SAINT LOUIS UNIVERSITY.

Spatial Analysis as a Process

Sourav Bhadra, Ph. D.

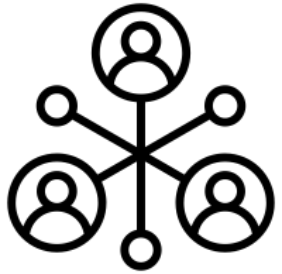
GIS4120/5120 Geospatial Analytics
9/3/2024

Pragmatic decisions are often based on a series of common questions

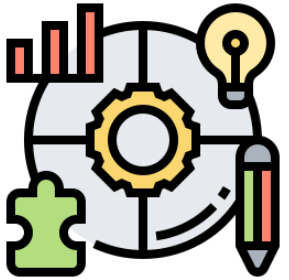
How well defined is the problem I am seeking to address?

How much time, money and resources can I afford to apply to this problem?

What research has previously been carried out on problems of this type and what strengths and weakness?



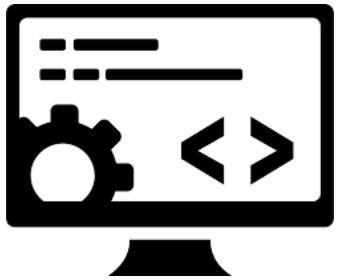
Who will be the recipient of the results and what are their expectations and requirements?



What implications does my examination have on the selection of techniques, data and what caveats?

Pragmatic decisions are often based on a series of common questions

How will I deal with data inadequacies?

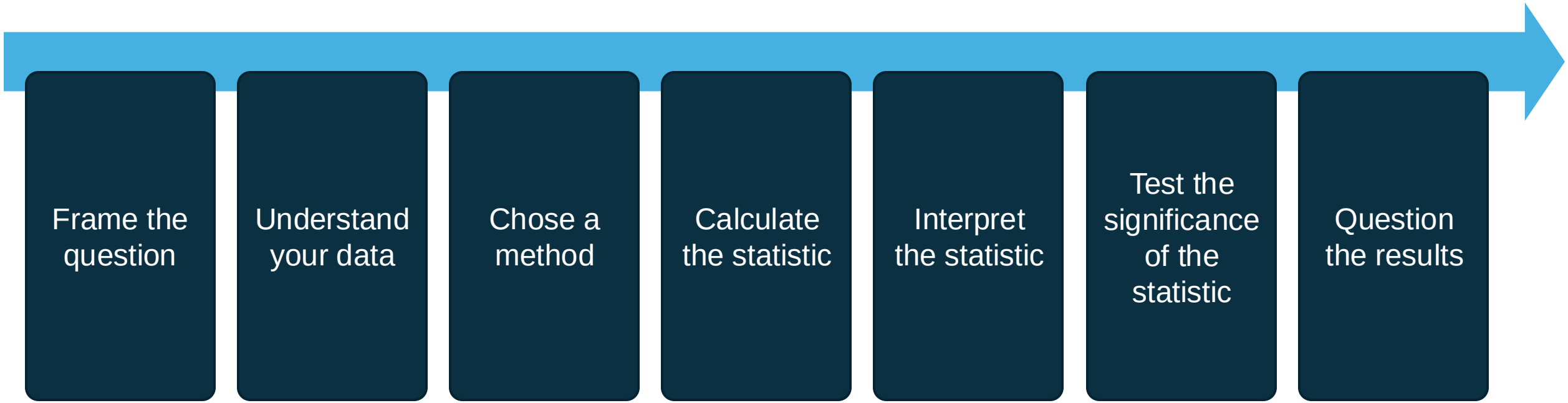


How will I deal with limitations and errors in the software I have chosen?

