Lab 3 - Part 2 Lab Report

Title: Lab 3 - Part 2
Notice: Dr. Bryan Runck
Author: Ben Landers
Date: 11/26/2024

Project Repository: https://github.com/blanders98/GIS5571/tree/main/Lab3
Google Drive Link:
Time Spent: 8 hours

Abstract

Part 2 of this lab focuses on building a real-time data visualization and analysis workflow to interpolate temperature data from the North Dakota Agricultural Weather Network (NDAWN) website for the past 30 days. The goal is to develop an ETL pipeline to retrieve temperature data from NDAWN stations and visualize it on a map, displaying average monthly temperatures. The lab also compares and contrasts three interpolation methods— Inverse Distance Weighting (IDW), Kriging, and another of choice in which I chose a natural neighbor method —using the ESRI decision guide to justify method selection. The final deliverable includes a notebook capable of generating interpolated temperature maps in real-time, along with a literature review supporting the chosen interpolation techniques.

Problem Statement

Accurate spatial interpolation of temperature data is essential for analyzing climate patterns in agriculture. NDAWN provides location-specific temperature data, but this data is limited for broader spatial analysis. The challenge in part 2 of this lab is to develop a real-time workflow to interpolate the last 30 days of temperature data from NDAWN stations and generate continuous temperature maps. A key point here is the ETL pipeline we create should be able to consistently pull the most recent 30 days of temperature data, not just a single one time download of a random recent 30 days of data. The final task is to compare three interpolation methods, including IDW, Kriging, and a third of choice to identify the most effective technique for accurate temperature mapping.

Table 1. Data Used in this Lab

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Temperature data	NDAWN 30-Day Temperature Data	Point	.csv	[NDAWN Temperature Data]()	
2						
3						
4						

Input Data

Table 2. Data Used in this Lab

#	Title	Purpose in Analysis	Link to Source
1	NDAWN 30-Day Temperature Data	Points with temperature data for interpolation	NDAWN Daily Weather Data
2			
3			
4			

Methods

Figure 1. Part 2 data flow diagram (Mermaidchart.com)

To begin Lab 2 Part 2, I developed an ETL pipeline that retrieves the most recent daily average temperature data from NDAWN for each station over the past 30 days. This data is fetched using an API request and stored in a pandas DataFrame. I then created a feature class of these station points from the pandas dataframe. I made sure to add a column for station name, average temperature, and date. This gave me a layer for station points with 30 points per location, one for each day with an average temperature for that day. Since the interpolation analyses that will be conducted require a single point per location, I needed to aggregate this layer further. So I created a new layer which aggregated the initial station points layer's data so there is only a single point per location with columns for average temperature, max temperature, and minimum temperature for that station over the last 30 days. I then mapped the aggregated station points (*Figure 2.*) and labeled the average temperature column, limiting the label to one decimal place for readability.

Next I conducted six interpolations on the temperature data, one for high temperature and one for low temperature for three different interpolation methods. The first interpolation method I ran was for inverse distance weighting in which I called upon ArcPy's spatial analyst's functionality. I used arcpy.sa.Idw and used the aggregated station point layer and both high and low temperature separately to output a raster with interpolated contours (*Figures 3 and 4.*). Then I used ArcPy's spatial analysts's kriging functionality for a kriging interpolation. I used arcpy.sa.Kriging with the aggregated station point layer and both high and low temperature separately to create a kriging raster output (*Figures 5 and 6.*). I specified a spherical model to be used with a search distance of 0.01. For my final interpolation method I chose to do a 'many predictions per location' simulation based method. I chose to do a nearest neighbor interpolation for this where I called upon ArcPy's spatial analyst's natural neighbor functionality to create a natural neighbor output. For this I used arcpy.sa.NaturalNeighbor and inputted the aggregated station points and both high and low temperature data separately to create two output rasters (*Figures 7 and 8.*).

