Lab Report

Title: Lab 1
Notice: Dr. Bryan Runck
Author: Emily Cavazos
Date: Oct 6, 2021

Project Repository: https://github.com/cavaz020/GIS5571/
Google Drive Link: N/A
Time Spent: 15 hours

Abstract
This work involved building an ETL to request data from the NDAWN API of the last 30 days of temperature data from all of their stations. Using this data, I created points on the map and then performed three different interpolation methods on the data. The three methods I used were Inverse Distance Weighted, Kriging spherical, and Global Polynomial Interpolation. I then go on to discuss the differences between the three methods and what I found to be the most accurate.

Problem Statement
The problem at hand involved building an ETL to pull the last 30 days of temperature data from the NDAWN site for all of the NDAWN stations. Then, the goal was to compare and contrast three interpolation methods by using them to create an interpolated temperature map for this data.

Table 1. Interpolation methods to be compared in this lab

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	IDW	Interpolation method 1	Will be used to interpolate temperature data	N/A	N/A	N/A
2	Kriging	Interpolation method 2	Will be used to interpolate temperature data	N/A	N/A	N/A
3	GPI	Interpolation method 3	Will be used to interpolate temperature data	N/A	N/A	N/A

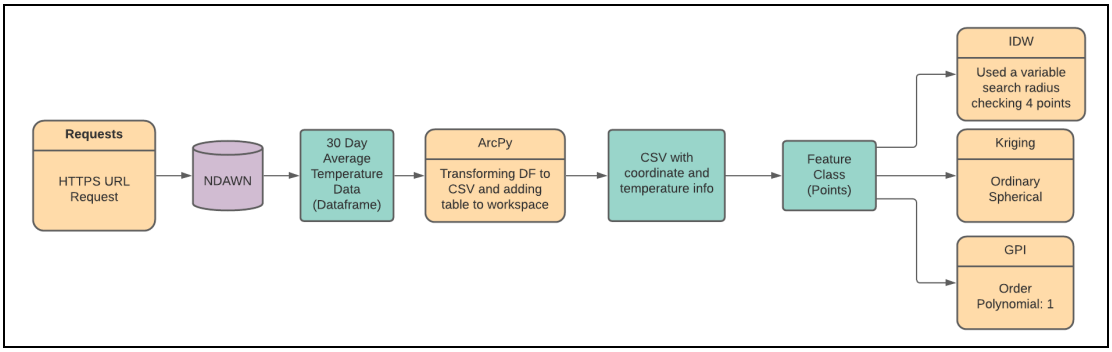
Input Data
This data is actually an output after sending a request to the NDAWN API.

Table 2. Data used in interpolations

#	Title	Purpose in Analysis	Link to Source
1	NDAWN Temperature Data for the last 30 days (11/05/21-12/04/21)	Dataset of the last 30 days of average temperature at each of the NDAWN stations. This data will be used to interpolate temperature for the overall area around the stations within Minnesota using 3 different methods.	Data retrieved using ETL but can be found here: https://ndawn.ndsu.nodak.edu/weather-data-monthly.html

Method:

Figure 1. Data Flow diagram of the process for preprocessing the data and each interpolation method



The methods for this lab were very straightforward. The ETL we used to get the data was built for us and we simply used the function to create a request for our dates and for all of the locations as seen in my code.

The next steps included taking the pandas dataframe that the data was held in and creating a CSV of the data. I then added the table to the workspace and created a feature class of points for the stations at each date. Finally, I ran the three different interpolation methods and visualized the outcomes. The output of each interpolation is a raster color coded by the temperature category. The resulting rasters are shown below.