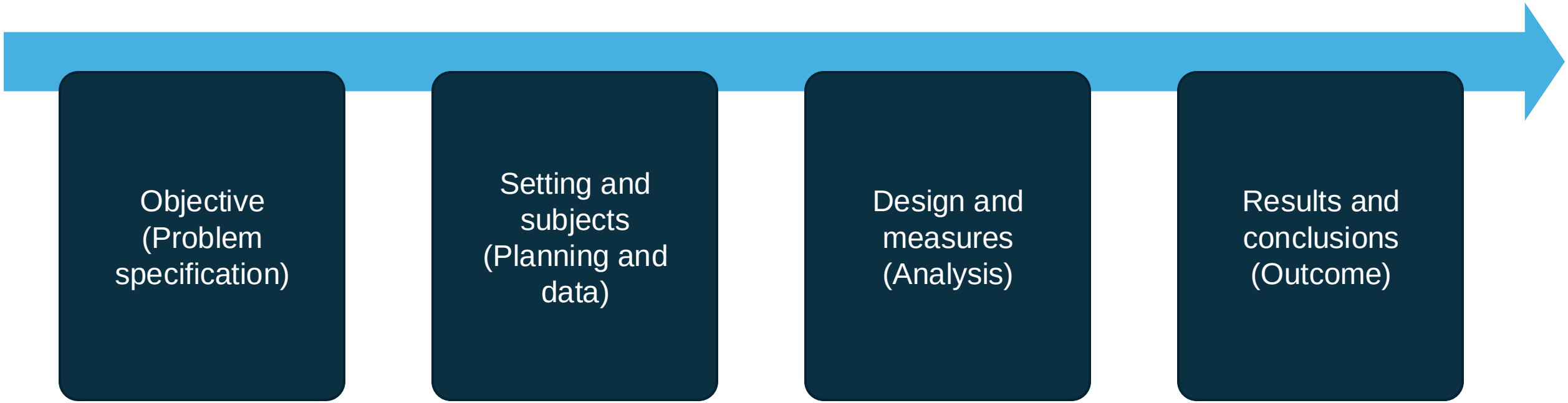
What are the implications of producing wrong or misleading results?

Are there independent and verifiable means for validating results obtained?

