Lab 3 Report

Title: Building a fully functional real-time data visualization and analysis workflow and comparing three types of interpolation
Notice: Dr. Bryan Runck
Author: Zain Ul Abdin Siyal
Date: 12/02/2022

Project Repository: https://github.com/Zain1443/GIS5571.git
Google Drive Link:
Time Spent: 12 hours

Part 02

Abstract

It is nearly impossible to have values for each location in any area of interest. This is where interpolation uses vector points with known values to estimate values at unknown locations. It basically follows the spatial autocorrelation phenomena and Tobler's first law of geography. The aim of lab 3 part 2 is to build a fully functional real-time data visualization and analysis workflow (basically an ETL) that can pull the latest monthly temperature data from the North Dakota Agricultural Weather Network site for all the NDAWN stations. Also, using four different interpolation techniques, namely Inverse Distance Weighting (IDW), Natural Neighbors, Ordinary Origins/Kriging, and Spline, to generate continuous surface based on temperature data. Finally, the similarities and differences in their results were discussed.

Problem Statement

Since it is nearly impossible to get temperature data for each location within a particular study extent; therefore, interpolation techniques are used. They basically measure the similarity between nearby observations and generate a continuous temperature surface. Most interpolation methods handle large numbers of sample points efficiently, which is suitable for variable and dense point spacing. For this particular lab, four different interpolation techniques are used to estimate the temperature at all of the NDAWN stations and results are compared and contrasted.

Table 1. Required Data to interpolate temperature at all of the NDAWN stations.

#	Requirement	Defined As	(Spatial) Data	Dataset	Preparation
1	Weather Data	Latest monthly Raw Weather data from the DNAWN site for all of the NDAWN stations	Discontinuous Data	Temp, elevation, wind, humidity	NDAWN

Input Data

The input dataset that is required here is the latest monthly raw data for identifying monthly average, minimum and maximum temperatures at all NDAWN stations.

Table 2. Input dataset.

#	Title	Purpose in Analysis	Link to Source
1	Latest monthly Weather Data	Latest monthly Raw data for identifying monthly average, minimum and maximum temperatures at all NDAWN stations.	NDAWN

Methods

The methodology to build a fully functional real-time data visualization and analysis workflow that can pull the latest monthly temperature data from the DNAWN site for all of the NDAWN stations is summarized in the data flow diagram, shown in Figure 1. In order to build an ETL that can pull temperature data from any website, the python libraries such as requests, pandas, and sys are imported. Subsequently, base_ndawn_url is imported after setting up the necessary parameters (selecting all weather stations and the latest monthly daily data) on the NDAWN website. A request to the NDAWN page is made and CSV file is opened using requests.get and with open commands.