Results

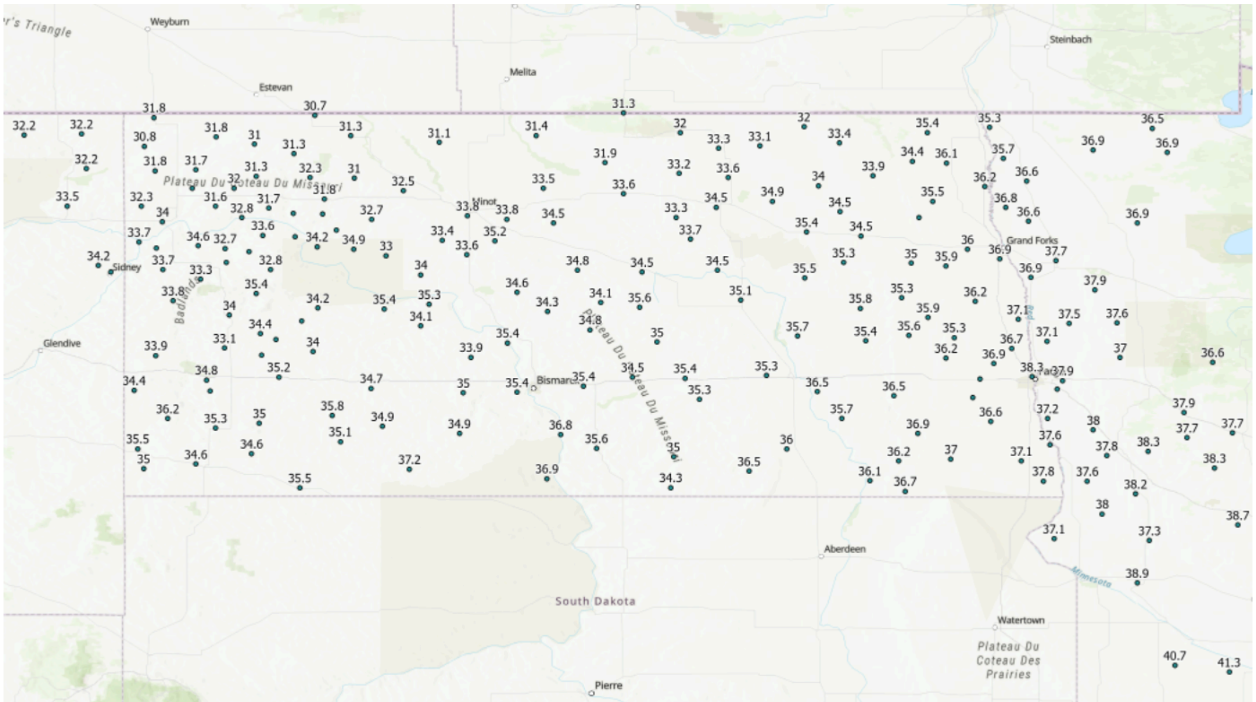

Figure 2. Aggregated station points with average daily temperature over the most recent 30 day period.

