Ethan Tabbert

Geospatial Programming

Professor Caglar Koylu

11 May 2022

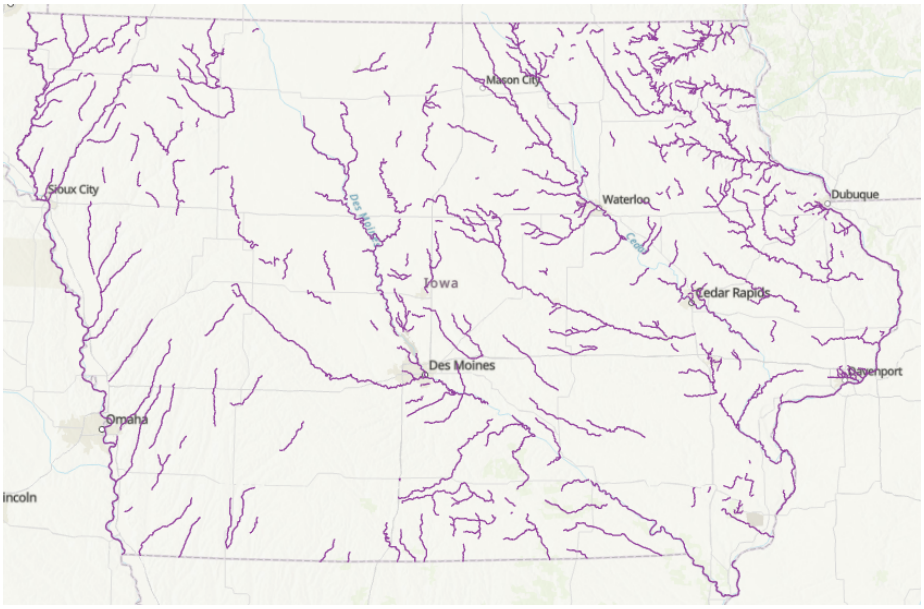
Impaired Streams and Cropland Correlation in Iowa

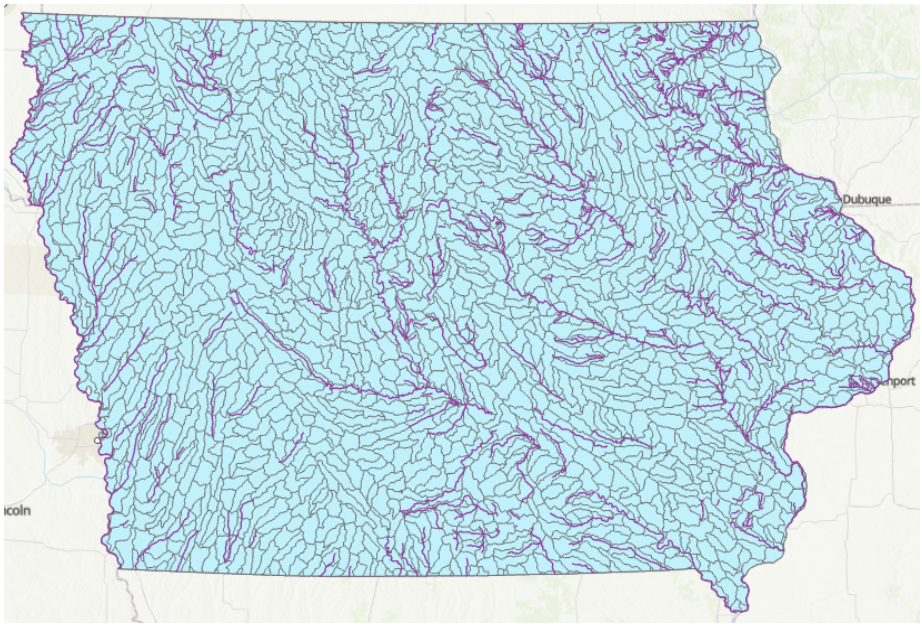
When it comes to coming up with overarching goals for GIS projects, I often struggle to come up with ideas. This is one of the bigger obstacles for me in these classes. For this class though, I decided to do some research on something that I am personally curious about. I've always heard people talk about how runoff from croplands negatively impacts rivers, streams, and other bodies of water. This is likely due to the sprays and chemicals used in order to preserve crops against pests. I decided I wanted to look into the data and see if I could find a correlation to indicate that this is true. So my ultimate question was is there a correlation between impaired streams and croplands in Iowa.