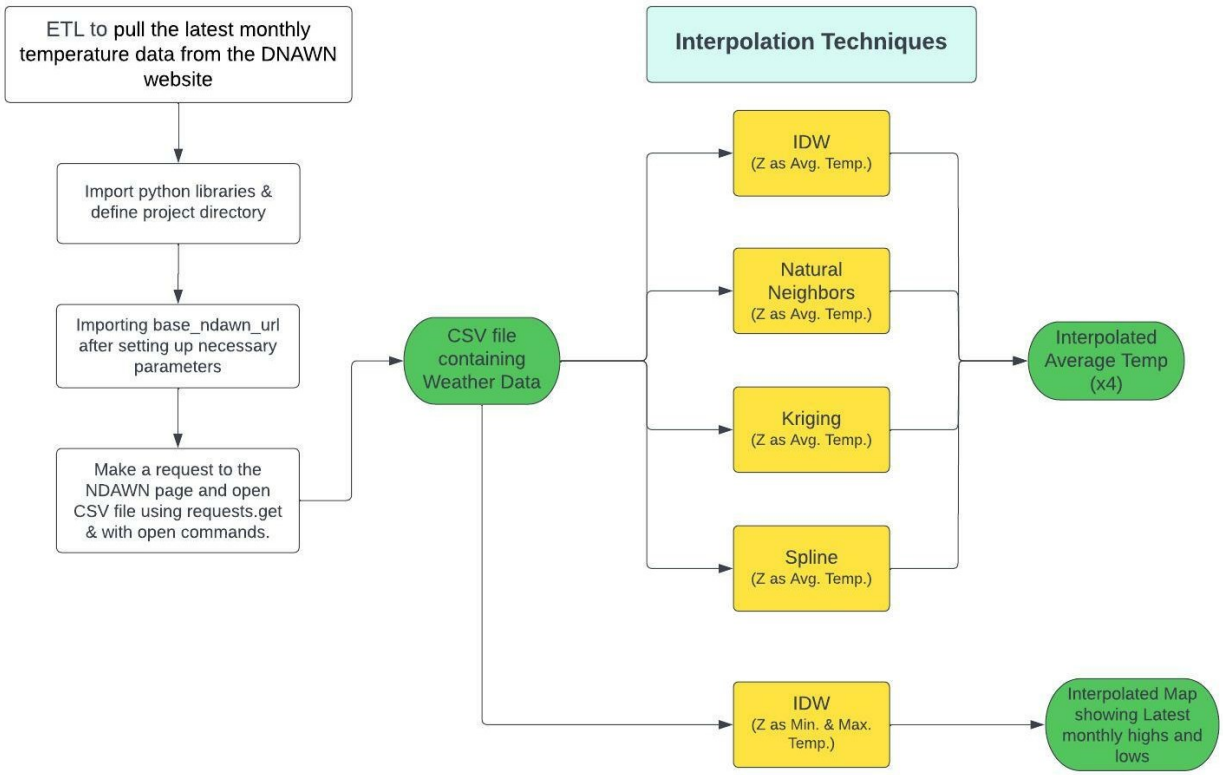


Figure 1: Flow Diagram of the entire Methodology.

The CSV file is imported as points in ArcPro using XY Table to point tool. Lastly, various interpolation techniques are used to interpolate average temperature, and results are compared and contrasted.

Results

The methodology described in figure 1 is followed sternly to get the Interpolated average temperature maps. It is quite evident that IDW did a pretty decent job as compared to Ordinary Origins/Kriging, Spline and Natural Neighbors. For this reason, it can be said that the IDW approach is the best option for interpolating this set of average temperature data because it will have a statistical-sounding relation. According to this method, the influence of the variable being mapped diminishes as it gets further away from the location where it was sampled.

	Station Name	Latitude	Longitude	Elevation	Year	Month	Day	Max Temp	Max Temp Flag	Normal Max Temp	Departure from Normal Daily Maximum Air Temperature	Departure from Normal Daily Maximum Air Temperature Flag	Min Temp	Min Temp Flag	Normal Min Temp	Departure from Normal Daily Minimum Air Temperature	Depart from Nor D Minim Temperat F
1	Ada	47.32119	-96.51406	910	2022.0	11.0	1.0	70.610	NaN	44.94	25.670	NaN	27.538	NaN	25.68	1.858	N
2	Ada	47.32119	-96.51406	910	2022.0	11.0	2.0	68.216	NaN	44.34	23.876	NaN	51.998	NaN	25.18	26.818	N
3	Ada	47.32119	-96.51406	910	2022.0	11.0	3.0	54.500	NaN	43.74	10.760	NaN	32.760	NaN	24.77	7.990	N
4	Ada	47.32119	-96.51406	910	2022.0	11.0	4.0	40.240	NaN	43.13	-2.890	NaN	16.214	NaN	24.27	-8.056	N
5	Ada	47.32119	-96.51406	910	2022.0	11.0	5.0	44.958	NaN	42.53	2.428	NaN	15.710	NaN	23.77	-8.060	N
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4226	Zeeland	46.01351	-99.68768	2070	2022.0	11.0	26.0	38.952	NaN	34.43	4.522	NaN	22.595	NaN	13.87	8.725	N
4227	Zeeland	46.01351	-99.68768	2070	2022.0	11.0	27.0	34.257	NaN	33.99	0.267	NaN	8.978	NaN	13.55	-4.572	N
4228	Zeeland	46.01351	-99.68768	2070	2022.0	11.0	28.0	31.883	NaN	33.49	-1.607	NaN	13.964	NaN	13.18	0.784	N
4229	Zeeland	46.01351	-99.68768	2070	2022.0	11.0	29.0	17.845	NaN	33.01	-15.165	NaN	10.004	NaN	12.87	-2.866	N
4230	Zeeland	46.01351	-99.68768	2070	2022.0	11.0	30.0	14.882	NaN	32.57	-17.688	NaN	0.806	NaN	12.50	-11.694	N