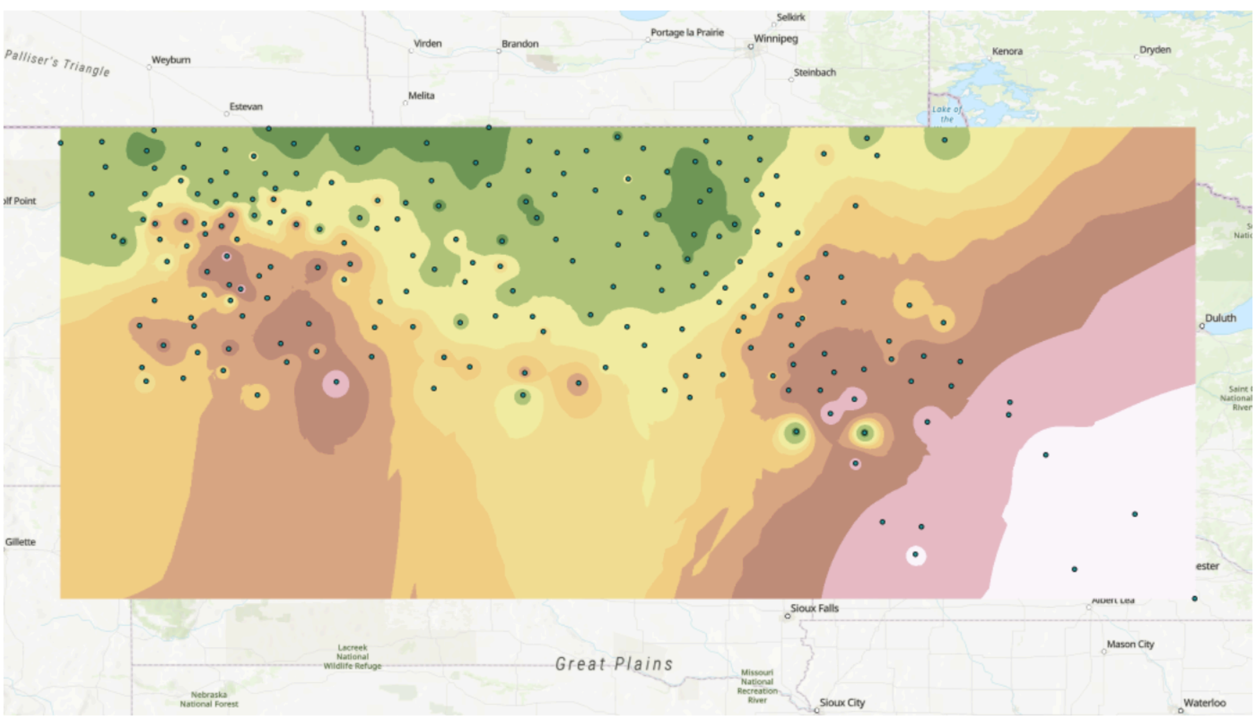

Figure 3. IDW Interpolation of High Average Temperature

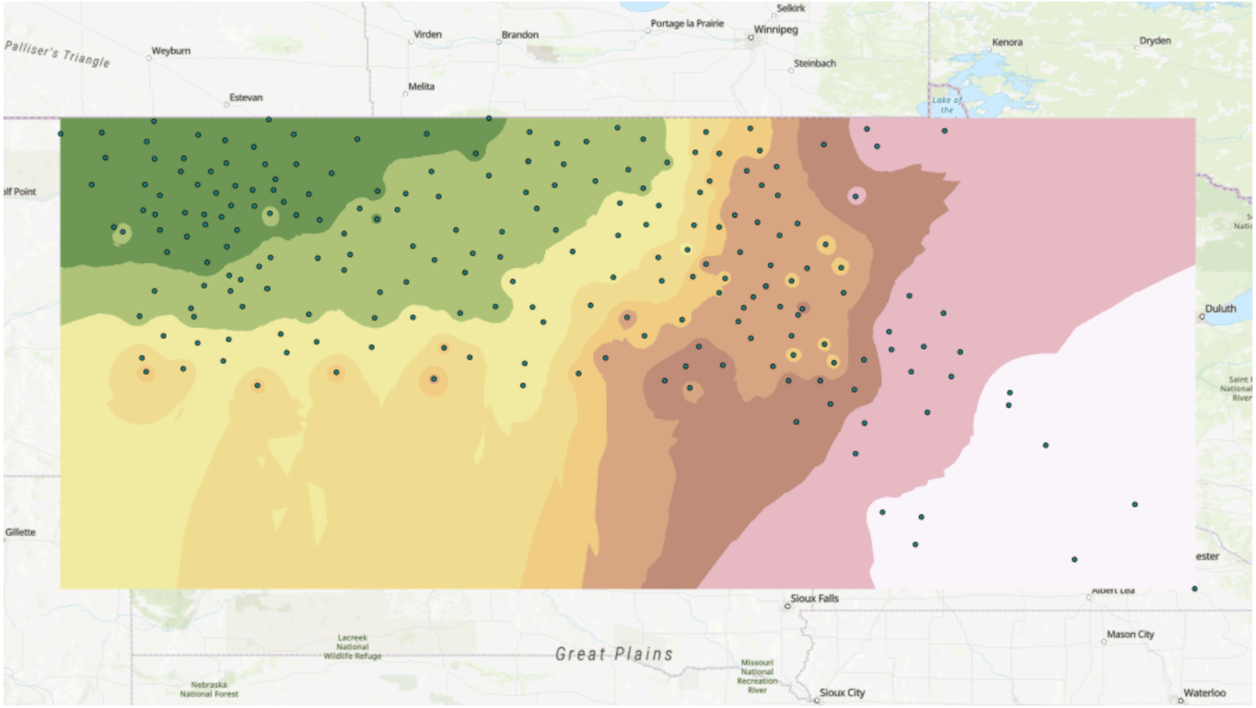

Figure 4. IDW Interpolation of Low Average Temperature

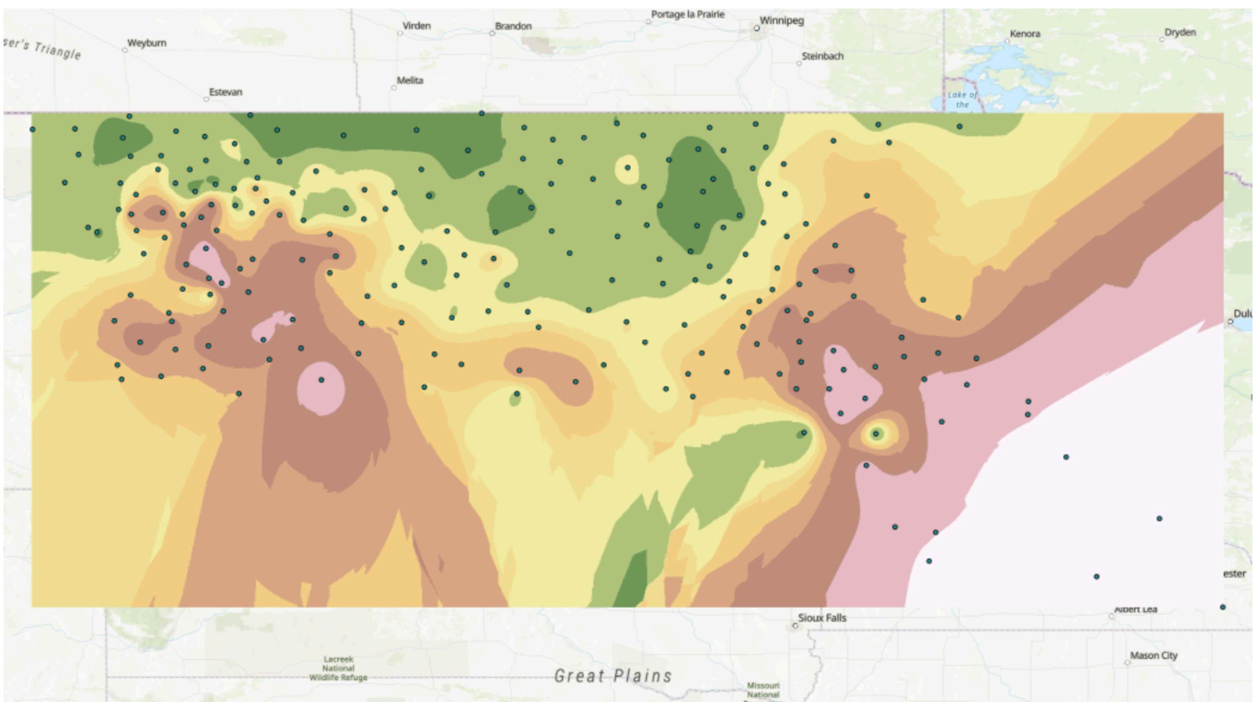

Figure 5. Kriging Interpolation of High Average Temperature

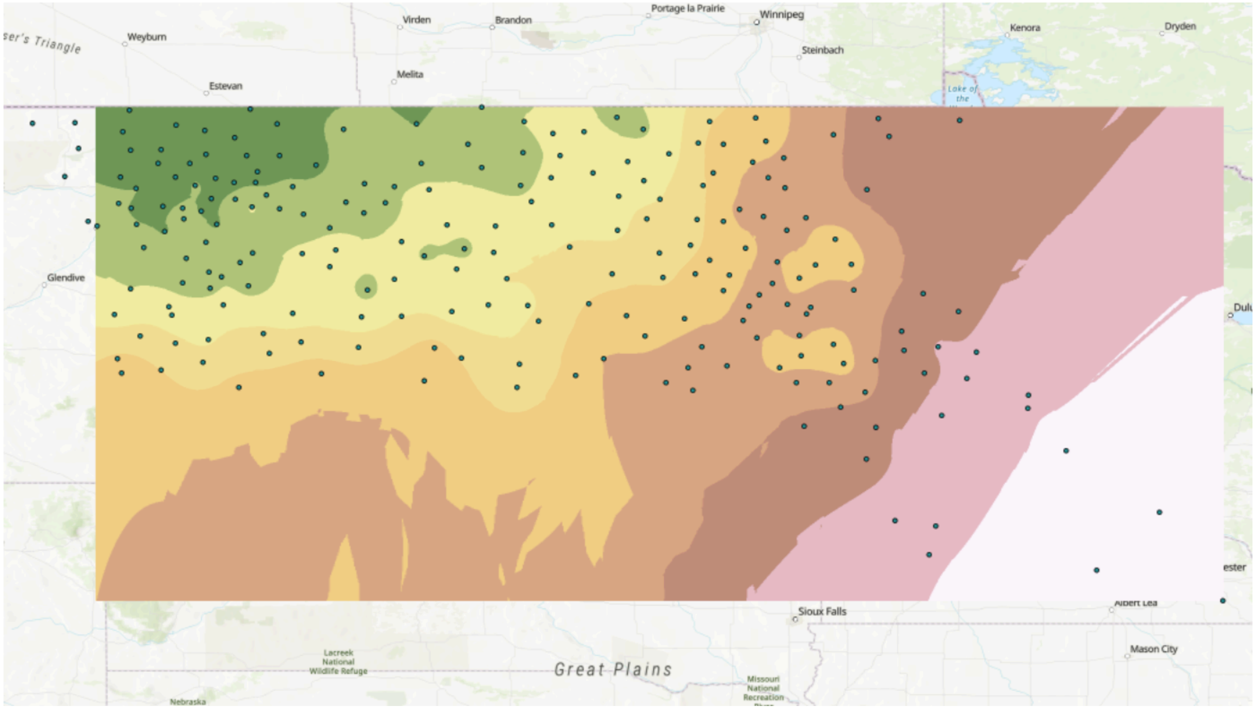

Figure 6. Kriging Interpolation of Low Average Temperature

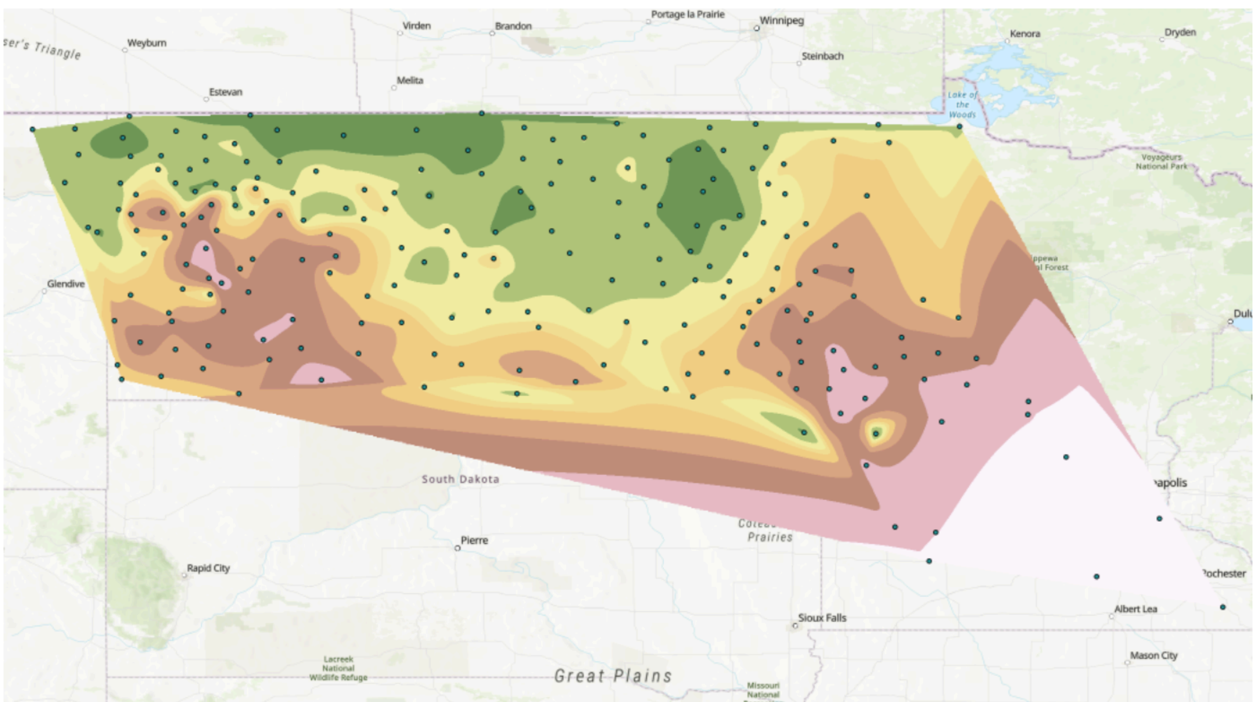

Figure 7. Natural Neighbor Interpolation of High Average Temperature

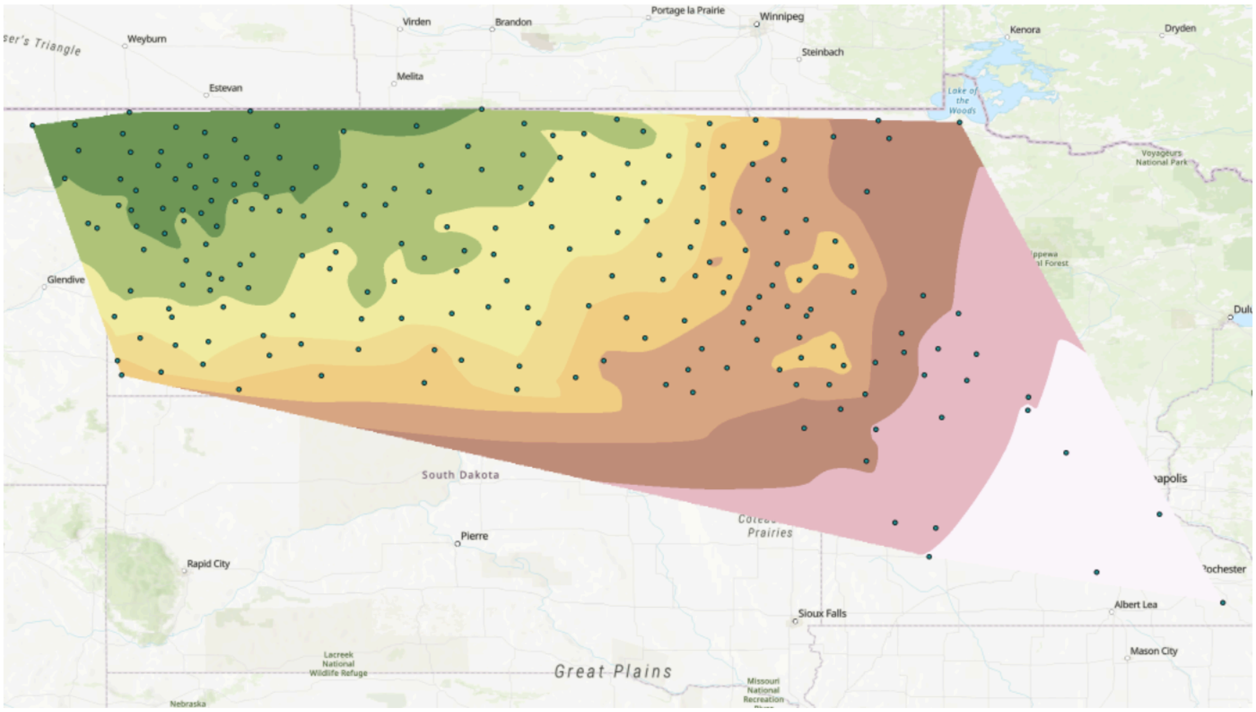

Figure 8. Natural Neighbor Interpolation of Low Average Temperature

Results Verification

To verify the results of my station data, I visually reviewed the inputted temperature data checking for completeness of 30 days of input data for every station, and checked for errors. To verify the results of my interpolation methods, I visually inspected the interpolated rasters for consistency with expected temperature patterns. I looked for significant outliers, glitches, and any obvious signs of an error in the computational analysis.