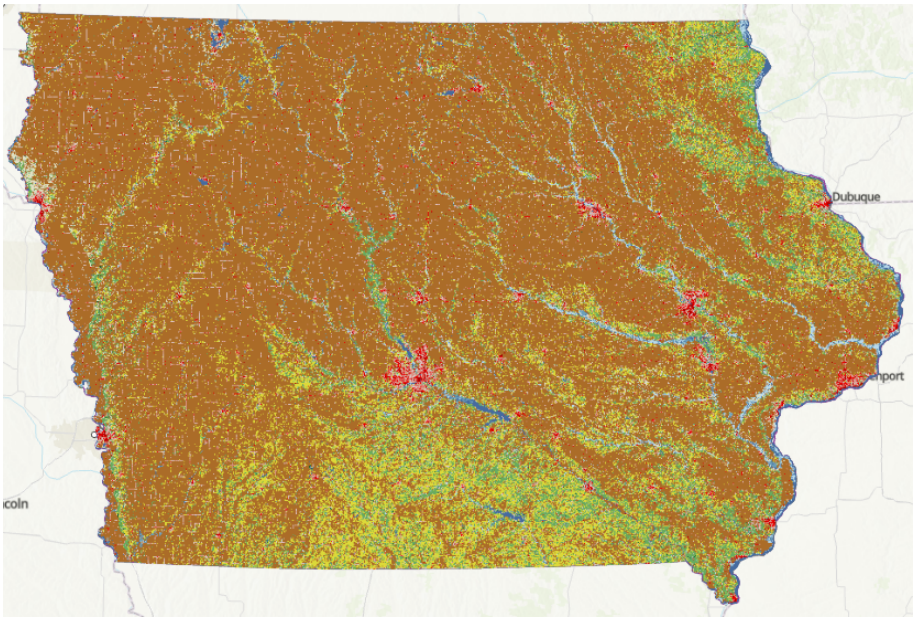
I also set another goal for myself when starting this project. As mentioned, since I tend to struggle a bit with forming good GIS questions, I wanted to not only make sure I learned and used the arcpy library for coding in a notebook, but I also wanted to incorporate some standard Python code in order to more closely analyze the data produced by the Arcpy notebook. This led me to create an additional Python program in which I import the resulting CSV file to do some analysis of the data.

For this project, I needed a few different features. First, I downloaded a polyline shapefile that contained all the impaired streams in Iowa in 2020. This allowed me to see where those streams are located and eventually I could use it to select watersheds. Second, I needed the watershed data for the state of Iowa. This download provided me with three different shapefiles

that contained different scales of watersheds. With this, I can visualize from larger watersheds down to smaller, more localized watersheds. I ended up going with the smaller watersheds, as the scale looked like it best fit the data I was working with. Finally, I needed ground cover information so I could see how much land was covered in crops as opposed to other types of ground cover, such as urban or forested areas. This was obtained in a raster format from 2019. Here are some images of the data sources.



