Introduction

The introduction should provide an overview of the work you set out to do and provide structure for the remainder of the document.

Background

- (coal ash report - car) coal one of the most dangerous combustible fossil fuels is comprised of a long list of dangerous chemicals – including substances such as arsenic, radium, other carcinogens, metals that can impair developing children's brains, toxins dangerous to aquatic life, etc [@Kelderman2019]
- power plants produce 100mil tons of coal ash every year, which is dumped into landfills and waste ponds [@Kelderman2019]
- only recently (2015) have complaints and lawsuit arisen in which certain ecological organizations have attempted to sue the EPA to regulate disposal of coal ash [@Kelderman2019]
- this coal ash rule has forced power companies to make publicly available data regarding chemical concentrations in 265 coal plants containing ponds and landfills (about 3/4 of all coal power plants across the US) [@Kelderman2019]
- environmental agencies have concluded that the groundwater under basically all coal plants are contaminated [@Kelderman2019]
- HOWEVER this might be overstated? we wanted to investigate whether or not if this was true.

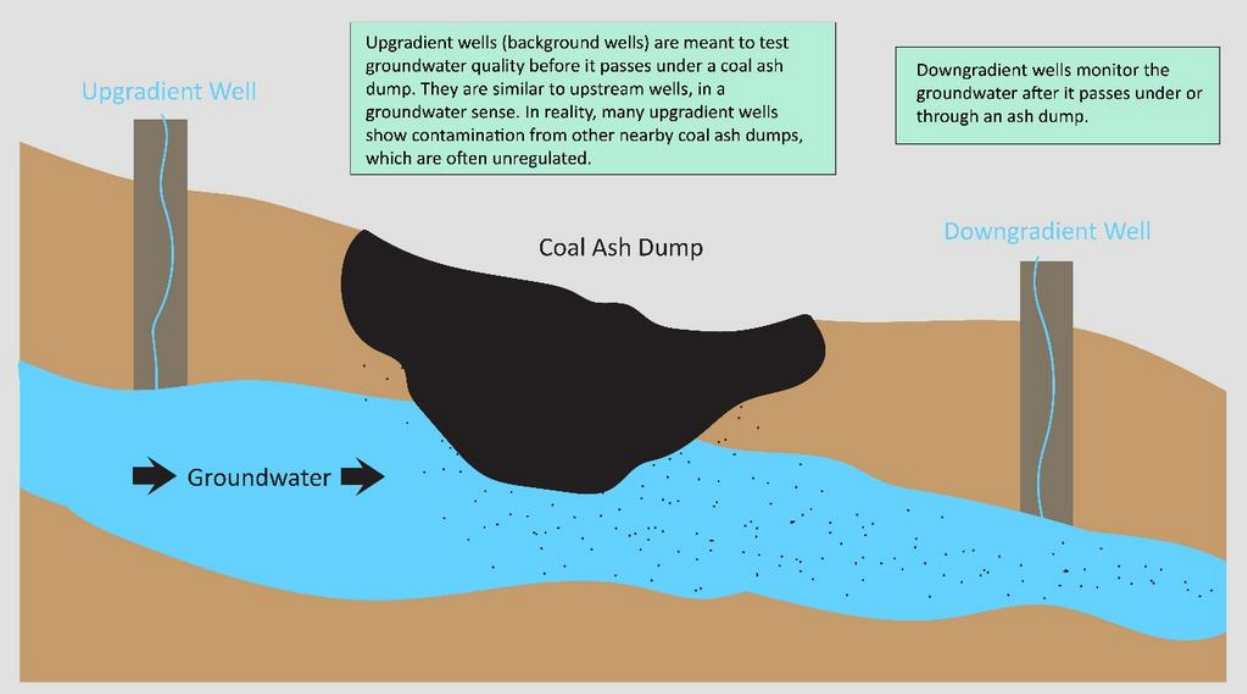

Figure 1: Difference Between Upgradient and Downgradient Wells

- upgradient wells (background wells) measures groundwater chemical levels BEFORE passing through a coal ash dump while downgradient wells monitor the groundwater AFTER it passes through an ash dump

- we have reason to believe that many chemicals are NATURALLY OCCURING and as such, the statement made my environmental agencies regarding all groundwater being contaminated may be overstated
- typically, we would estimate the amount of chemical contamination, for this example – arsenic – caused by a coal ash dump with the equation: downgradient arsenic concentration minus upgradient arsenic concentration
- however, because there may be retired/unregulated upgradient wells that are occasionally contaminated already, this might be inaccurate
- END GOAL IS TO CORRECT THE CONTAMINATED VALUES
- we have already conducted a preliminary investigation using a variety of machine learning techniques to aid us in identifying potential contaminated upgradient wells
- we have also utilized bootstrapping and imputation techniques to correct for their measurements through by accounting for the innate contamination which may be caused by factors such as retired and unregulated wells
- our methodologies have yet to account for another problem however, involving limit of detection problem which arises from the measuring devices' inability to obtain chemical concentrations smaller than a certain threshold amount

Data

Coal Ash Rule

- A large coal ash spill at the Tennessee Valley Authority (TVA) which occured 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 [@Car2020]
- This was known as the Coal Ash Rule, passed on December 19, 2014 [@Car2020]
- Changes were made to the Coal Ash Rule over the years in the form of 'amendments,' one of which made required facility information and data to be made publicaly available to the public (April 15, 2015 rule change) [@Car2020]

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 [@EIP2020]
- these reports are in PDF format and are thousands of pages long, which makes it difficult for individuals to look through the data in a meaningful way [@EIP2020]
- the EIP wranged 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 [@EIP2020]
- they 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 [@EIP2020]

Variables

- a coal ash site consists of multiple disposal areas
- within these disposal areas lie multiple wells
- each observation represents a well
- wells are split into 2 different types - upgradient and downgradient wells
- variables consist of information regard chemical contaminant concentrations and specifics regarding the well
- from the 19 different contaminants (antimony, arsenic, boron, etc. ...) a major problem is that some wells only have measurements for certain chemicals and don't have them for others
- we are currently using information from plants within illinois but there is data for all the states in the US