Lab Report

Title: Lab 3
Notice: Dr. Bryan Runck
Author: Yaxuan Zhang
Date: 11/29/2023

Project Repository: https://github.com/YaxuanSeanZhang/MGIS_ARCGIS/tree/main/GIS%205571/Lab3
Google Drive Link:
Time Spent: 8 hrs

Abstract

stands for Extract, Transform, Load, and it is a crucial process in data integration and data warehousing. ETL is used to gather, process, and transfer data from various sources to a data warehouse or other target systems for analysis and reporting. In this lab, we will build ETL pipeline to extract data from different APIs for different data types. By comparing different ETL workflows, we will gain a deeper understanding of ETL process.

Problem Statement

In this lab, we will go through the ETL process for different data types, e.g., shp, geojson, and csv. Through practicing decomposing interfaces for spatial web API's into informal conceptual models, we can compare contract different web API's using informal conceptual models and custom-built ETL routines. We will build an ETL pipeline with ArcPro Jupyter Notebook and integrate two datasets via spatial join.

Table 1. Main Steps

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Extract Data from API	Raw input dataset pulling from different API (i.e., NDAWN)	Point	Avg temperature, Min/Max temperature	NDAWN	Define API, and pull data
2	CSV to Shp	shapefile	Point	Various	NDAWN	Convert csv to shapefile and then transform the coordinate
3	Interpolation	Three interpolation methods (i.e., idw, kriging, spline)	Point to raster	Various	NDAWN	

Input Data

We will use data from NDAWN. Data from NDAWN are csv tables. We need to customize the API to define the range of the data, and then pull data via customized API. To obtain real-time data, we can define the date by using datetime package in python.

Table 2. Required Dataset

#	Title	Purpose in Analysis	Link to Source
1	Average Weather	Daily average weather of all stations	NDAWN
2	Minimum Weather	30-day minimum weather of all stations	NDAWN
3	Maximum Weather	30-day maximum weather of all stations	NDAWN

Methods

IDW interpolation methods

Figure 1 ETL and Spatial Join Workflow

Results

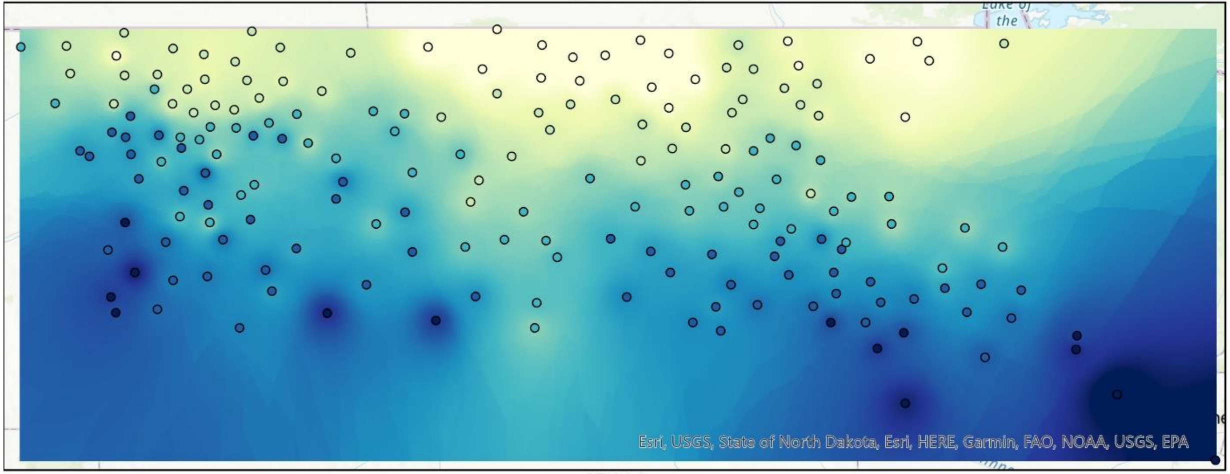
The results are the comparison among three interpolation methods for average, minimum, and maximum temperatures:

IDW: The assigned values to unknown points are calculated with a weighted average of the values available at the known points.