4230 rows × 21 columns

Figure 2: CSV file containing weather data for 4230 stations.

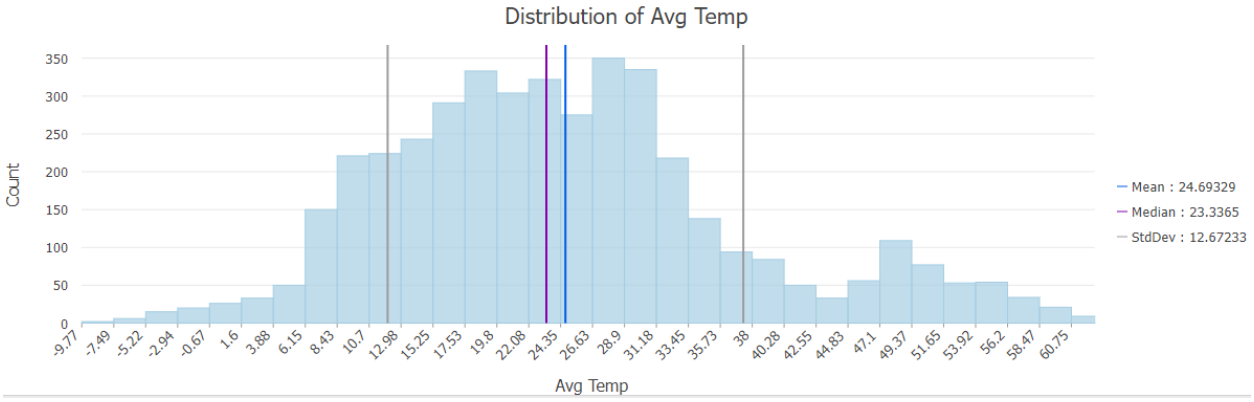
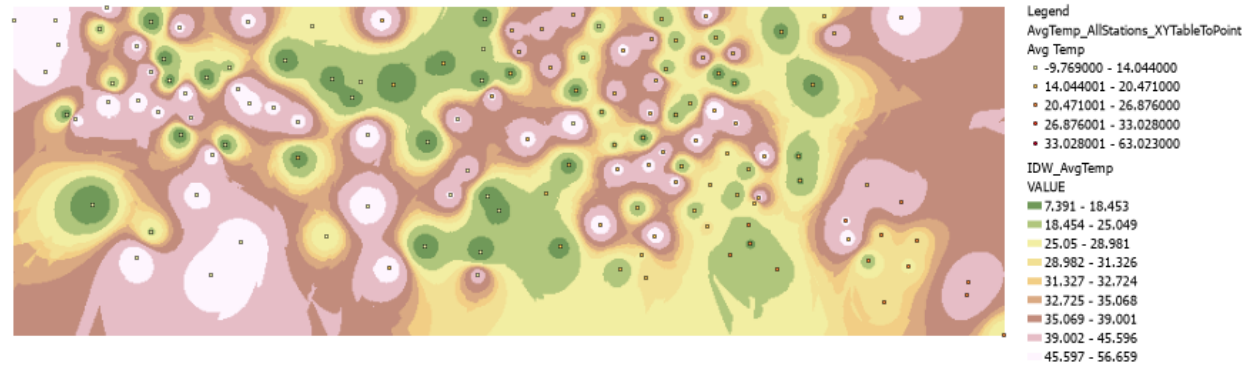
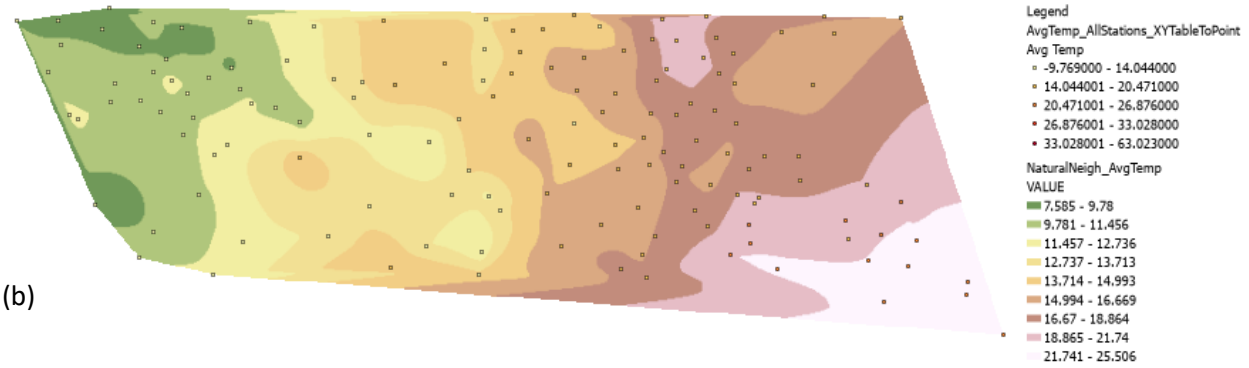