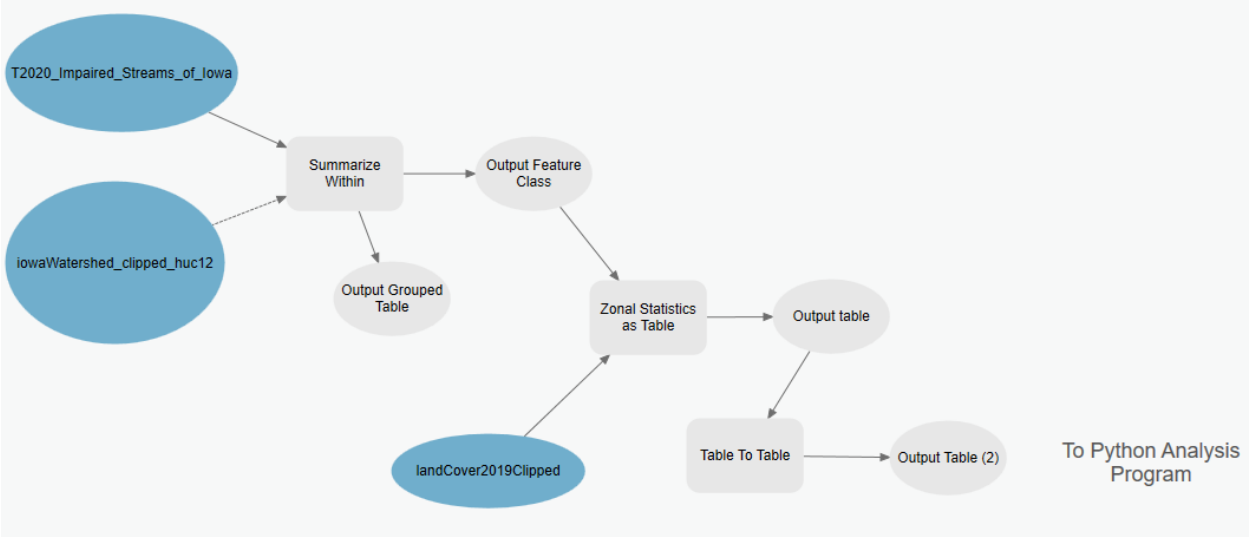




Top to bottom: Impaired streams in Iowa in 2020, HUC 12 watersheds with impaired streams overlaid, and ground cover raster from 2019.

Each of the data pieces were obtained from different websites listed in the references. In the case of the watershed and impaired stream data, those were obtained via surveying and software analysis. For the ground cover raster, information was obtained using Landsat imagery. All this data was gathered within about a two year timespan, considering I have some data from 2019 and some from 2020. This appears to be close enough for an accurate analysis.

In the case of the ground cover raster, I needed a bit more information in order to utilize the data. Luckily, a legend was provided along with the imagery that showed what each pixel value meant. This allowed me to piece apart the data and only analyze what I needed. For example, I was primarily interested in croplands, which corresponded with the pixel value 82. This way I knew what each pixel indicated based on its value/color.



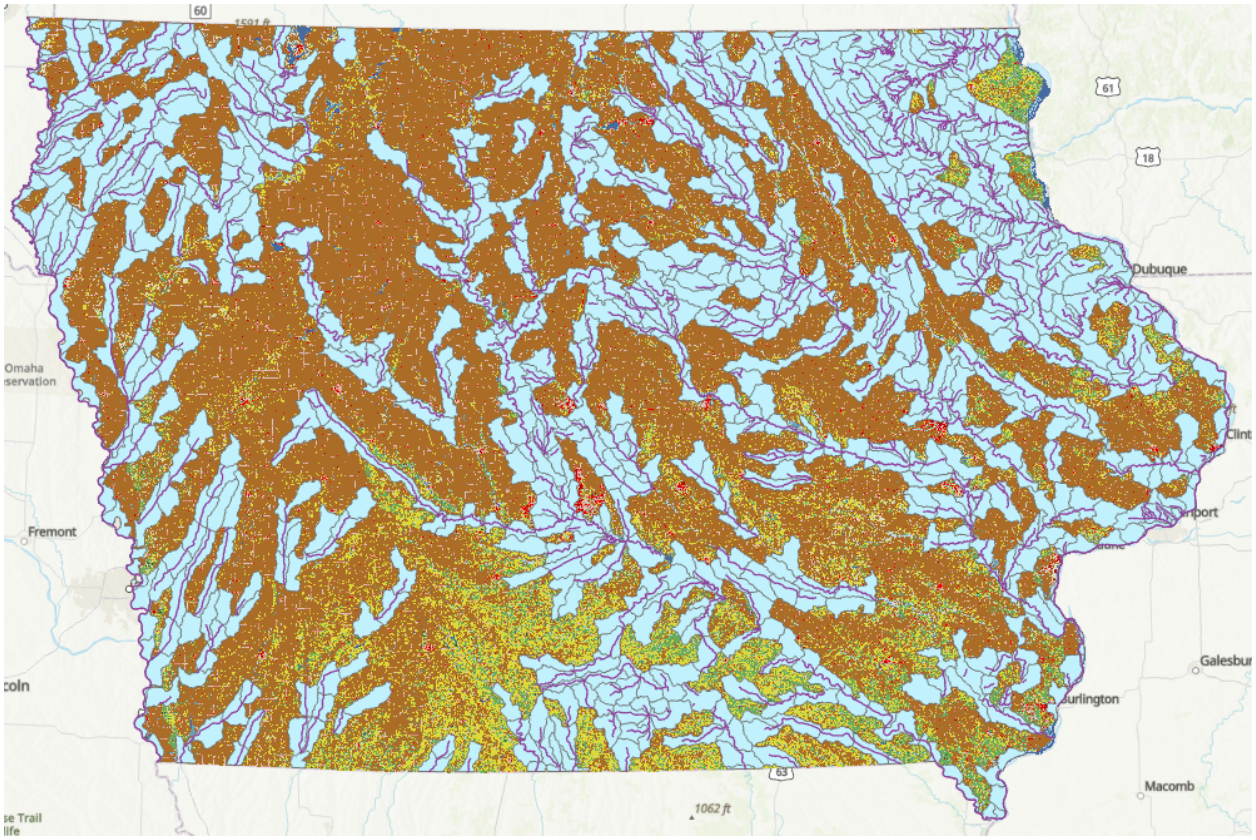
This model displays a simplified version of my workflow. The main parts it is missing is I did a lot of clipping and converting in the beginning to get the data down to the state of Iowa. Once the data was clipped, I put the 2020 impaired streams and the HUC 12 watershed into the summarize within so I could get a new shapefile that only contained the watersheds that had impaired streams running through them. Once that was done, the output then goes into the zonal statistics as table tool along with the clipped land cover data. This will examine all the pixel values within the watershed polygons and add specific pieces of that data to a table. In my case, I wanted to see what the majority of pixels were in each watershed. I also set it up so it would show me which pixels were in the 95 percentile within those watershed polygons. After that is completed, the table then gets converted to a CSV file and goes through my dataAnalysis.py file.

