Ann Fritz
GEOG 777 – Project 1 Report
October 13, 2020

Project Goal

I built an interface that allows a user to enter different values of K to explore the relationship between nitrate concentration in groundwater to rates of cancer occurrence in Wisconsin. The Wisconsin Department of Natural Resources Environmental Division or Wisconsin Department of Health Services may need to take action if a definitive relationship between nitrate levels in groundwater and rates of cancer occurrence are found.

Background

High concentrations of nitrate in drinking water pose a risk to human health, especially infants. The U.S. Environmental Protection Agency standard for nitrate in drinking water is 10 milligrams of nitrate (measured as nitrogen) per liter (mg/L). Consuming too much nitrate can affect how the blood carries oxygen throughout the body and cause a condition called methemobloginemia, also known as "blue baby syndrome." Recently, however, studies indicate there are also potential health concerns for adults, including a possible increased cancer risk. The magnitude of that risk is not fully known, hence the need for this project.

The Wisconsin Department of Natural Resources (WDNR) has resources that can provide insight into the relationship between nitrate in groundwater and cancer occurrence in Wisconsin. The WDNR has a database of nitrate concentrations in groundwater from wells constructed throughout the state. For the last ten years, the WDNR has also been assembling a database of cancer occurrences throughout the state. I can use these two datasets to explore the relationship between nitrate in groundwater and cancer occurrence.

The Data

1. cancer_county.shp contains rate of cancer occurrences aggregated by county. *Source: WDNR*
2. cancer_tracts.shp contains rate of cancer occurrences aggregated by census tract. Cancer rates are defined as total cancer occurrences divided by the total population of the block group and aggregated by tract. *Source: WDNR and US Census tracts*
3. well_nitrate.shp contains nitrate concentrations at sampled wells. *Source: WDNR*

Implementation and Design

I used ArcGIS Desktop 10.7.1 to develop the initial geoprocessing workflow. Nitrate concentration and cancer rates needed to be in the same geometry to examine the relationship between the two data sets. I used the inverse distance weighted method in the Spatial Analyst toolbox to interpolate nitrate concentrations within the state and then used "zonal statistics as table" to determine an interpolated nitrate concentration for each census tract polygon. I joined the resulting table to the census_tract shapefile to get both cancer rate and nitrate concentrations in the same dataset. Finally, I used ordinary least squares regression in the Spatial Analyst toolset to complete the regression analysis, as well as Moran's I statistic to measure degree of spatial autocorrelation of residuals.

I used ModelBuilder and Python 2.7 to develop initial models that were later exported as .py files. The .py files were combined to run in a standalone python script. Tkinter, a standard graphical user interface (GUI)

package included with Python, was used for to display and to provide user-interaction. The decision to use Tkinter and Python was ease of use as well as forcing myself to relearn long forgotten python .

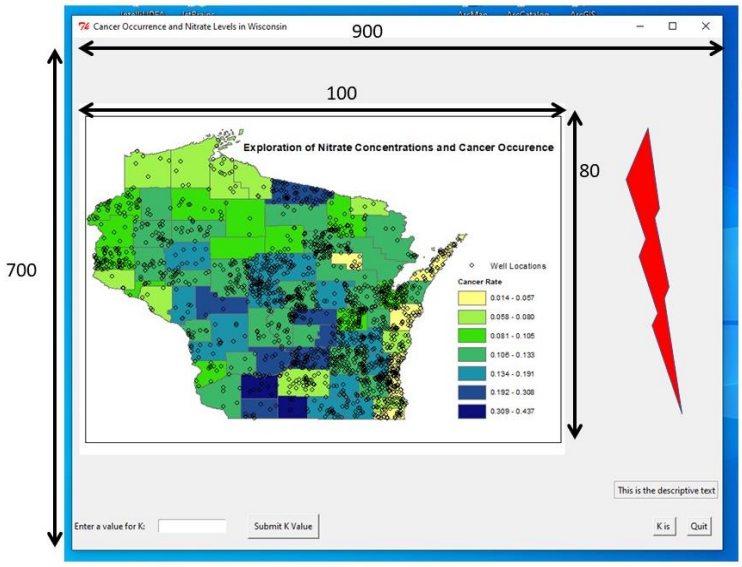

Figure 1. How do I get the text to show up where the red lightning bolt is? A lot of trial and error!

