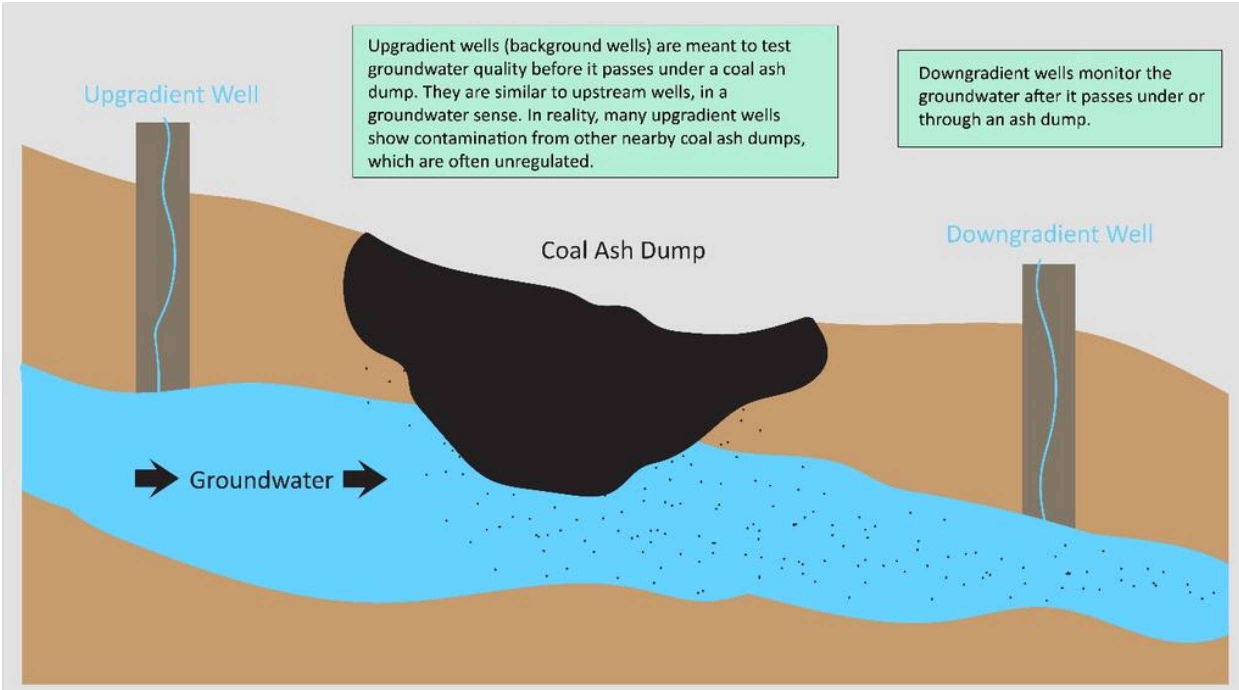

Figure 1: Difference Between Upgradient and Downgradient Wells

mean contamination of a contaminant, say, arsenic, was above the health-based threshold considered necessary to ensure safety, then that well would be marked as an "exceedance," and deemed to be contaminated.

@Kelderman2019 notes the possibility of contamination being caused by an external factor, unrelated to the coal ash, and provides a stipulation as how to account for this. They excluded tallying the wells in which the mean downgradient values were lower than the mean upgradient values, as this would mean that the contamination was not caused by the coal plant itself.

Our end goal remains the same as the @Kelderman2019: to identify contaminated groundwater in coal plants. In their report, it is not mentioned if left censored values were being accounted for in the calculations of the average concentrations of the downgradient and upgradient wells. As such, it may be prudent to check to see if the proportion of wells in the U.S. in which contamination is present would be altered if we used our techniques to calculate the average concentrations of the contaminants.

The limit of detection problem stems from the measuring devices' inability to obtain chemical concentrations smaller than a certain threshold amount, thus affecting the measurements recorded.