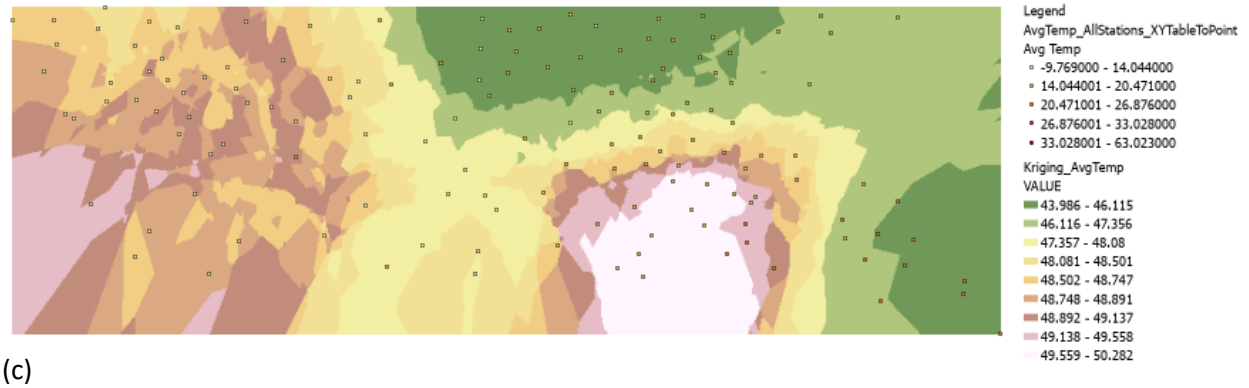
Figure 3: Distribution of Average Temperature