Image 1. IDW

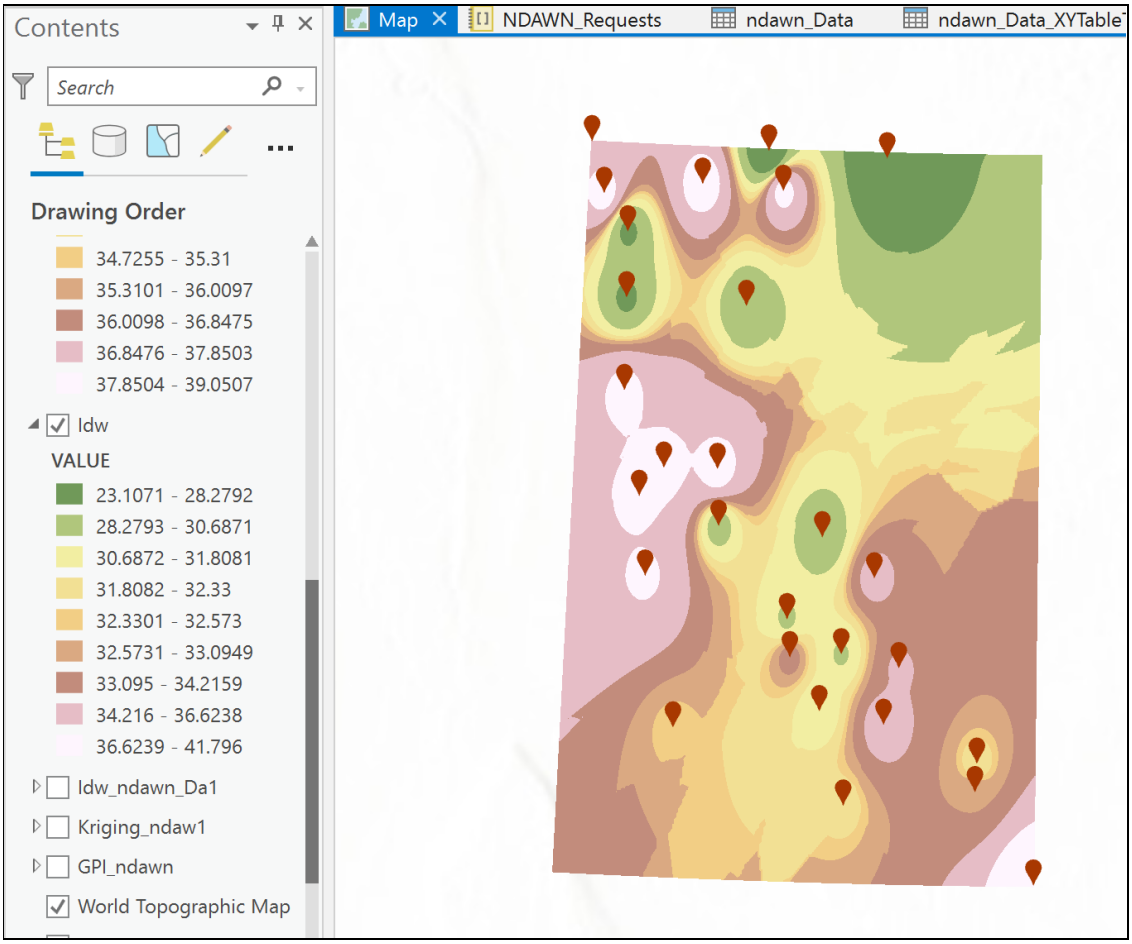

Contents
Search
Drawing Order
34.7255 - 35.31
35.3101 - 36.0097
36.0098 - 36.8475
36.8476 - 37.8503
37.8504 - 39.0507
Idw
VALUE
23.1071 - 28.2792
28.2793 - 30.6871
30.6872 - 31.8081
31.8082 - 32.33
32.3301 - 32.573
32.5731 - 33.0949
33.095 - 34.2159
34.216 - 36.6238
36.6239 - 41.796
Idw_ndawn_Da1
Kriging_ndaw1
GPI_ndawn
World Topographic Map
Map
NDAWN_Requests
ndawn_Data
ndawn_Data_XYTable

