Lab Report

Title: **Lab 3 – Part 2 – NDAWN Interpolation**
Notice: Dr. Bryan Runck
Author: Kyle Smith
Date: 11/26/2024

Project Repository: kylejsmith4/GIS5571/Lab3
Time Spent: 76 hours (Lab 3)

Abstract

This lab develops an ETL pipeline using temperature data from the North Dakota Agricultural Weather Network (NDAWN) and focuses on reviewing a variety of interpolation methods. Average temperature data from October 2024 at all 216 NDAWN stations was extracted, mapped, and analyzed using three different interpolation methods: Local Polynomial Interpolation (LPI), Inverse Distance Weighted (IDW), and Ordinary Kriging. These three methods were compared using a variety of observations and statistics, and academic literature on interpolation of temperature data was reviewed. While findings showed that Inverse Distance Weighted (IDW) was the most effective interpolation method for this project, it is generally agreed that a form of Kriging is optimal for unevenly distributed data points, such as weather stations.

Problem Statement

This lab uses average temperature data from October 2024 at all 216 NDAWN stations in a ETL pipeline. As shown in the chart below, several analytical steps were required using ArcPy / Python and ArcGIS Pro. Ultimately, this project asks which selected interpolation method is optimal for this dataset, and what does academic literature generally say about interpolating temperature data?

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Extract data to .csv file	Call NDAWN API	NDAWN stations and temperature data	Location and temperature data	NDAWN Dataset	Prepare and clean CSV file
2	Create map of all stations	Point Data	X Y spatial data	Station name, Average min & max temp		XY Table to Point Feature Class, and to map
3	Create interpolation surfaces	Min & max temperature maps	Rasters of each interpolation method for avg. min and avg. max temperatures	Location and temperature data		IDW, Ordinary Kriging, LPI
4	Data validation to review each interpolation method	Validation and error calculations	Point data	True (measured data from NDAWN) vs Predicted temperature data for each interpolation method		Calculated error data to points

Input Data

#	Title	Purpose in Analysis	Link to Source
1	NDAWN station location and average temperature data for all 216 station sites for October 2024 (31-day period)	Location data for mapping of stations, and average min and max temperatures for each site for interpolation and analysis.	User created API: NDAWN Dataset

Methods
As shown in the data flow diagram which follows, the ETL pipeline used here made a call to the user defined NDAWN API, extracted data to a csv, mapped station locations via a point feature class, and then performed interpolation using ArcPy and ArcGIS Pro Geoprocessing Tools. Each of the three chosen interpolation methods can be defined by the following:

- **Inverse distance weighted** (IDW) interpolation explicitly assumes that things that are close to one another are more alike than those that are farther apart.

- **Local Polynomial Interpolation** (LPI) fits many polynomials, each within specified overlapping neighborhoods.

- **Ordinary kriging** assumes the model $Z(s) = \mu + \varepsilon(s)$, where μ is an unknown constant. One of the main issues concerning ordinary kriging is whether the assumption of a constant mean is reasonable.

Additionally, the three major statistics used to analyze interpolation effectiveness here can be defined:

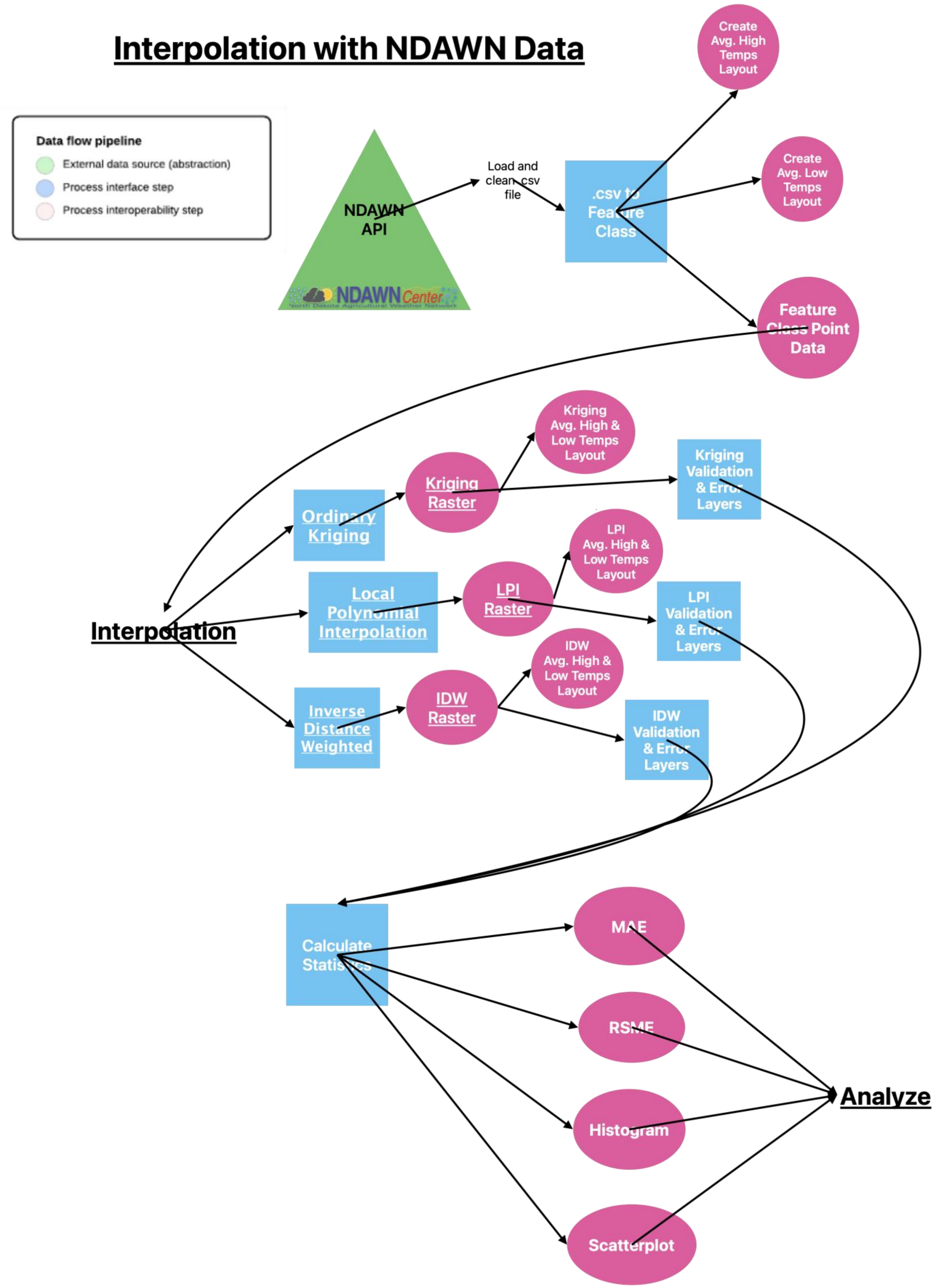
- **Mean Absolute Error** (MAE) measures the average of the absolute differences between observed values and predicted values. A lower MAE indicates better model accuracy, as errors are smaller on average.