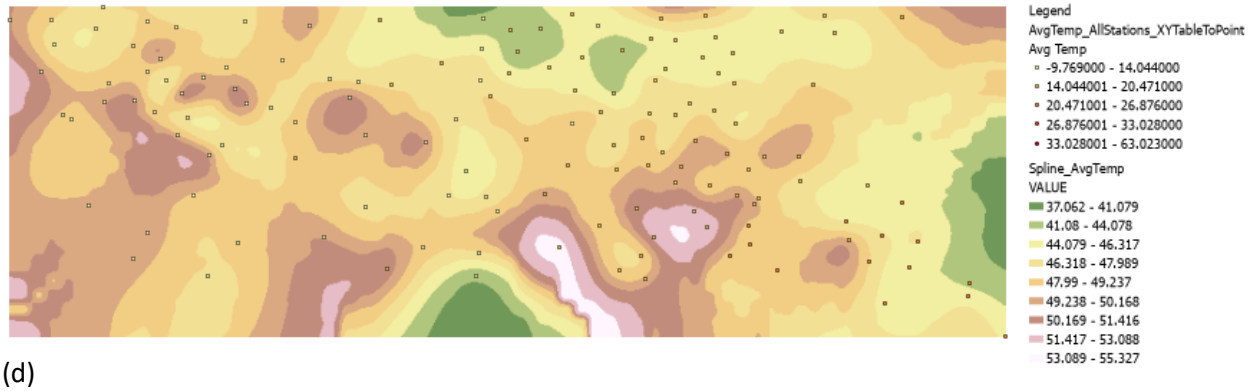
(a)



(b)



(c)

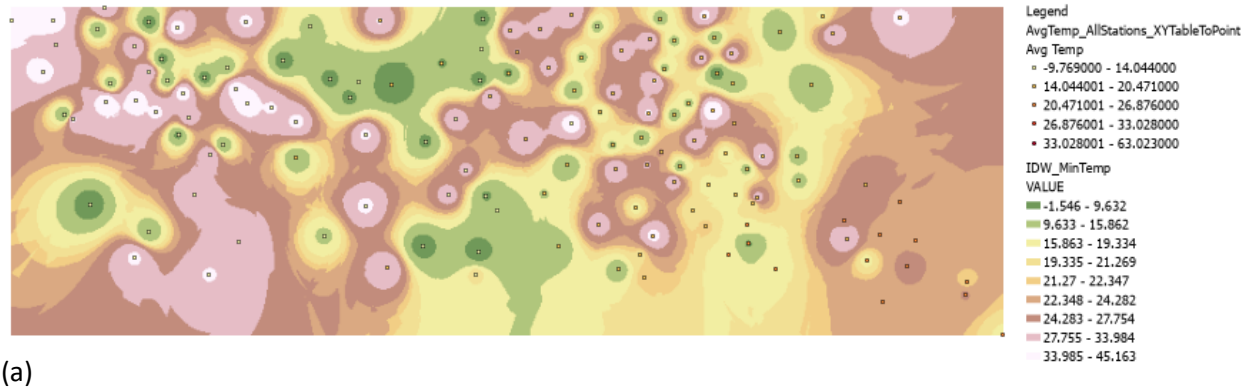


(d)

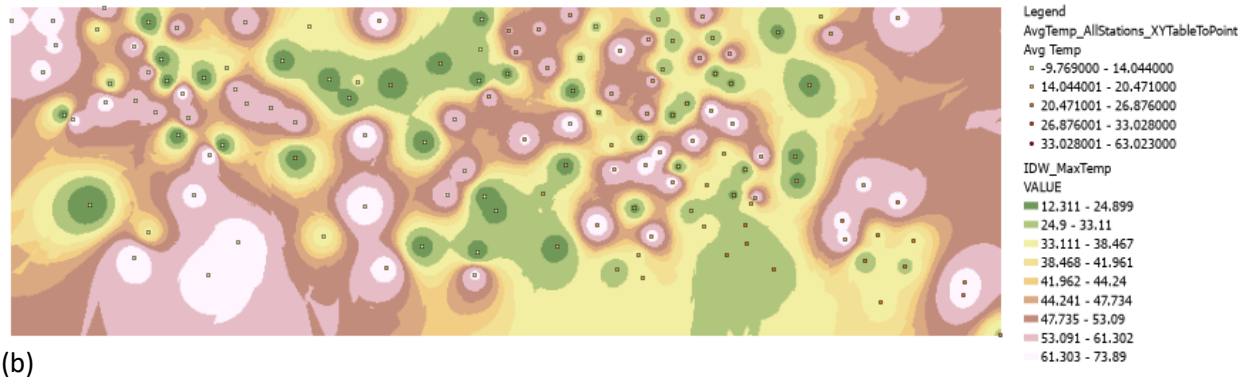
Figure 4: Interpolated Map layers with Z as avg. temperature (a) IDW, (b) Natural Neighbors, (c) Ordinary Origins/Kriging and (d) Spline.

