GEOG 777 Project 1 Report
Daniel Fourquet
3/2/2021

Project Scenario

The instructions for Project 1 ask the students to imagine that they work for the Wisconsin Department of Natural Resources (WDNR). WDNR is studying the relationship between the nitrate concentration in drinking water and the risk of cancer. High nitrate concentration in drinking water is already considered a health risk, but the WDNR has recently discovered that it can also pose a cancer risk.

The department collected a database of ground water nitrate levels from well water tests as well as cancer data on the county and tract level. As a GIS Analyst working for the WDNR, I was asked to create a tool that will use this data to determine the relationship between nitrate levels and cancer rates. The department requested an application with a GUI that will allow the user to set the K value for IDW interpolation, run the analysis on the data, and view the results.

Implementation

I decided to complete this task using ArcGIS Pro via its python library, ArcPy. I built the GUI using Tkinter and used Pillow to complete image manipulation tasks. Tkinter and Pillow are both python libraries that come packaged with ArcGIS Pro's python environment. This makes it easy to share the tool with anyone in the department who has ArcGIS Pro installed without having to worry about installing additional libraries.

To keep this project organized, I divided the project into three .py files. These files are each dedicated to their own task:

- Main.py - This file is the user's entry point to the program and controls the GUI. It allows the user to set the K value and to view the results of the analysis. Main.py calls functions from the other 2 files to complete the required tasks.
- RunAnalysis.py - This file is a collection of functions that run ArcPy's geoprocessing tools.
- GenerateReports.py - The maps showing the results of the analysis are created with these functions.

Data Analysis in ArcPy

In order to find the relationship between ground water nitrates and cancer rates, the first task is to consolidate the data into a single layer with a common geographic unit between the two. The WDNR requested the analysis to be completed at the tract level. Since they provided a tract shapefile with cancer rate data, it makes sense to find the average ground water nitrate concentration in each tract. I achieved this with the following steps:

1. The nitrate data that the WDNR provided is a point layer representing well water tests. I converted this data into a continuous raster layer using IDW interpolation. This allows you to estimate the nitrate concentration in any point throughout the state. The K value that the user enters in the GUI is used to adjust the Power parameter in the IDW tool.
2. ArcPy's *Zonal Statistics as Table* tool will take the nitrate raster file and produce a set of statistics based on another input vector file. I used this tool to find the mean nitrate concentration in each tract. The resolution of the IDW output needed to be adjusted to ensure that each tract had at least one cell within its geometry. This is from a limitation with esri's zonal statistics tool. It finds raster values from the center point of each cell in the raster file. If there are no cells centered within a given feature's geometry, it is not given a value. This was the case for many of the smaller tracts using the default resolution from the IDW tool. The problem was solved by using a finer resolution.
3. The output of the zonal statistics tool was joined to the tract shapefile.

This fulfills the requirement of having both nitrate concentration and cancer rates in the same layer, allowing me to compare the two values. The next step was to run a linear regression analysis on this data to determine if there is a relationship between the two. This was completed using the ordinary least squares tool. The nitrate concentration was used as the explanatory variable and the cancer rate was the dependent variable. Finally, the results of the OLS analysis are only considered accurate if there is no spatial autocorrelation in its output's residuals, so the data analysis is completed with the Moran's I tool.

Generating Output Maps

The output maps were straightforward to make. I created a Pro project within the python program's folder that contains a map and a layout for both the IDW and OLS outputs. After I designed the layouts in Pro, it was easy to access them via ArcPy's mp module to change the data sources and labels to match the latest output from the tool.

The GUI

I created the graphical user interface using Tkinter. My goal was to create a GUI that is easy to use and informative. It needed to fulfill the requirements of setting the K value for IDW interpolation, run the analysis, display the results, and provide the option to save the results to a file.

Making the tool easy to use is important, especially if the tool is to be shared with people who are generally unfamiliar with python. This tool only uses libraries that come with ArcGIS Pro's default python environment, so any user with Pro installed can use it. I created a batch file that will start the script using the correct python environment.

*Run_Project1.bat - Notepad

```
"C:\Program Files\ArcGIS\Pro\bin\Python\envs\arcgispro-py3\python.exe" "Main.py"
```

This simple batch file ensures that the script is run from Pro's python environment.

It will make using the tool effortless for non-python users.

The main window is split into two, with the setup frame on the left and the results frame on the right. The setup frame allows the user to adjust the K value and start the analysis. The results frame shows an image of the IDW and OLS results and allows the user to save them both as PDF files. It also has buttons to show the detailed OLS stats and the results of the autocorrelation analysis.