Data

It is important to note where the data used originated from, before we delve into the details of our case study. As such, a brief history regarding the Coal Ash Rule and its origin will be explored, alongside details regarding our coal ash dataset.

Coal Ash Rule

A large coal ash spill at the Tennessee Valley Authority (TVA) which occurred on December 22, 2008 in Kingston, TN – prompted the Environmental Protection Agency (EPA) to propose a set of standardized regulations and procedures to address the concerns regarding coal ash plants nationwide in the US. This was known as the Coal Ash Rule, which passed legislation on December 19, 2014 [@Car2020].

Changes were made to the Coal Ash Rule over the years in the form of 'amendments,' in which the most relevant to this study, required coal facilities to publish information and data regarding the concentrations of contaminants in the wells.

Source of Data

The data used in the study are from the results published in "Annual Groundwater Monitoring and Corrective Action Reports," which were made available to the public in March 2018 as a result of the Coal Ash Rule. These reports are in PDF format and are thousands of pages long, which makes it difficult for individuals to parse through data in a meaningful way.

The @EIP2020 obtained the data from an online, publicly available database containing groundwater monitoring results from the first "Annual Groundwater Monitoring and Corrective Action Reports" in 2018 which was collected from coal plants and coal ash dumps under the Coal Ash Rule.

They wrangled the data into a more accessible machine-readable format which contains information from over 443 annual groundwater monitoring reports posted by 265 coal ash plants, downloadable from the EIP's website, which we use in our case study.

Table 1: Data dictionary for the coal dataset.

Variable	Variable Name	Description
State	state	The state where the site is located.
Site	site	The name of the site as it is presented in its groundwater monitoring report.
Disposal Area	disposal.area	The name of the disposal area(s) as they are presented in the groundwater
Type of Well	type	The type of disposal unit. SI = surface impoundment, L= landfill, M = mixed
ID of Well	well.id	The identifier given to each monitoring well in the groundwater monitoring rep
Gradient Type	gradient	The location of the groundwater monitoring well relative to the regulated ash d
Sample Date	samp.date	The date the well was sampled.
Contaminant Name	contaminant	The contaminant name. These have been standardized to allow for analyses acr
Measurement Unit	measurement.unit	The concentration units. These include mg/l, ug/l, pCi/l, and standard units (S
Below Detection	below.detection	LOD status for the concentration, '<' Indicates that the concentration was belo
Concentration	concentration	Concentration of contaminant

Variables

Specifics regarding the variables in the coal dataset can be viewed in Table 3.1. Each observation in the dataset represents a well which measures the concentration of the contaminant in question. Most of the variables are explanatory, such as the state, site, and disposal area in which the well is located in. However, there are several variables specific to groundwater data collection which are important to note.

There are four different types that each well can be classified as, which is represented in the `type` variable. These consist of: `L`, `M`, `SI`, and `U` which stand for "landfill," "mixed," "surface-impacted," and "uniform

The coal dataset contains information regarding chemical concentrations at coal plants. A coal plant consists of multiple disposal areas for the coal ash that it produces. At each disposal area, there are specific locations that groundwater is being measured, known as wells, which represent an observation in the dataset.

Plan of Action

The investigation conducted by @Kelderman2019 mentions certain restrictions within the data that we believe may have caused their analysis to potentially be inaccurate. Specifically with the limit of detection problem arises when measuring devices used to measure chemical concentrations are unable to detect below a certain threshold, causing large numbers of observations to be considered "below detection." These values are often encoded as NA or even mistakenly marked as 0. There are no mentions of any attempts to account for or handle these missing values, which is the motivation for this case study.

Our investigation works with methods on handling this missing data through the techniques detailed in chapter 1 in order to obtain more accurate and precise estimates of the mean concentrations of contaminants at coal ash sites. Through our newly obtained estimates, we hope to see if they might lead us to different beliefs than the claims made by @Kelderman2019 regarding the most contaminated wells in the U.S.

Application