Image 2. Kriging spherical

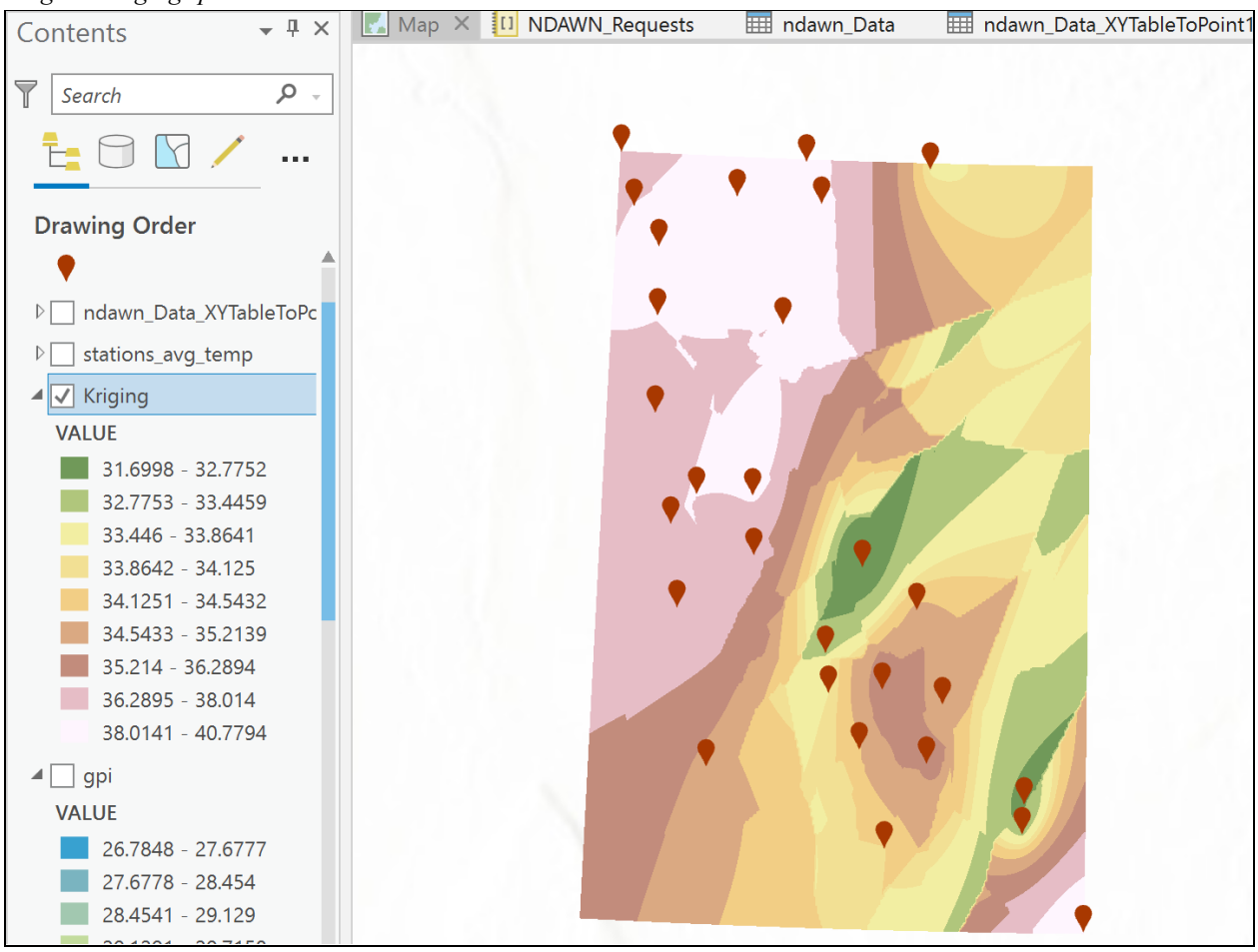
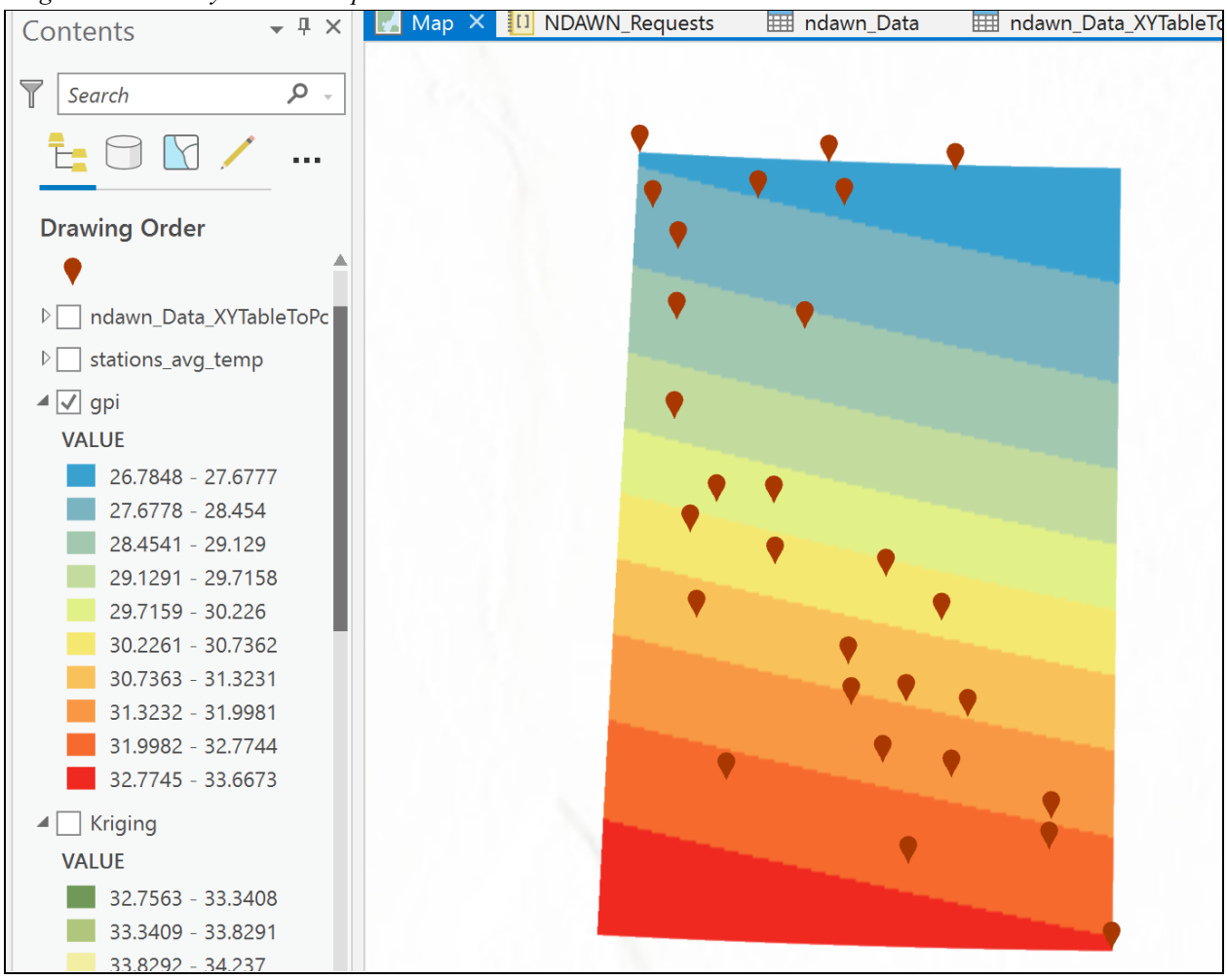