- **Root Mean Square Error** (RMSE) measure the average of squared differences between observed and predicted values. It is more sensitive to outliers than MAE. A lower RMSE indicates better model performance.

- **Coefficient of Determination (**R^2) is a statistical measure that indicates how well the predicted values explain the variation in the observed data. It ranges from 0 to 1, with 1 representing a perfect fit.

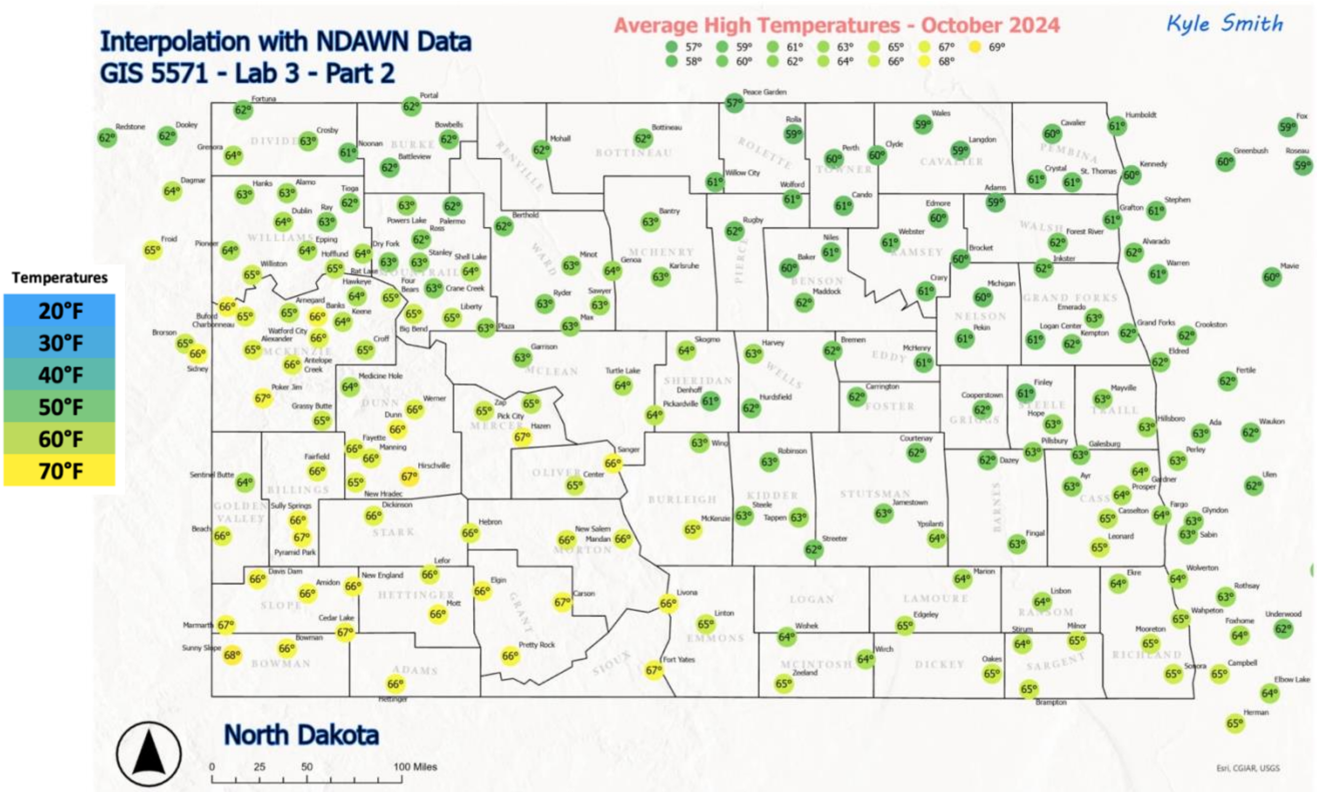
Finally, histogram and scatterplot diagrams were prepared for analysis.