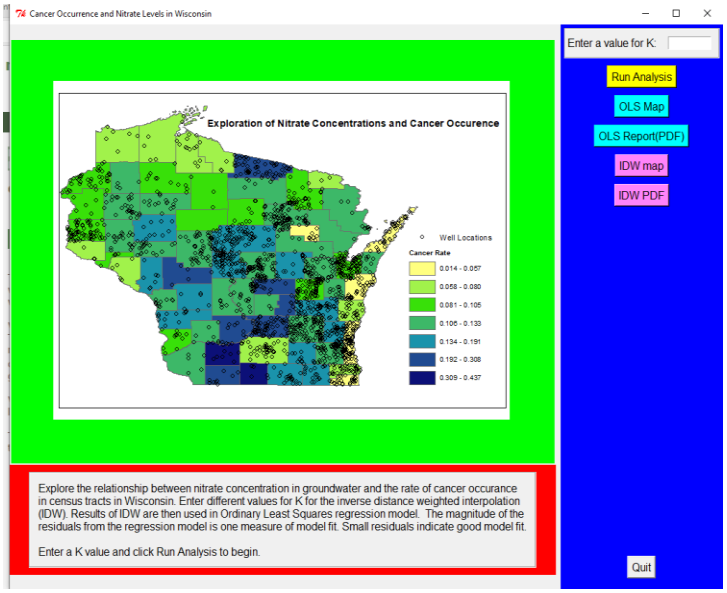

Figure 2. Even though it looks like clown vomit, using background colors, bg="red" or bg="green" for example, was very helpful in the frame design and button placement process of the user interface.

Implementation timeline is shown in Table 1. Creation of the Tkinter GUI was more problematic than I anticipated, especially the layout of the widgets within the frames, examples shown in figure 1 and 2. Much of my problems focused on finding consistent and clear documentation of the Tkinter package and all the options. Not every site claiming to have instructions about Tkinter included all the options available, or they did not indicate which version of Python was being used.

Results

Results using a K value of 2 are shown in figures 3 and 4. Higher concentrations of nitrate are present in central and south central Wisconsin; it should be noted that the interpolated concentrations are above the 10 mg/L nitrate standard in much of south central Wisconsin. There is a positive correlation between nitrate concentrations and cancer occurrences in Wisconsin, especially in south-central and central areas. These data are not normally distributed, and there is significant clustering (figure 5). These results point to further analysis to determine if there are other potential significant variables that could impact nitrate concentrations and cancer occurrences. Other variables to explore could include land-use practices

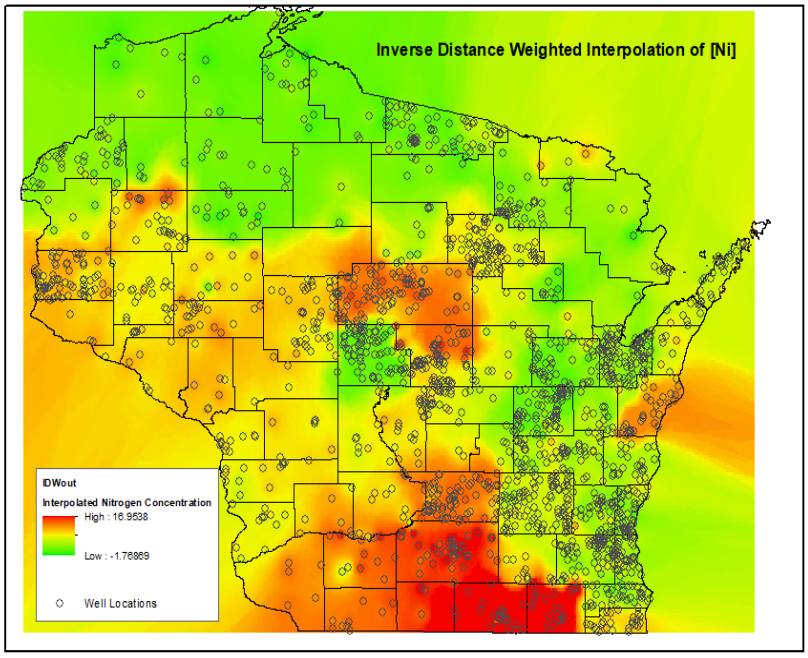

Figure 3.Interpolated nitrate concentrations using IDW, K = 2

that could contribute to higher nitrate concentrations; use of private septic systems in a county (note the four counties in south central in red), what acquifer the water is being drawn from, well construction material or if the well needs maintenance, or if there are specific types of cancer that have a positive correlation with nitrate concentration.