Image 3. Global Polynomial Interpolation



Results:
The table below shows my understanding of the differences between the three methods in performing the same function.

Table 1. Comparison between the three Interpolation methods.

Interpolation Method	Information required	Requires measurement or model of spatial autocorrelation?	Output Type	Level of assumptions/ complexity	Type of interpolation	Smoothness of the output	Provides Uncertainty of the predicted values	Processing Speed
IDW	One prediction per location	Implicit (Implicitly assumes spatial autocorrelation)	Prediction	Few	Exact	Not Smooth	No	Fast
Kriging	One prediction per location, quantile value, or predicted values and errors	Yes	Prediction error or probability	Many (Kriging methods offer options to make the data comply with the assumptions)	Inexact	Intermediate	Yes	Fast
GPI	One prediction per location	No	Prediction	Few	Inexact	Smooth	No	Fast

Results Verification
This assignment did not require much output but was more about the experience of working with three interpolation methods so the results verification is more so just speaking on my experiences and findings of working with these.

Discussion and Conclusion
The first interpolation method I used was Inverse Distance Weighted (IDW). As we learned in class, the idea behind IDW is that the observation points close to the target point are considered more important and are thus given a larger influence on the target value than those far away. This was a very simple method. I chose to use 4 surrounding points to verify but I feel like I should have chosen more. The accuracy can be dependent on what inputs you put in so this method is usually less accurate.

The literature seemed to be in agreement with recommending using some form of kriging. This may be because it takes in more assumptions which means it takes a greater effort to build the model but methods that have many assumptions tend to be more flexible and can produce better results. I found an article that found that Kriging-Spherical and Kriging-Exponential were found to have the highest accuracy. This informed my decision to use Kriging-Spherical for my results above. Simple kriging is an interpolation method where the idea is to give a weight to every observed point and then compute the target value from the mean plus a linear combination of the weights times the difference between the mean and the value in the observation points.

For my last interpolation method I chose Global Polynomial Interpolation (GPI) which is usually used when the surface in which the sample points reside slowly changes from region to region which I thought would make sense for the points because temperatures usually tend to get warmer closer to the south. GPI uses polynomials and apparently the more complex the polynomial, the more difficult it is to make meaning of the results. For my polynomial I entered the order of 1.

References

1. Wenjing Cao, JinXing Hu and Xiaomin Yu, "A study on temperature interpolation methods based on GIS," 2009 17th International Conference on Geoinformatics, 2009, pp. 1-5, doi: 10.1109/GEOINFORMATICS.2009.5293422.
2. Cronqvist, F. (2018). Interpolation of temperature data for improved weather forecasts (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-229731
3. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

Self-score

Category	Description	Points Possible	Score
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	**26**
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	**20**
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	**26**
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	**20**
		100	**92**