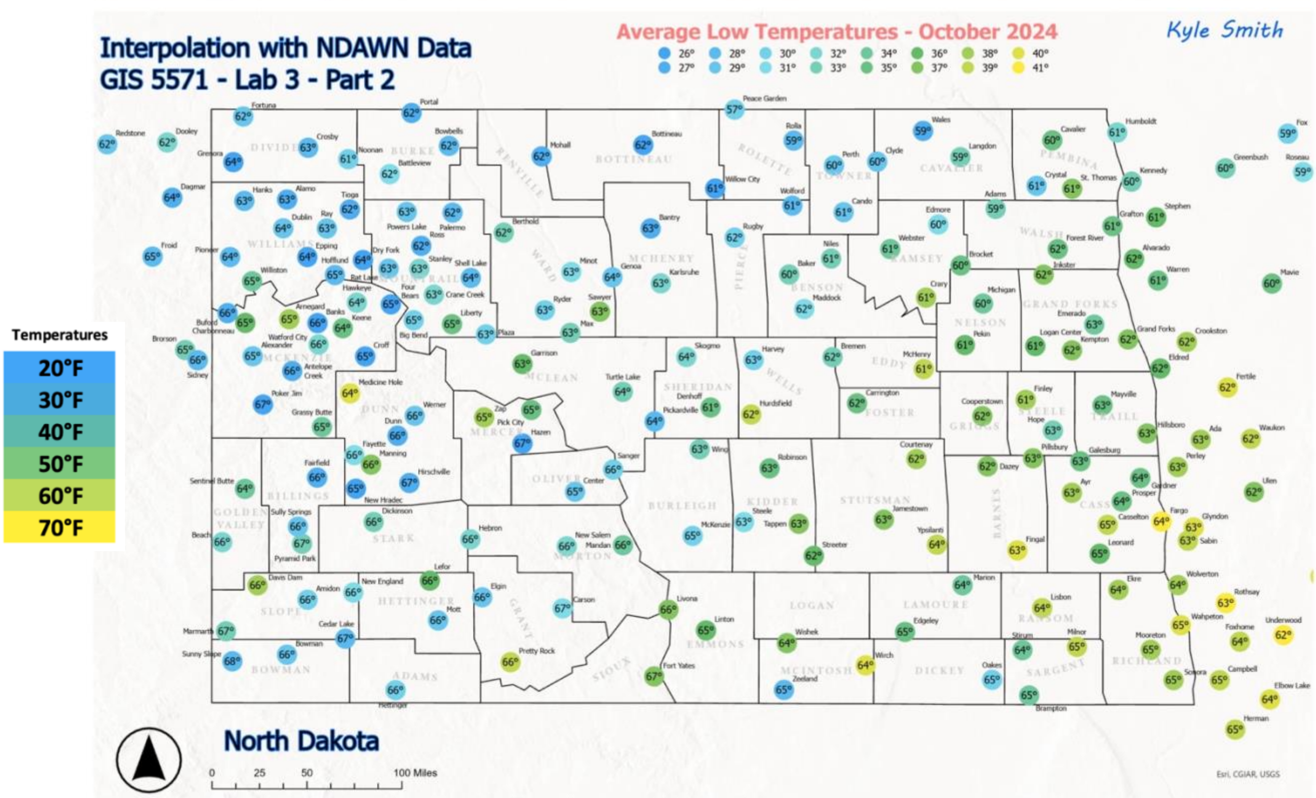
Interpolation with NDAWN Data

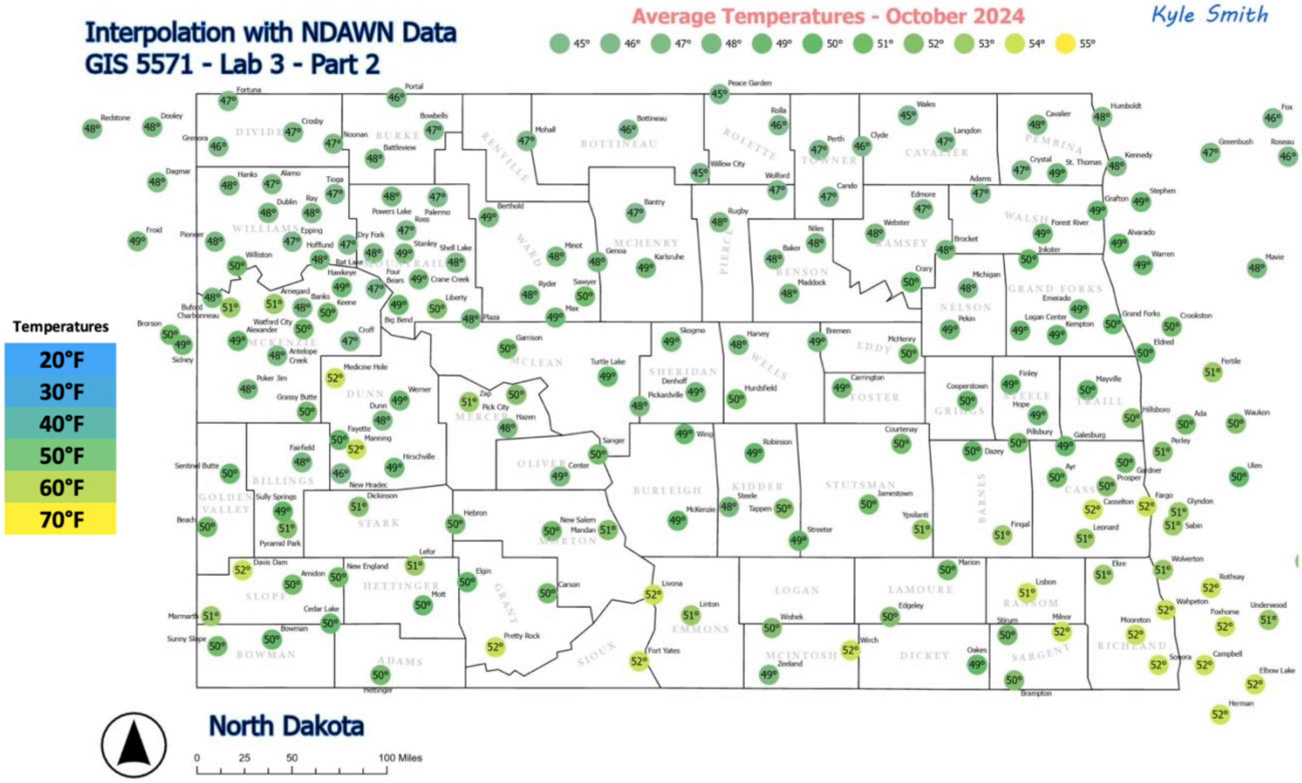


Results

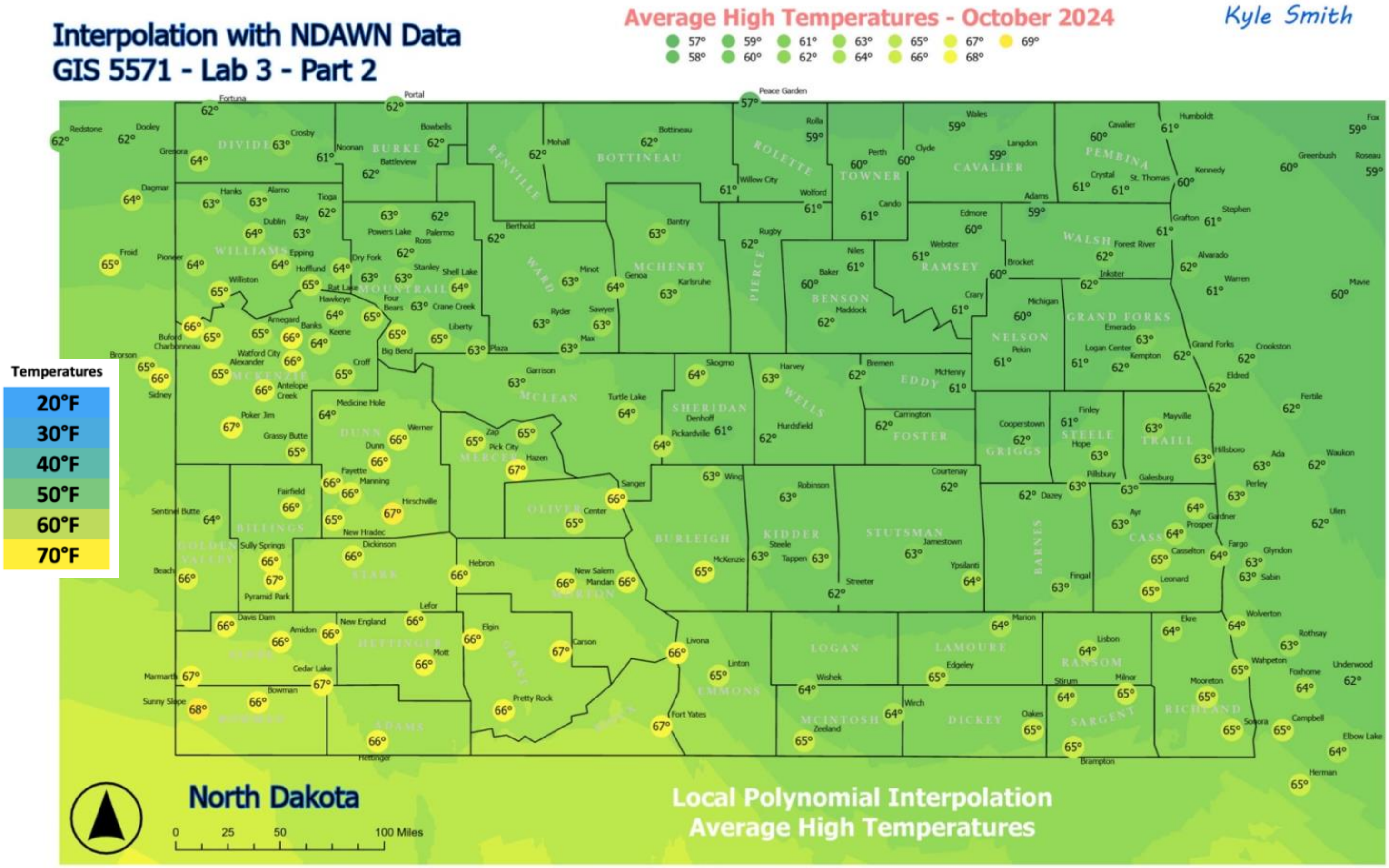
<u>Maps of Average High, Low, and all Temperatures for each NDAWN station site for October 2024</u>