Discussion and Conclusion

In conclusion, this lab demonstrated the process of building a real-time data visualization and analysis workflow for temperature data from the NDAWN stations. Through the development of an ETL pipeline, I successfully retrieved the most recent 30 days of temperature data and visualized station points with aggregated temperature values. The ETL pipeline I built would successfully retrieve the most recent 30 days of data no matter what day it is used from here on out.

Using this data, I conducted three different interpolation methods—Inverse Distance Weighting (IDW), Kriging, and Natural Neighbor—on both high and low temperature data. Each interpolation method was applied using ArcPy's spatial analyst tools, and the results were compared to evaluate their effectiveness in creating accurate temperature maps. The IDW method is commonly used for temperature data due to its simplicity and effectiveness in capturing local variation, while Kriging offers a more statistically sophisticated approach, ideal for handling spatial autocorrelation. The Natural Neighbor interpolation method, chosen for its simulation approach, provided an alternative to predict many values per location.

ESRI's decision guide for interpolation methods gave me valuable insight for choosing methods. The decision trees showed me that IDW predicts one value per location, while kriging predicts quantile values. Since both of these categories were covered, I decided to use simulation for my third method so I can compare a method which predicts many values per location. The decision guide also showed me that for output type IDW is a prediction type, while kriging is a probability type, and simulation is a full distribution of possible value. The guide also indicated IDW and simulation methods would not have smooth outputs and kriging