Running the geoprocessing tools can take a long time. It's a good idea to keep the user informed of what the tool is currently doing and how much more is left to be done. I created a progress bar widget as a subclass of tk's Frame class. The progress bar was created in its own .py file and can be very easily reused in future Tkinter projects. After importing and instantiating the ProgressBar class, the widget will fill the area of the parent object. All of the sub-widgets are created using relative positioning and size. This project covers the entire window area with the progress bar, but it would look good if its parent is a smaller frame as well. The progress and status message are updated with the .set_status() and .set_prog() methods, as seen in the screenshot below.

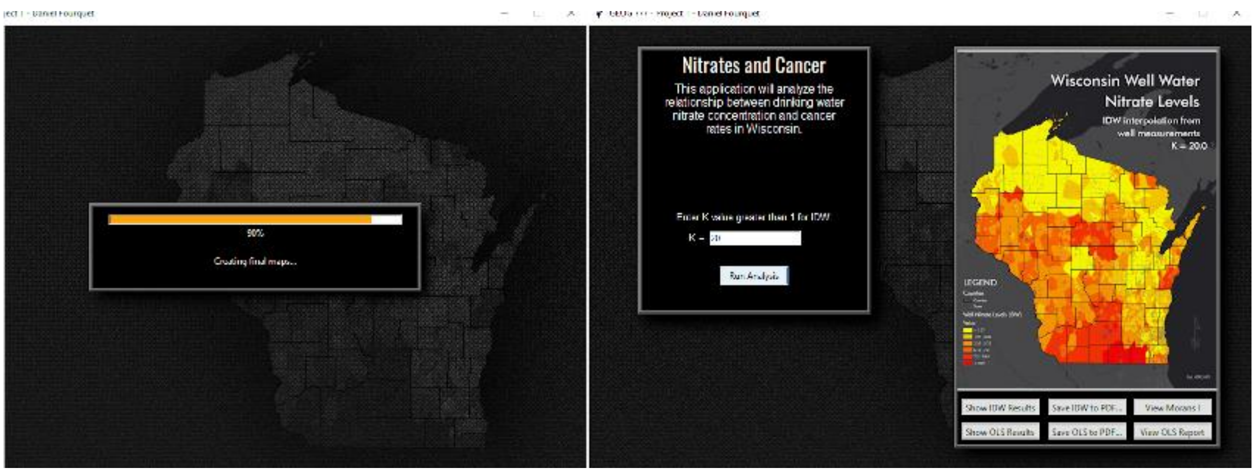
```
prog.set_status("Summarizing results of the nitrate interpolation at the tract level...")
prog.set_prog(0.3)
nitrateVals = ra.get_average_nitrate_dict(tracts, "GEOID10", idwOutput, k)

prog.set_status("Updating nitrates field in tracts...")
prog.set_prog(0.5)
ra.update_nitrates_field(nitrateVals, tracts)

prog.set_status("Running ordinary least squares linear regression...")
```

The ProgressBar "prog" in action in Main.py

One thing that I didn't really like about Tkinter was how ugly the 90s-style GUI is. I found that using themed widgets (eg ttk.Button instead of tk.Button) improves the look, but I decided to have a little fun creating a background image for the tool. I created it in GIMP using a desaturated version of the IDW output and a web texture. I used the drop shadow filter to create a 3D effect for the frames and progress bar. Since these elements move, I had to create a background image for each position that the frames take and control which image one is shown using code. Honestly the result still looks quite outdated, but I'll bet it would've looked awesome 25 years ago!



The progress bar and results with bckground image.

Conclusions

The conclusions of the analysis comparing nitrates and ground water can be found in the results of the OLS tool. OLS linear regression attempts to build a model that predicts the values of the dependent variable (cancer rates) based on the explanatory variable (ground water nitrates). The OLS tool creates a pdf report which can be opened via the results frame in the GUI. The report shows a positive coefficient and statistically significant robust probability score. This means that there is a positive relationship between nitrates and cancer rates. The higher the ground water nitrates tend to be, the higher the cancer rates tend to be.

This doesn't tell the whole story though. The adjusted R-squared value measures the performance of the model on a scale from 0 (the explanatory variable has no relationship to the dependent variable) to 1 (the explanatory variable perfectly predicts the dependent variable). In the case of this analysis, the adjusted R-squared is 0.02, meaning that the nitrates only account for about 2% of the observed cancer rates. Further, the Moran's I analysis found that the standard residuals are very spatially autocorrelated. If nitrates were a significant explanation for cancer rates, then the Moran's I would have shown that the standard residuals are randomly distributed. It's clear that there are other missing explanatory variables that have a greater impact on the cancer rate than ground water nitrates.

One problem with this study is that it only looks at ground water and doesn't take treated municipal water sources into account. Adding nitrate level tests from municipal sources to the model would be a logical next step.