The quality of the sample point set can also influence the choice of the interpolation method. The surface might not accurately depict the actual topography if the sample points are sparse or have a poor distribution, which might be the case here. Since there is a short distance between sample points, IDW visually seems to be the best fit.

In order to create an interpolated real-time temperature map for the highs and lows of the last 30 days from NDAWN, IDW technique was used, whose results are shown in figure 4.



(a)



(b)

Figure 5: Interpolated temperature map for the (a) lows and (b) highs of the last 30 days from NDAWN.

Results Verification

For each interpolation technique, I used the same parameters & ran the tools. In order to validate the results, I thought of using the Cross-validation method for all the data to estimate the trend and autocorrelation models by iterating through every weather station data and leaving one out, and then repeating that many times.

I tried to do that but wasn't able to do it because of uncertainty regarding how to perform cross-validation on that many sample points and my computer's limited processing power. Then I tried to verify the results visually as IDW covers the avg. Temperature range well than the other temperature maps.

Discussion and Conclusion

The best interpolation method to use relies on a variety of criteria. It relies on the variable's nature and the time scale on which it is represented; there is no universal approach that works for all issues. In order to find the optimal interpolation method for a particular project, all interpolation techniques should be investigated, and their outcomes should be compared rather than assuming one interpolation approach is superior to another. The literature suggests simple interpolation using a kriging approach is the best option for interpolating temperature data because it will have a statistical-sounding relation (articles that support this claim are cited). Moreover, if there aren't enough sample points, add extra in places where the landscape varies dramatically or frequently before attempting to use Kriging. However, IDW did a pretty decent job as compared to Ordinary Origins/Kriging, Spline and Natural Neighbors for interpolating this set of average temperature data.

References

1. NDAWN Station Location. NDAWN. Website. November 28, 2022. NDAWN - North Dakota Agricultural Weather Network (nodak.edu)
2. Going on the Grid -- An Intro to Gridding & Spatial Interpolation. neon. Website. November 25, 2022. https://www.neonscience.org/resources/learning-hub/tutorials/spatial-interpolation-basics
3. Choosing the Right Interpolation Method. GIS Resources. Website. November 30, 2022. https://gisresources.com/choosing-the-right-interpolation-method_2/
4. Classification trees of the interpolation methods offered in Geostatistical Analyst. ArcGIS Desktop. Website. November 21, 2022. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm
5. Comparing interpolation methods. Esri. Website. November 24, 2022. https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/comparing-interpolation-methods.htm
6. Natural neighbor interpolation. Wikipedia. Website. November 25, 2022. https://en.wikipedia.org/wiki/Natural_neighbor_interpolation
7. How to interpolate temperature correctly? Stack Exchange. Website. December 1, 2022. https://gis.stackexchange.com/questions/30905/how-to-interpolate-temperature-correctly
8. How to validate the interpolation result in ArcGIS. ResearchGate. Website. November 30, 2022. https://www.researchgate.net/post/How_to_validate_the_interpolation_result_in_ArcGIS
9. How IDW works. ArcGIS for Desktop. Website. November 20, 2022. https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-idw-works.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	28
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	23
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	27
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	18
		100	96