would be moderately smooth, which I did find to be true considering my smoothest raster output was low temp kriging and the rest were fairly unsmooth.

I researched common techniques for the interpolation of temperature data and found kriging to be widely used, accepted, and recommended. After reading "Mapping Climate Zones of Iran Using Hybrid Interpolation Methods" by Asadi et al. (2022), I found that the authors recommend using kriging for temperature data interpolation because it effectively captures spatial variability, enhancing the accuracy of temperature maps. By integrating kriging with regression models (regression-kriging), their method improves the precision of climate zoning and spatial structure analysis of temperature and precipitation patterns across Iran. This hybrid approach ensures more reliable and detailed climatic mapping. Although I did not employ a regression model alongside my kriging method, I believe this supports my use of kriging in this lab and for future work.

References

Asadi Oskouei, E., Khaki, B. D., Kouzegaran, S., Navidi, M. N., Haghighatd, M., & others. (2022). Mapping climate zones of Iran using hybrid interpolation methods. Remote Sensing, *14* (11), 2632. https://doi.org/10.3390/rs14112632

Esri. (n.d.). *Classification trees of the interpolation methods offered in Geostatistical Analyst*. ArcGIS Desktop. Esri. Retrieved November 26, 2024, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

OpenAI. (2024). ChatGPT (October 8 Version) [Large language model]. Available at https://chat.openai.com

Self-score

Fill out this rubric for yourself and include it in your lab report. The same rubric will be used to generate a grade in proportion to the points assigned in the syllabus to the assignment.

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	**24**

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	**28**
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	**16**
		100	**94**