PROJECT PLAN: KANSAS WATER QUALITY AND CONSERVATION PROGRAM SPENDING

ABSTRACT

This project seeks to model the relationship between the Kansas Department of Agriculture (KDA) cost-share programs for conservation and data from the Kansas Department of Health and Environment (KDHE) on surface water quality. As a second step in analysis, the project looks to model the relationship between characteristics that appear in upstream water sources and the characteristics that appear in municipal drinking water. Finally, the project seeks to understand whether the conservation cost-share programs have unintended consequences (positive or negative) for municipal drinking water.

1 INTRODUCTION

The goal of this project is to evaluate the relationship between the conservation program spending managed by the Kansas Department of Agriculture (KDA) and surface water quality which is monitored by the Kansas Department of Health and Environment. An additional goal of the work is to understand whether there is a relationship between surface water quality and municipal water quality, and whether there are downstream effects of conservation program spending on municipal water quality.

The water quality portal (https://www.waterqualitydata.us/) "is the premiere source of discrete water-quality data in the United States and beyond." The data about surface water quality testing in the Water Quality Portal is submitted to the EPA by KDHE. The Kansas Data Science Consortium (KDSC) has also received data on surface water quality testing directly from KDHE that supplements what is housed at the Water Quality Portal. Broadly, the information on surface water quality includes tested characteristics (e.g. Barium, Strontium, E-coli) from point-based data (latitude/longitude) which KDSC analysts have processed to include geographic identifiers such as hydrologic unit code (HUC), county, and census tract.

This project seeks an economic model to establish the cost municipalities face in cleaning their water. Municipal water sources come from surface water and from groundwater. They must treat the drinking water to ensure that water's safety for human consumption. Given that the sources of water for different municipalities may come with varying levels of characteristics in the water, municipalities face different costs to ensure safety. To better understand these varying costs, it is important to model how levels of characteristics in water sources relate to levels of those same characteristics that appear on the CCRs for municipal drinking water.

2 Baseline or Initial Analysis

The key data for this project comes from multiple sources, including the Cost-Share Information Management System (CSIMS) at the Kansas Department of Agriculture, the Consumer Confidence Reports submitted by municipal water systems and maintained by the Kansas Department of Health and Environment, as well as data from the Kansas Department of Health and Environment surface water quality monitoring program.

An important step for this project to move forward is to match the latitude and longitude coordinates for tests recorded in the Water Quality Portal to recognized geographies such as HUC codes, county identifiers, and census tract codes. Additionally, work will need to be undertaken to match the KDA CSIMS data on conservation spending to surface water quality tests at a common geography. KDA CSIMS data include information about the county, the river basic, and HUC14 codes.

The Consumer Confidence Reports (CCRs) have been collected and processed from Kansas Department of Health and Environment's website where they are stored as word documents. This data must be linked with the water quality data from the Water Quality Portal. This must be done with an understanding of water flows. To connect the data in the CCRs with the data in the Water Quality Portal, the analyst will need to consider how water from surface-level water sources (lakes, streams, and rivers) arrives in municipal water sources. The analyst will also need to consider the share of municipal water each city obtains from surface water and from groundwater.

3 Final Analysis

The final analysis will use statistical methods like regression to evaluate the impact of conservation spending on surface water quality. In particular, use the project type and the practice, variables included in the KDA CSIMS data, to think about the kinds of conservation projects that might have an impact on water quality. There are 61 unique values in the project type variable. The analyst working on this project should consider what kind of similarities the different project types may have and which may impact surface water quality through what types of channels. There are 108 unique values in the practice variable. Similar to the project type, the analyst should draw links between the practice used and specific water quality characteristics.

When conducting the analysis to examine whether certain project types and practices have an impact on surface water quality characteristics, it is important to consider the amount of money attached to the project, which can be found in the computed cost and amount requested variables.

Additionally, the final analysis for this project will include a set of regression estimates that identify the relationship between water quality characteristics in surface-level water and those same characteristics in municipal water sources. These regression estimates should be put into tables following the standard format for displaying regression estimates. Include robustness checks, heterogeneity tests, and visualizations that investigate the validity of these estimates.

4 Final Goals & Evaluation

The final goal of this project is a report and slide deck that summarizes the data work, the exploratory analysis, and the final statistical analysis to establish whether a relationship exists between trade association density and NTM density by country and industry. The analysis should be presented at the Kansas Data Science Consortium conference on May 2nd, in

Manhattan, KS. The project report and accompanying script files and should be sent to the project sponsors at the completion of the semester.

5 Related Work

Read, E. K., L. Carr, L. De Cicco, H. A. Dugan, P. C. Hanson, J. A. Hart, J. Kreft, J. S. Read, and L. A. Winslow (2017), Water quality data for national-scale aquatic research: The Water Quality Portal, Water Resour. Res., 53, 1735–1745, doi:10.1002/2016WR019993.

Josset, L., Allaire, M., Hayek, C., Rising, J., Thomas, C., and Lall, U. (2019). The U.S. water data gap—A survey of state-level water data platforms to inform the development of a national water portal. Earth's Future. 7, 433–449. https://doi.org/10.1029/2018EF001063

Zheng, J. G., Wang, P., Patton, E. W., Lebo, T., and McGuinness, J. S. L. D. L. (2011). A Semantically-Enabled Provenance-Aware Water Quality Portal. In Environmental Information Management Conference.

Environmental Protection Agency. Drinking Water Treatment Technology Unit Cost Models. Last Updated on May 5, 2023. Accessed on February 8th, 2024.

Kansas Department of Health and Environment. Consumer Confidence Reports. https://www.kdhe.ks.gov/531/Consumer-Confidence-Reports. Accessed on February 8th, 2024.

6 Data & Technical Requirements

Analysis should be done in Stata, R, or Python and code should be stored in script/do files. The analysis should be "full stack" in that the script files load in a working data set and produces a set of output such as figures and tables. Each do file should accomplish a defined and contained task (e.g. data construction or table 1 creation, etc).

The data for this project will come from three sources. The water quality data from upstream sources comes from the Water Quality Portal run by the Environmental Protection Agency (EPA). The water quality data for municipal drinking water comes from the Consumer Confidence Reports (CCRs) that are available on the Kansas Department of Health and Environment's (KDHE) website. It will also be necessary to find information about how upstream water flows into city water sources.