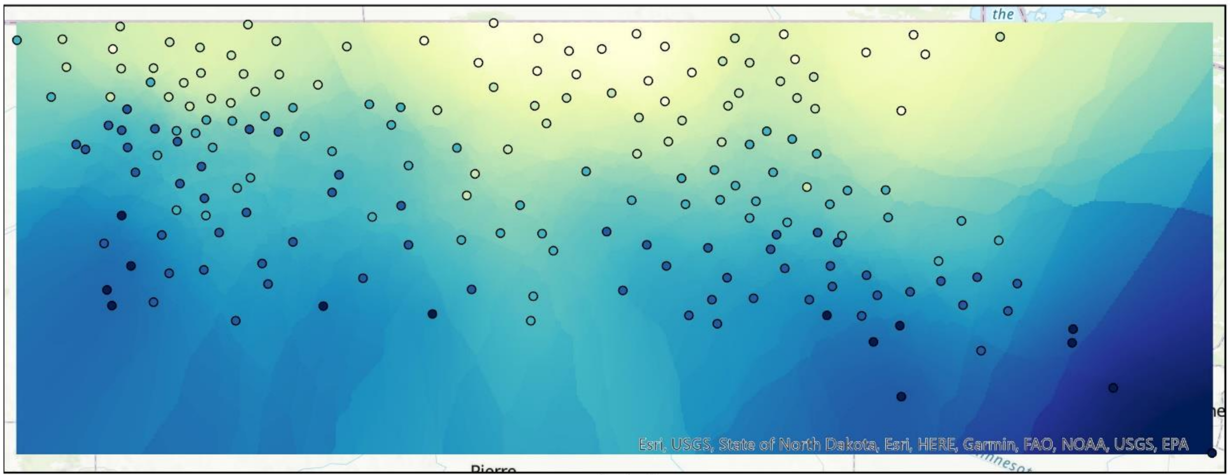
Kriging: Kriging is a geostatistical interpolation method that estimates values at unmeasured locations by incorporating spatial correlation information from observed data points, providing both predictions and uncertainty assessments.

Spline: Spline interpolation involves fitting a piecewise-defined polynomial function to the data points. Cubic splines, in particular, are commonly used for their smoothness and ability to capture more complex variations in the data.