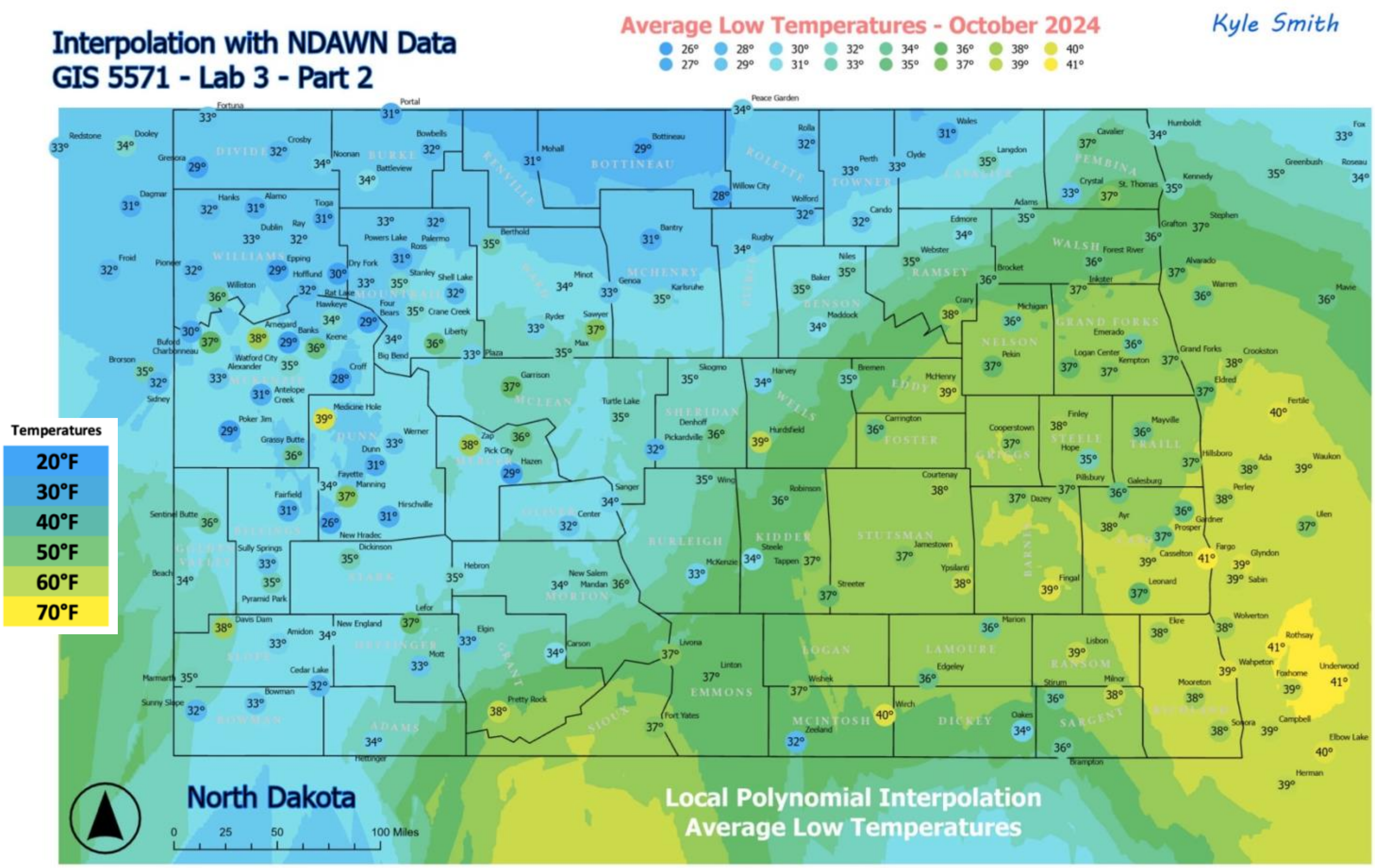




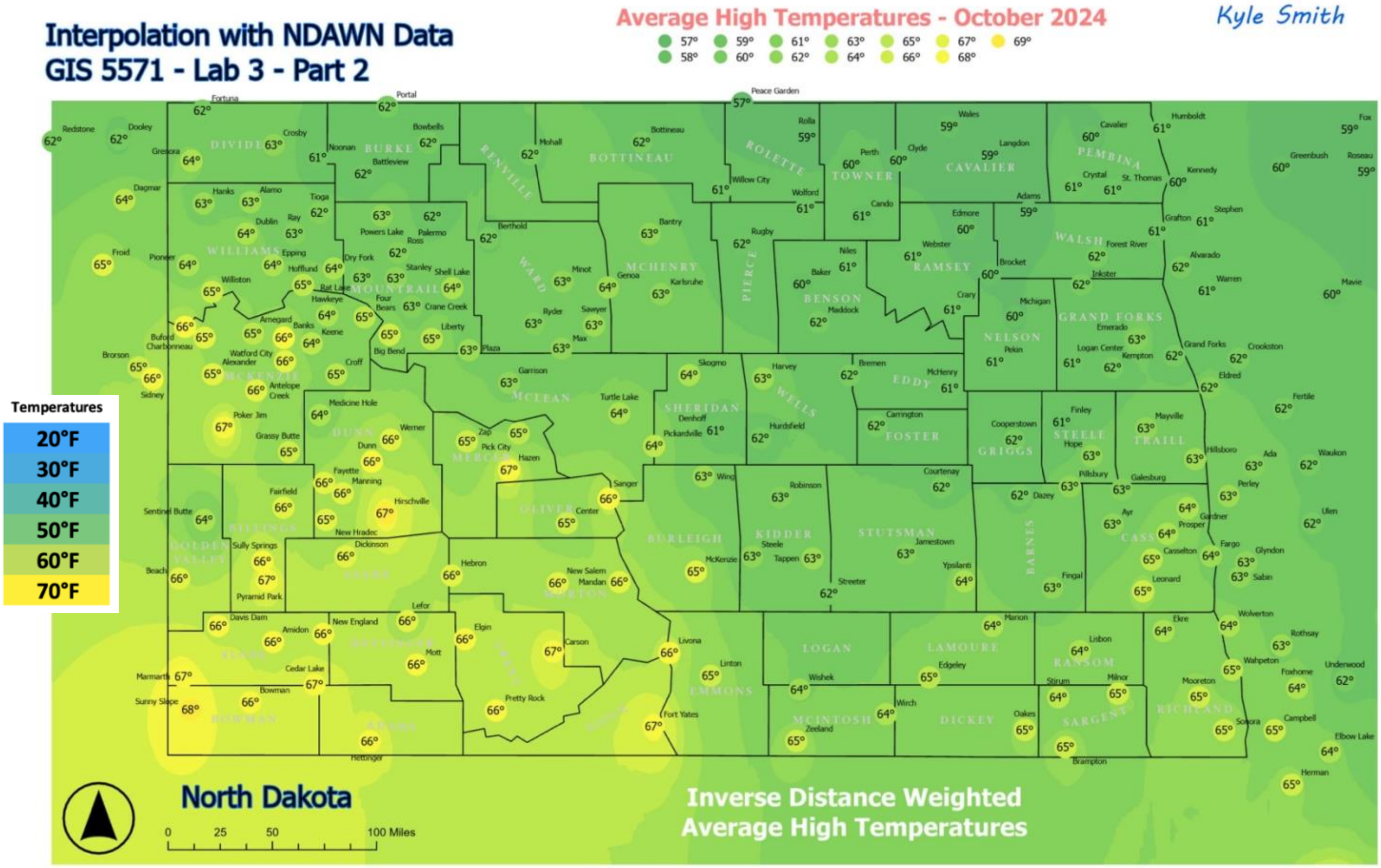

Interpolation with NDAWN Data
GIS 5571 - Lab 3 - Part 2
Average Temperatures - October 2024
Kyle Smith