In my Python program, the CSV is imported and parsed through, which pulls out the data needed for analysis. While it is being parsed, the program checks to make sure that the data is correct and usable, including first line detection so as to not include the field names. That data is then stored into a few different arrays. I then set the program to parse through those arrays and generate a few different pieces of data that give me some more insight into my initial questions.

In total, there were 713 watershed zones that had impaired streams running through them. Of those, 611 watershed zones had a majority of the land cover classified as cropland. 569 watershed zones are more than 95% cropland. Within the watershed zones that contain impaired streams, 85.69% are majority cropland and 79.8% have over 95% cropland. These values can be seen in my Python program output pictured below.

```
========================== RESTART: E:\dataAnalysis.py ==========================
>>> import_data("outputData.csv")
Total Count: 713
Zones where majority is cropland: 611
Zones where cropland takes up 95% or more of the zone: 569
Percent of zones with majority cropland: 85.69
Percent of zones with over 95% cropland: 79.80
>>>
```

I think while this data is somewhat generic, I learned a lot about the Arcpy library and how Python can be used to manipulate datasets in order to see outcomes to form questions. In my specific application, this data doesn't prove a specific issue or solution, but rather it shows me the correlation between cropland and impaired streams. For a program like mine, I would use it to first find a correlation and see that there is an issue. Now that I can see the issue, I would gather more in depth data and write more in depth code in order to find more specifics. The correlation between impaired streams and croplands is quite apparent through this data. Going forward, it would be good to examine specific streams and find out what exactly is impairing them. It would also be good to research what types of sprays and chemicals are being used in farms around Iowa. With more information like that, we could discover new issues and solutions revolving around the specifics behind impaired streams as a result of crop cultivation. I would not know to look further into it without a program like mine however, which helps to bring some transparency to the issue without spending too much time getting into specifics before knowing whether or not a problem exists to begin with.



Above: Impaired streams overlaid onto the watersheds they contact overlaid onto the ground cover data.

With that being said, the limitations of my project do not allow for in depth analysis of the how and why streams are impaired near farmlands. However, you could use my program to discover that a problem exists in the first place, and then present it to a group that has more resources to do deeper research. I don't think I would do anything differently if I could do the project again, mainly because I feel like I learned a lot and got out of it the skills that I needed to. Some students may be more focused on the geography portion and gain more from doing more research, working with more data, and seeing more in depth solutions. For me though, I gained significant experience with Jupyter notebooks, Python, Github, ArcGIS Pro, and the Arcpy library that I didn't have before. I ultimately think these skills will be more useful to me in my career path as a programmer.

Sources:

Iowa Geospatial Data, https://geodata.iowa.gov/.

"Multi-Resolution Land Characteristics (MRLC) Consortium." *Multi-Resolution Land Characteristics (MRLC) Consortium*, https://www.mrlc.gov/.

Links to Final Code:

https://github.com/geog3050/etabbert/blob/main/finalNotebook.py (Arcpy notebook code)

https://github.com/geog3050/etabbert/blob/main/dataAnalysis.py (Data analysis Python code)

https://github.com/geog3050/etabbert/blob/main/outputData.csv (Resulting data from notebook)