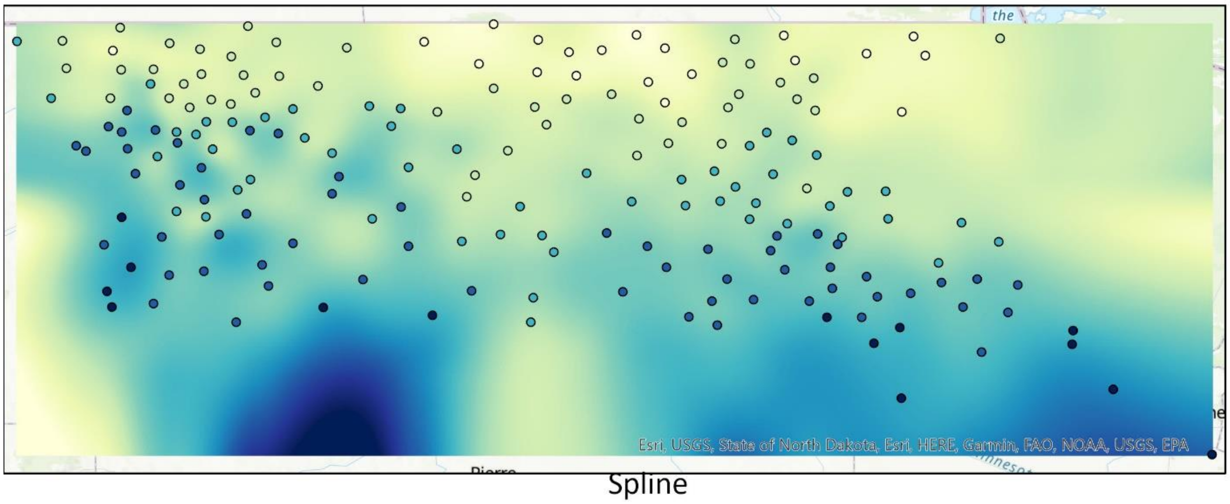
Average Temperature


Lake of
the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA

IDW

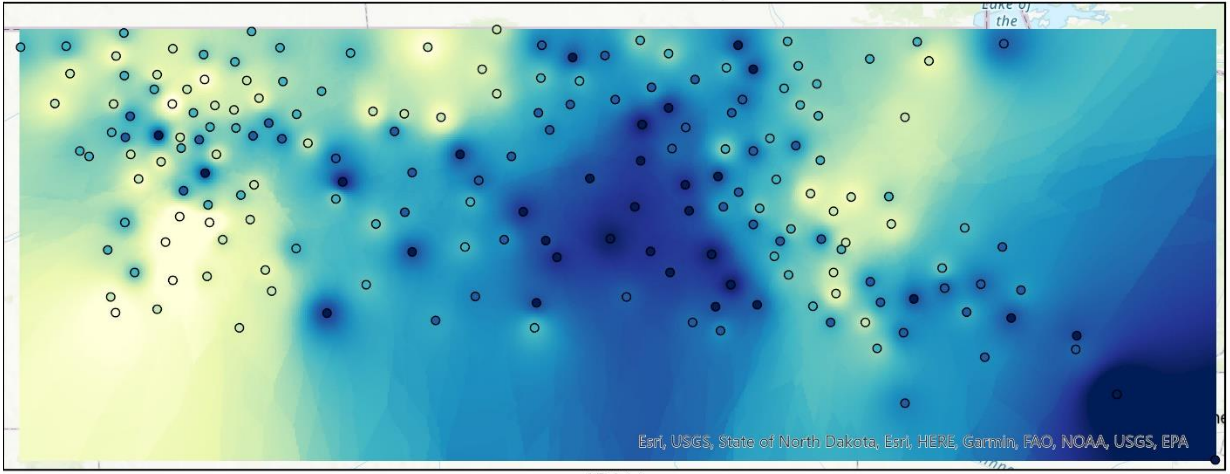

the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA
Pierre

kriging

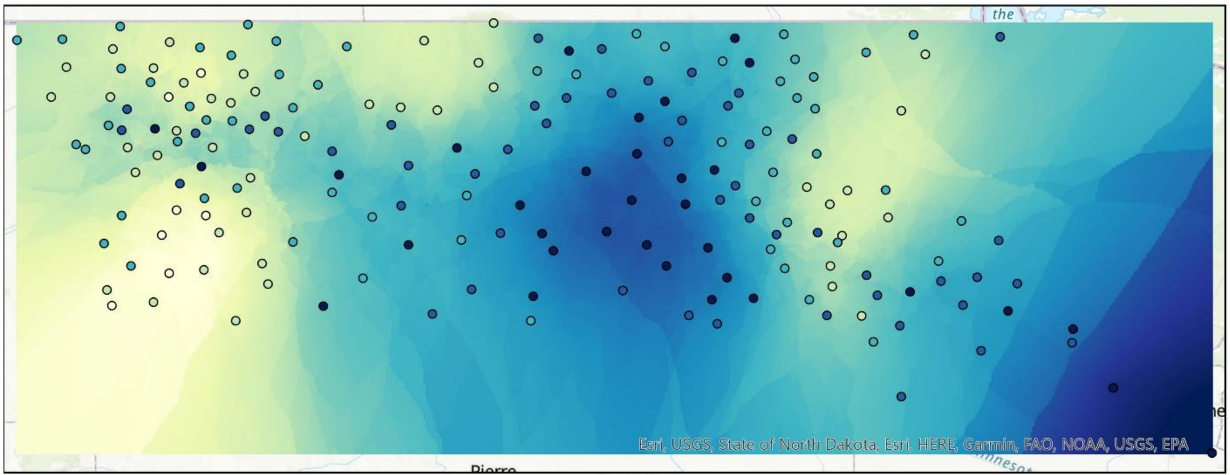

the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA
Pierre

Spline

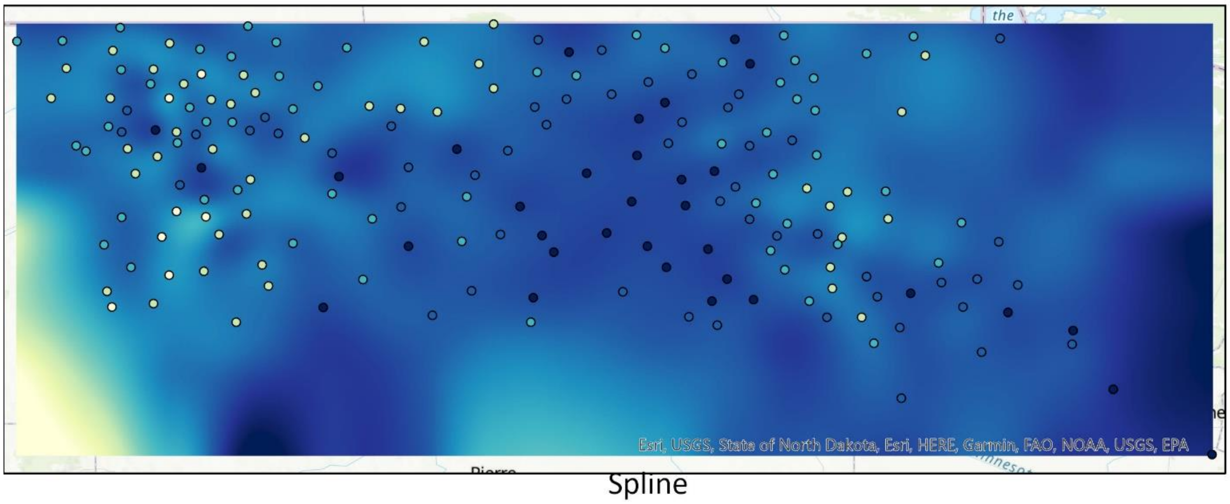
Maximum Temperature


Lake of the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA

IDW

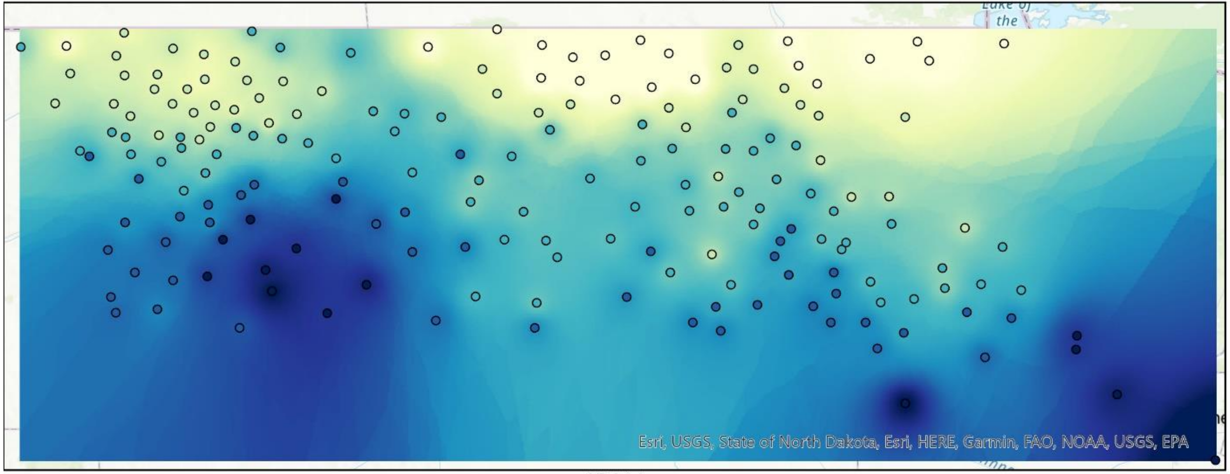

the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA
Pierre

kriging

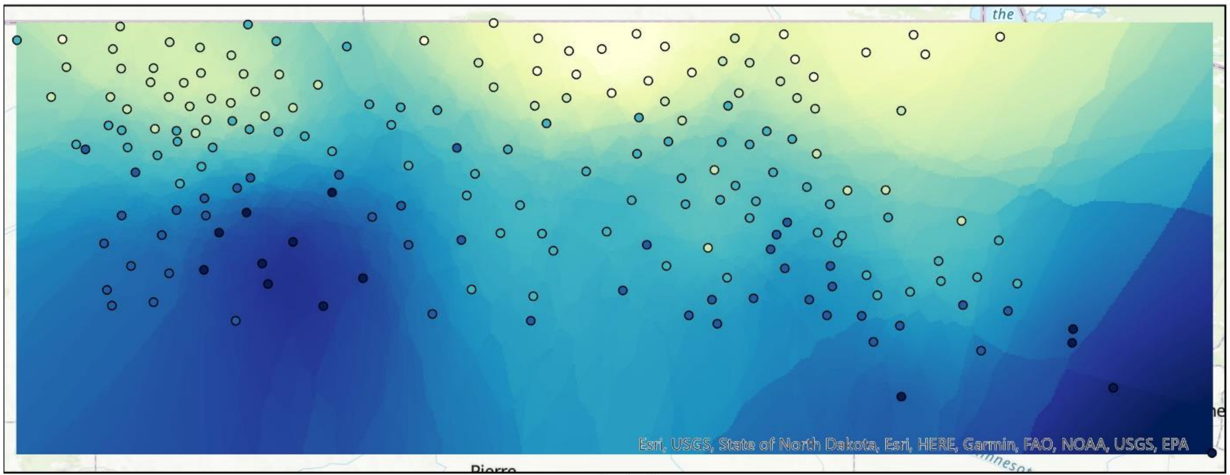

the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA
Pierre

Spline

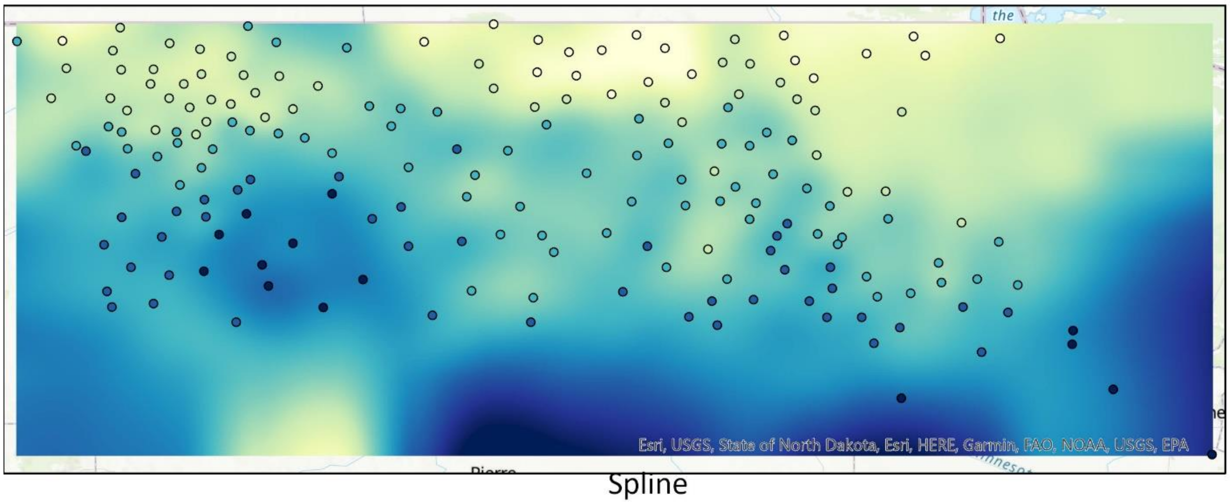
Maximum Temperature


Lake of the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA

IDW


the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA
Pierre

kriging


the
Esri, USGS, State of North Dakota, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA
Pierre

Spline

Results Verification

I made maps to show both true values (i.e., station temperature) and interpolated raster layer.

Discussion and Conclusion

To interpolate temperature, Cao et al. found that Kriging-exponential and Kriging-spherical interpolation methods are the highest-accuracy methods, inverse distance weight method is less accurate, and Kriging-Gaussian and Spline interpolation methods have the lowest accuracy.

Cao, W., Hu, J., & Yu, X. (2009, August). A study on temperature interpolation methods based on GIS. *In 2009 17th International Conference on Geoinformatics (pp. 1-5)*. IEEE.

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	**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	**20**
		100	**100**