45° 46° 47° 48° 49° 50° 51° 52° 53° 54° 55°
Temperatures
20°F
30°F
40°F
50°F
60°F
70°F
North Dakota
0 25 50 100 Miles

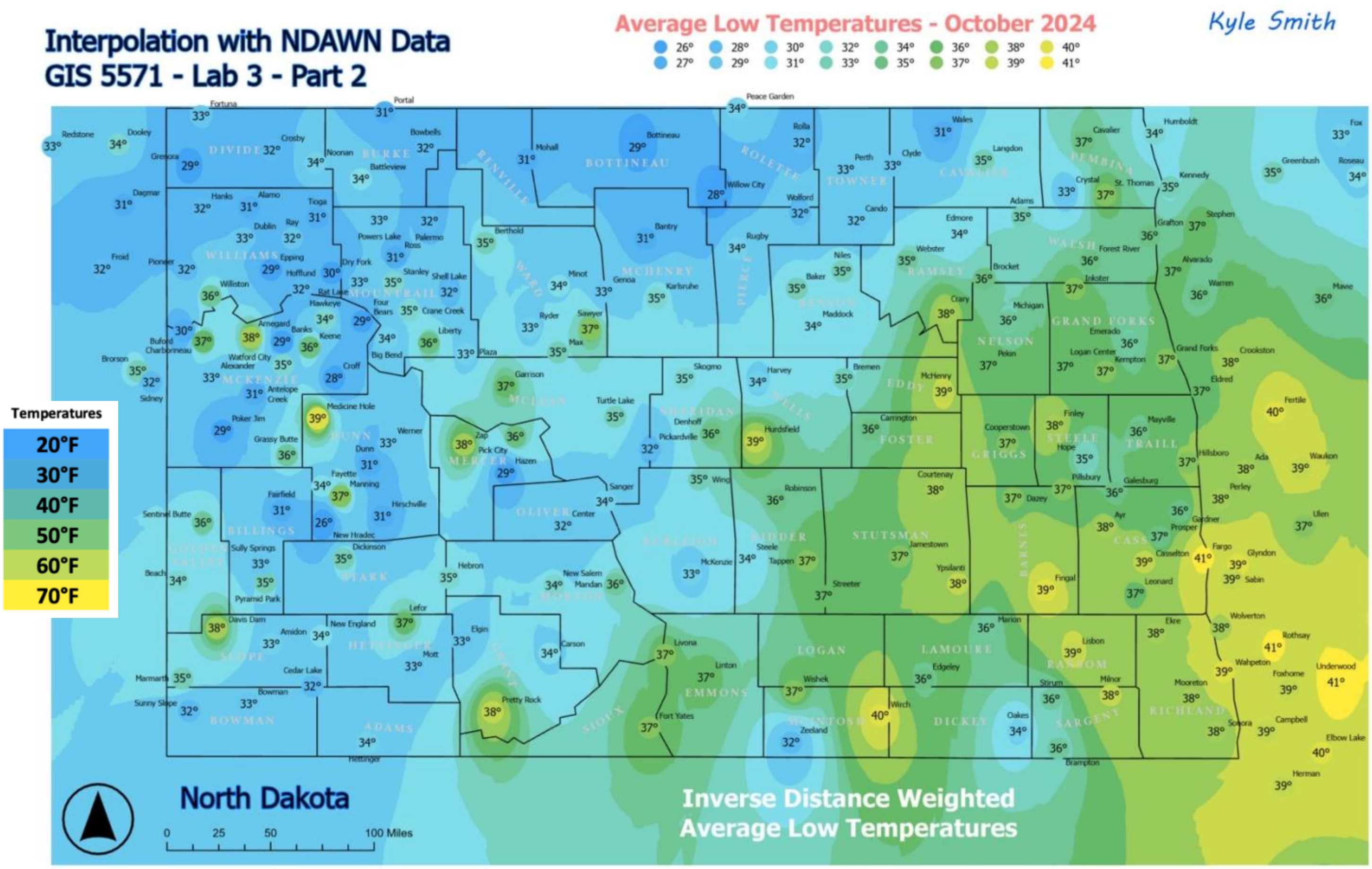
Avg High & Low temps using Local Polynomial Interpolation (LPI)