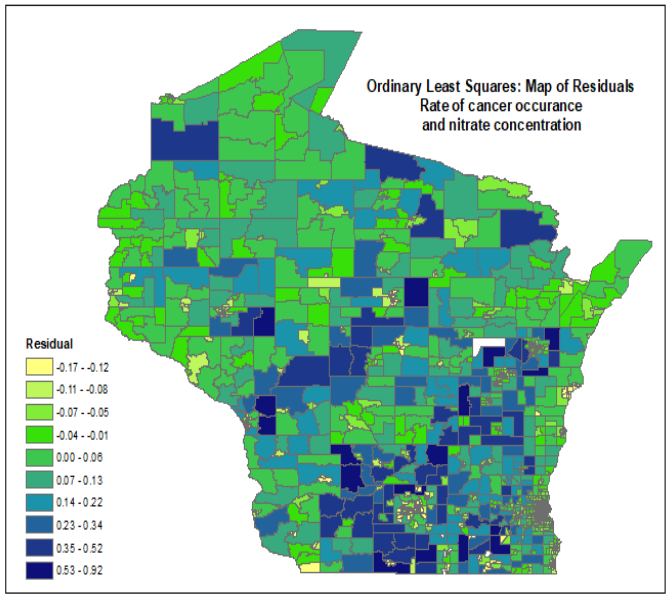

Figure 2. Ordinary least squares map of residuals.

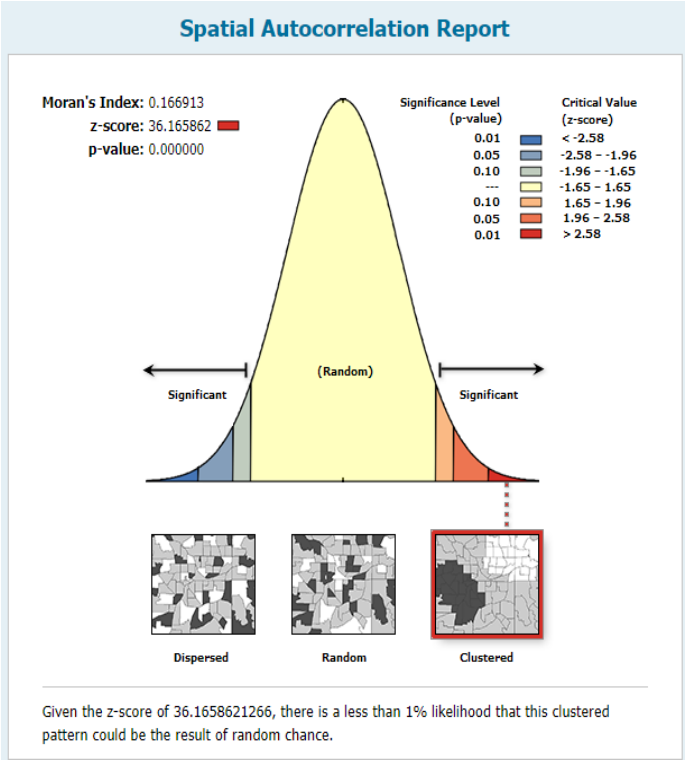

Figure 3. Spatial autocorrelation report on residuals of OLS shapefile indicates significant clustering of the points..

Table 1. PROJECT 1 IMPLEMENTATION PLAN

TASK	Milestone
Geoprocessing Logic	
Using ArcGIS Desktop 10.7.1, use geoprocessing tools to make sure geoprocessing and analysis logic is correct.	12 Sept. 2020
1. Interpolate nitrate levels in wells using inverse distance weighting (IDW from Statistical analysis tools), resulting raster of smooth surface of nitrate levels.	
2. Use "zonal statistics as table" and link the nitrate levels of the interpolated surface to GEOID of census tracts. Census tracts already have cancer occurrence per capita for that tract. This results in a nitrate concentration and cancer occurrence per capita value for each tract.	
3. Use ordinary least squares regression model on cancer rates and nitrate levels at the census tract level to see if there's a relationship between the two variables.	
Create Python script(s)	
4. Use ModelBuilder within ArcGIS Desktop to export tools to standalone Python scripts for geoprocessing tasks identified in steps 1-3.	14 Sept. 2020
5. Use PyCharm to create single script that will call each script as a function	20 Sept. 2020
Test python script functionality	
6. Testing and debugging of python scripts	Ongoing
Create GUI	
7. Use PyCharm and TKinter to create GUI for user-interface to change K and display results.	27 Sept. 2020
Pilot Test	
8. Beta demo complete	29 Sept. 2020
Testing and Debugging	
9. Test GUI	Ongoing until
10. Debug with multiple values for K to ensure error handling procedures are correct.	project completion
Final Project Deliverable	**13 Oct. 2020**