Residential Water Quality Problem Statement

Background

Human life is impossible without water, covering roughly 70% of the Earth's surface and accounting for 60% of an adult's body by weight. Clean drinking water is necessary for survival, but access to fresh water is also essential for irrigation of crops, basic hygiene, and medical care. The World Health Organization (WHO) reports that with access to safe water, children demonstrate much better health outcomes, enabling them to focus on education and achieve more in life. Water can be thought of as not just essential for human health, but for the health of a society as a whole.

Various global initiatives such as the Millennial Development Goals (MDG) led by UNICEF and the WHO have been implemented to improve water access and quality. These efforts have resulted in a dramatic increase in the percentage of the global population that has access to "improved water sources", which keep water supplies free from contamination, from 76% in 1990 to 91% in 2015. However, hundreds of millions still live without access to clean water, especially in rural areas.

While the populations with the least access to good water sources are largely clustered in developing countries, recent events such as the Flint water crisis in Michigan and the California drought remind us that even in economically advanced countries such as the United States, careless public planning and negligent environmental regulations can still threaten the public's access to clean and safe public water supplies. In Flint, scientific reports indicate that the inadequate addition of the appropriate corrosion chemicals to the water sources, led to increased lead exposure of many residents and particularly doubling that in children. In this study, they outlined that children have higher water-soluble lead absorption than that of adults, posing a significant threat to their development.

Industrial and technological revolutions which have resulted in advanced manufacturing processes, have slowly resulted in water pollution of nearby areas from the contaminants and by-products released into the natural water sources. As we proceed with the development of novel technologies, it is imperative that we continue to improve the quality of life and access to safe drinking water. The effects of pollution in water sources will trickle down and potentially lead to unprecedented changes of the local/global environment, ecology of species, and human development.

Your Task

Your goal is to analyze the chemicals, droughts, water usage and industry occupational data (described in detail below), potentially in combination with supplementary datasets, in order to increase the understanding of how various factors such as natural events and governmental

initiatives have influenced the environment through time, specifically pertaining to the quality of water in residential areas.

We have pre-cleaned several supplementary datasets for your use. Additional data is available, including details about educational attainment of the population and earnings by industry.

You are asked to pose your own question and answer it using the available datasets in the available time. What is important is the insightfulness and depth of your conclusions and analysis. **You need not be comprehensive; quality data analysis will be rewarded over breadth of the question posed.**

Submissions may be predictive, using machine learning and/or time series analysis to predict or model water supply trends. Submissions may also be illuminating, through use of thoughtfully chosen data visualizations or sound statistical tests.

Consider exploring one of the sample questions below, or creating your own variation. Creativity in formulating your own question generally has a positive effect on judges' assessment of your submission; **however, it should not be at the expense of analytical depth, precision, and rigor, which are far more important.**

Sample Question 1: How do water quality measures correlate with quality of life measures/SES factors?

Sample Question 2: What counties are most vulnerable in the event of a drought? Do droughts have an effect on industry specific earnings?

Sample Question 3: Does a relationship exist between the major type of industries (i.e; manufacturing) and the quality of water? Are there greater concentrations of potentially hazardous contaminants in areas of more industrial manufacturing?

Datasets

The provided datasets are are spread across six tables. Your team should only use the tables that are relevant to your chosen question/topic. The raw data sources are noted; however, we encourage you to use our tables since they have been organized and cleaned to "play nice" with each other.

chemicals
Data containing the measured mean concentration of a particular chemical or substance in community water systems throughout the counties in the United States from 2000 – 2016.
~882,000 rows & 12 columns. Size: ~100MB. Source: Centers for Disease Control and Prevention.

droughts

Data containing the particular percentage of various range of drought severities, indexed by counties for particular start-end periods throughout the United States.
~1.35 million rows & 11 columns. Size: ~100MB. Source: U.S. Drought Monitor.

earnings
Information about the industry specific median earnings (in that specific year's USD, inflation adjusted) indexed by counties for all of the United States, taken from 2010 – 2016.
21,999 rows & 31 columns. Size: ~5MB. Source: U.S. Census.

educational_attainment
Data containing the educational attainment of the US population by county from 1970 – 2000 and estimates of 2012 – 2016.
16,416 rows & 12 columns. Size: ~2MB. Source: U.S. Department of Agriculture.

Industry_occupation
Data containing the estimated working population (16 years and over) for the various industries indexed by counties, taken from 2010 – 2016.
5,712 rows & 18 columns. Size: ~0.7MB. Source: U.S. Census.

water_usage
Information about particular water usage (irrigation, public supply, crop, etc.) and thermoelectric power generated for counties that were found for the year 2010.
3,225 rows & 117 columns. Size: ~2MB. Source: U.S. Department of the Interior.