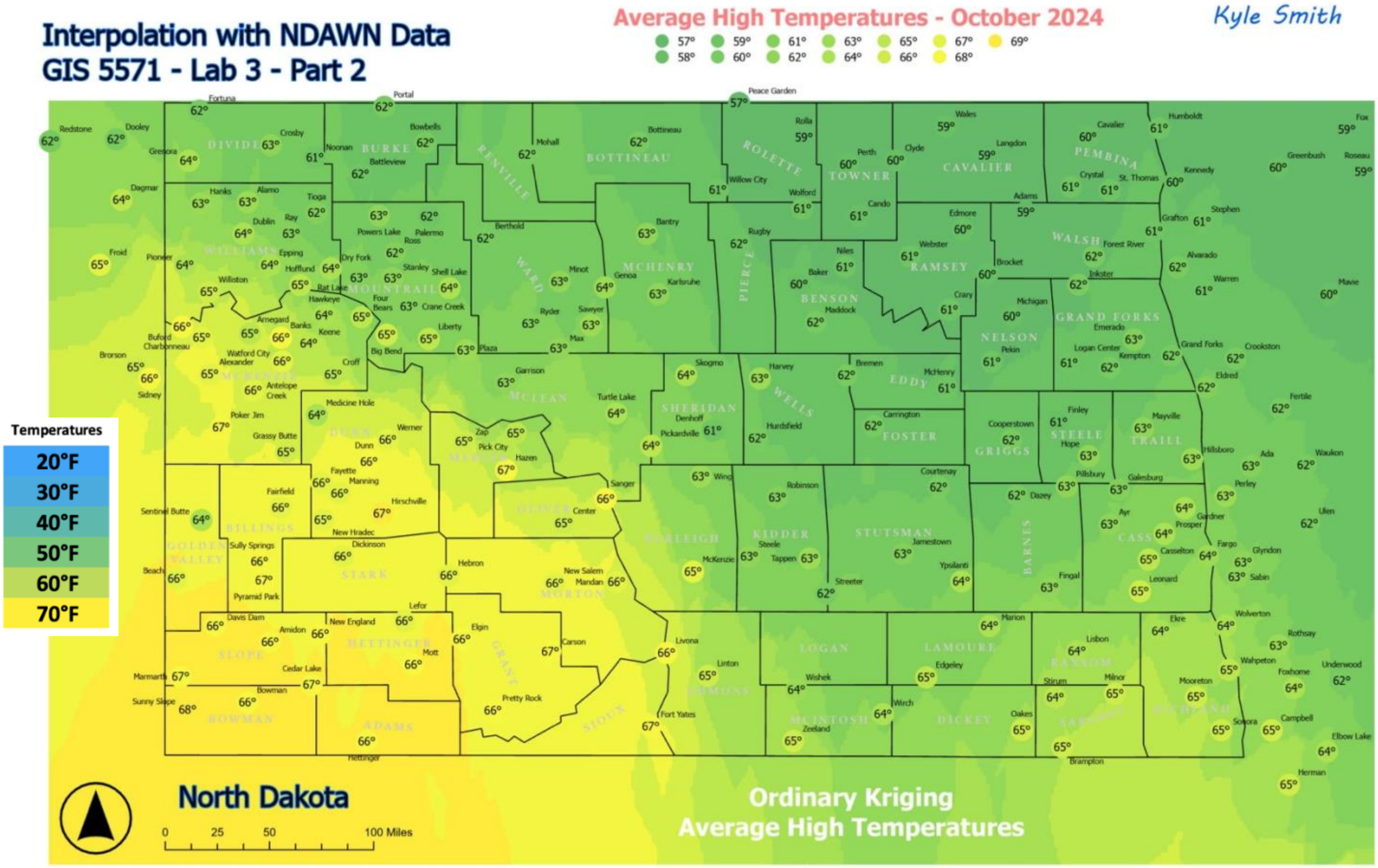


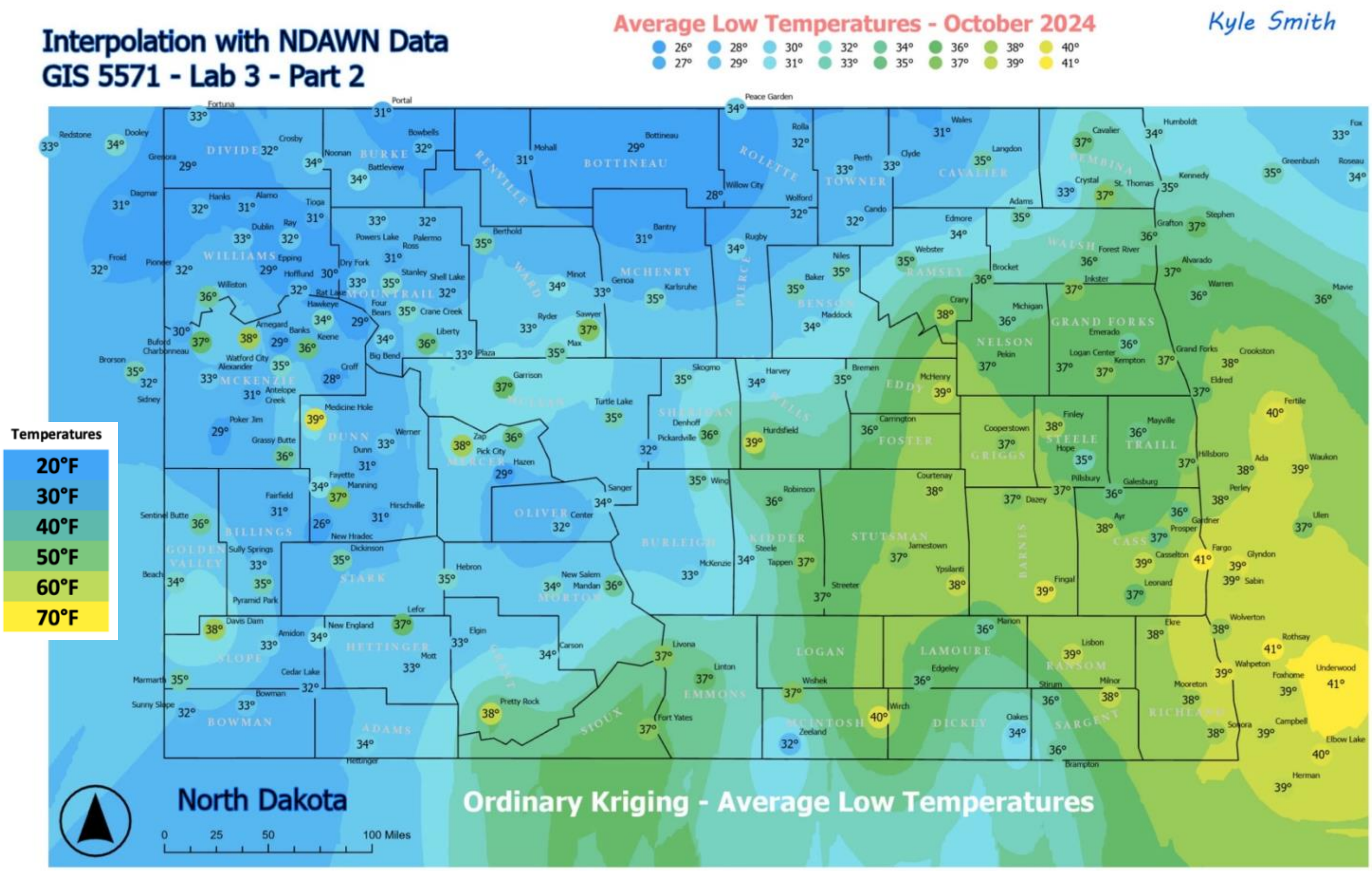
Avg High & Low temps using Inverse Distance Weighted (IDW)



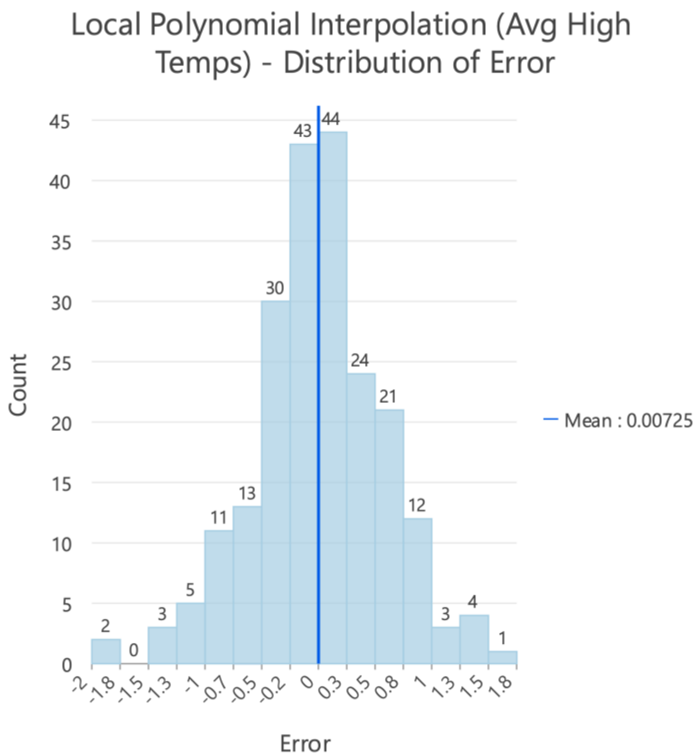


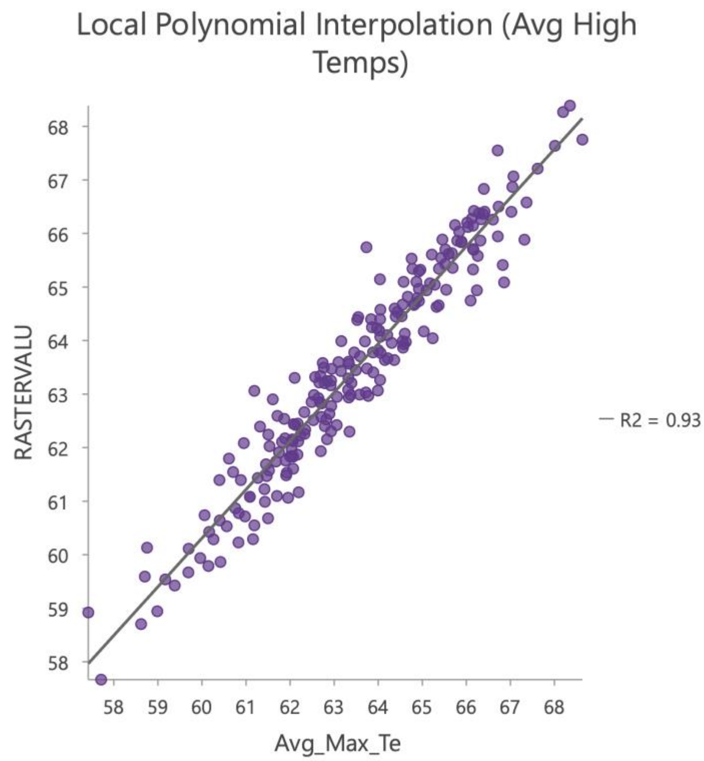
Avg High & Low temps using Ordinary Kriging



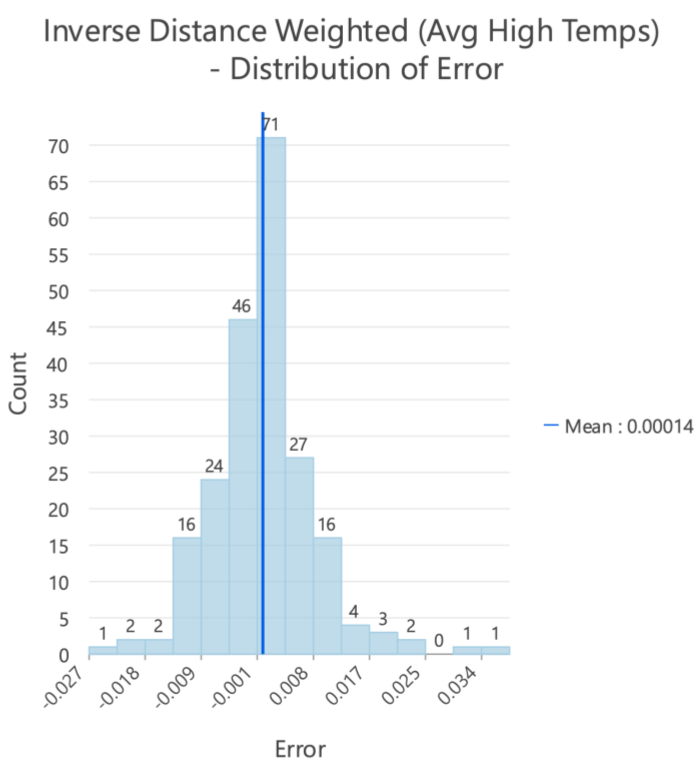


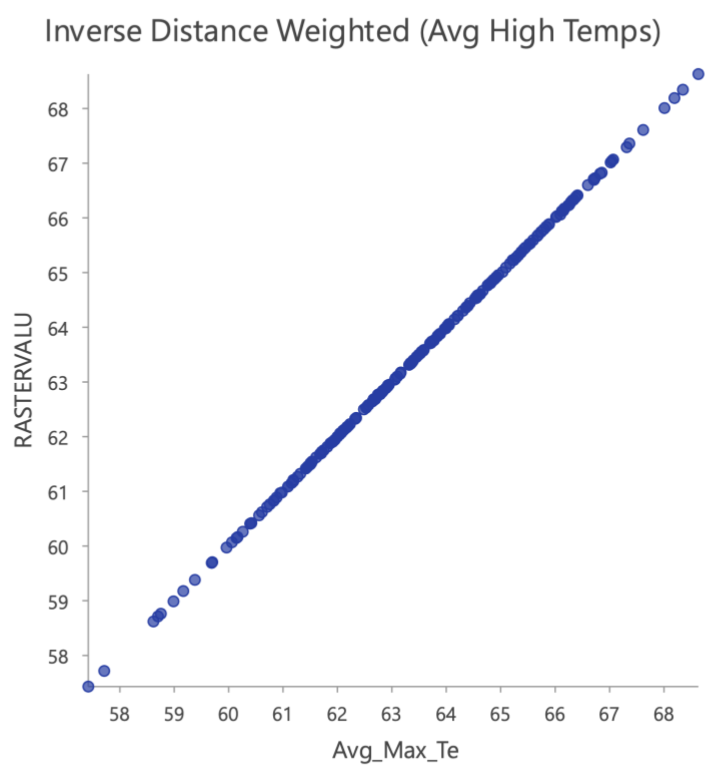
Local Polynomial Interpolation (LPI) – Error distribution histogram & scatterplot



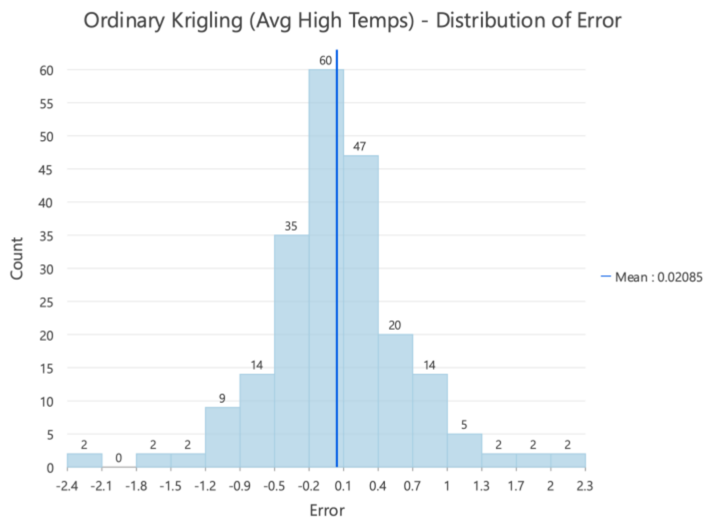


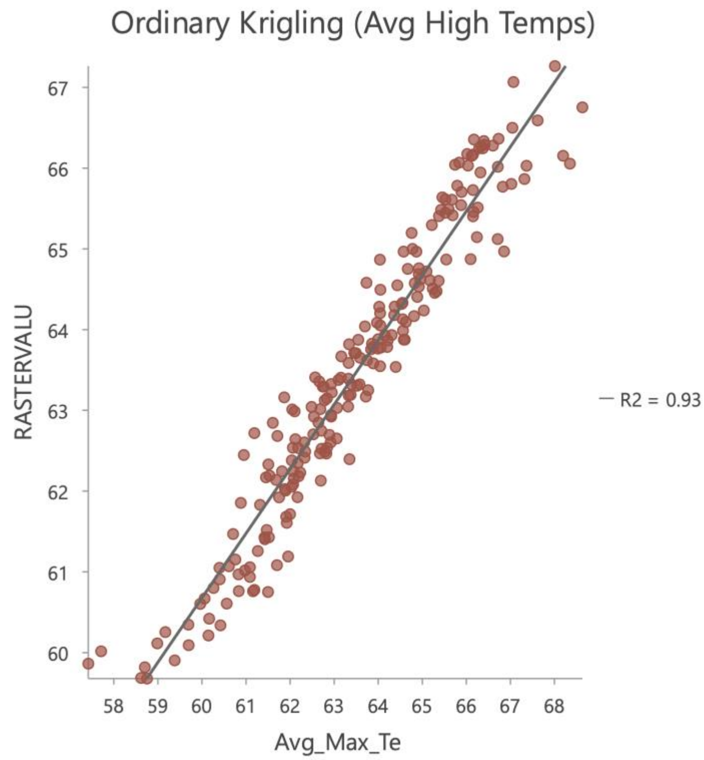
Inverse Distance Weighted (IDW) – Error distribution histogram & scatterplot





Ordinary Kriging– Error distribution histogram & scatterplot





Visual Observations

The average low temperature maps for October 2024 show warmer low temperatures in the southern areas of North Dakota, with cooler areas in the north. Low temperatures range from about 26 degrees to 42 degrees. The average high temperature maps show a similar gradient with warmer temperatures in southern areas and cooler temperatures in the north. High temperatures range from 57 degrees to 68 degrees. The IDW maps generally shows sharper transitions between temperatures, the Kriging maps show smoother transitions or gradients, and the Local Polynomial maps seems to be a middle ground.

Statistics Summary

Interpolation Method	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Coefficient of Determination (R2)
IDW	0.005431991	0.00795386	0.999988706
Ordinary Kriging	0.455569329	0.639612802	0.933042778
LPI	0.431978412	0.578990421	0.929576847

Based on the above statistics and charts, IDW produced the most accurate results, closely matching observed values. Ordinary Kriging and LPI provided smoother surfaces but less accuracy.

Results Verification

This lab involves analyzing three interpolation models of average temperature distributions at set stations across an entire state. The interpolation process is both a method to predict point data over a larger contiguous plane and a verification of results by calculating errors between actual and calculated values. While this project finds that Inverse Distance Weighted (IDW) was the most effective interpolation method, it is generally agreed that a form of Kriging is optimal for unevenly distributed data points. This is discussed in greater detail in the section below.

Discussion and Conclusion

This lab required the use of IDW interpolation, plus a type of Kriging, and one more. In making the decisions to choose the interpolations used here, I consulted the ESRI article "Classification trees of the interpolation methods offered in Geostatistical Analyst" found here. Reviewing this article was helpful in choosing interpolation methods for temperature data over a large region.

Using these decision trees, we can determine our model of temperature data needs the following:

- Predicted values and errors
- One value per location
- Prediction and prediction error values
- Complex level of assumptions
- Exact interpolation
- Smooth output
- Uncertainty values of predicted values provided
- Intermediate processing speed

Based on those criteria, **Inverse Distance Weighting** (IDW), **Ordinary Kriging**, and **Local Polynomial Interpolation** (LPI) are selected for this lab. Temperature interpolation requires predicted temperature values and errors at each location, an understanding of complex assumptions, and smooth understandable output. IDW can provide more exact predictions, but the output can lack smoothness, which is key to understanding a temperature map. Ordinary Kriging is accurate and can provide smooth outputs, however the uncertainty can be greater. LPI is a relative balance between smoothness, processing speed, and complexity. Academic literature tends to show that **Kriging methods are recommended for interpolation of temperature data**. Kriging can account for spatial autocorrelation and address areas with a void of true data. Regression-based techniques, such as Kriging, can address both spatial and physical factors influencing temperature distribution, such as terrain impacting the prediction of temperatures while maintaining a higher accuracy over other methods.

References

Esri. (n.d.). *Classification trees of the interpolation methods offered in Geostatistical Analyst*. Retrieved November 25, 2024, from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm

Wang, L., Che, M., & Li, J. (2014). Spatial interpolation and temperature information visualization. In Park, J., Pan, Y., Kim, C. S., & Yang, Y. (Eds.), *Future information technology* (Lecture Notes in Electrical Engineering, Vol. 309). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-55038-6_28

Hengl, T., Heuvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma, 120*(1–2), 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018

North Dakota Agricultural Weather Network (NDAWN). (n.d.). *North Dakota Agricultural Weather Network*. Retrieved November 25, 2024, from https://ndawn.ndsu.nodak.edu/

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	25
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	94