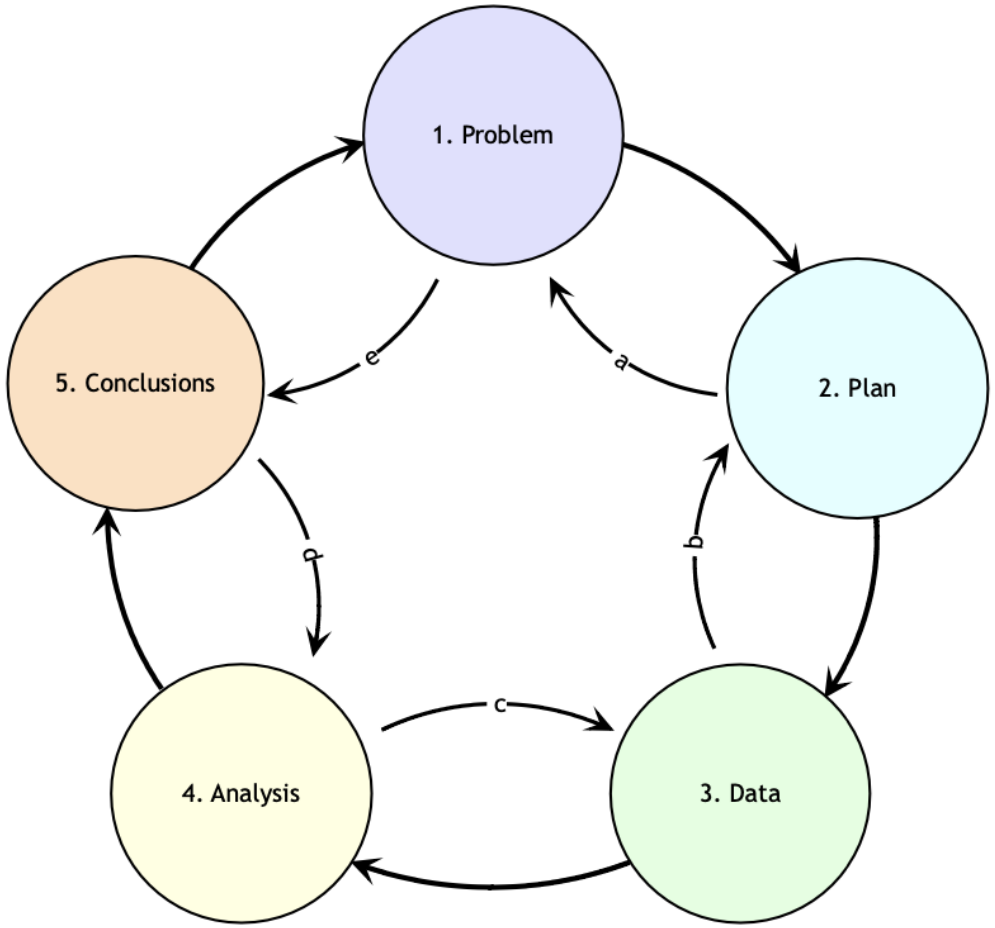
Mitchell's analytical process



Draper’s analytical process



Mackay and Oldford's PPDAC Model



- Geospatial analysis process as an iterative framework instead of a linear process
- The iteration can occur at any stage of the model
- Changes in one stage can influence other stages
- You can also start from anticlockwise (but be careful about your inferences)

Problem: Framing the question

- **Spatial Scale**
- Statistical Scale
- Spatial Arrangement
- Practicality
- Ecological Fallacy
- Atomistic Fallacy

What area are you studying, and how will changing it affect your data?

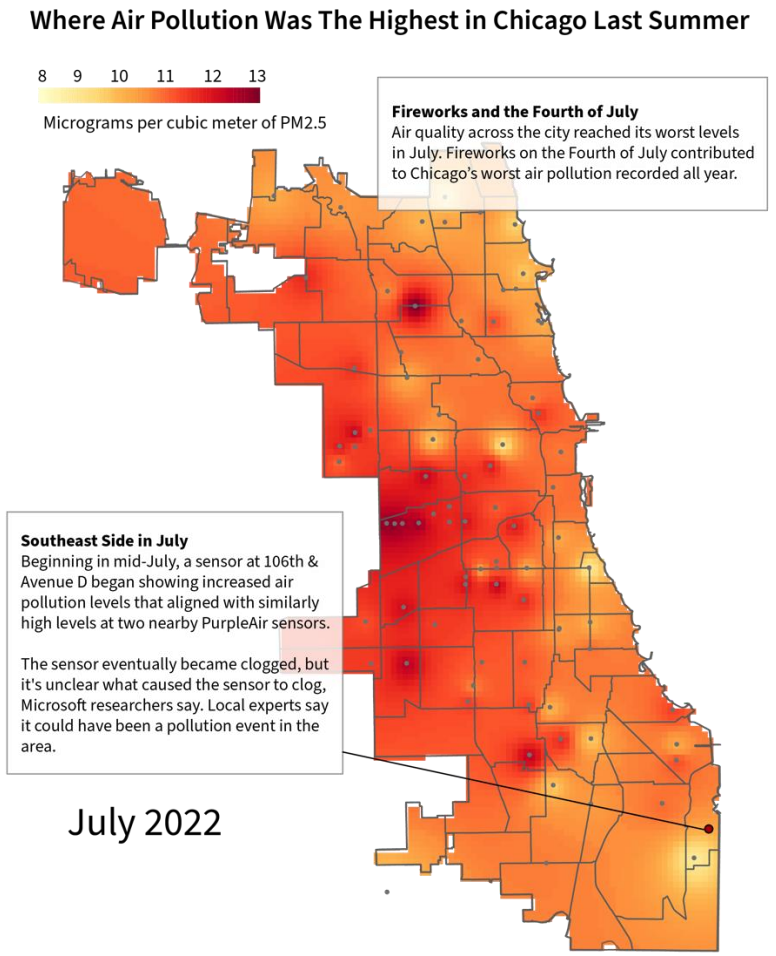
If you're studying urban heat islands in a city, focusing on just the downtown area might show high temperatures. But if you expand the study to include suburban areas, the average temperature might decrease, and the heat island effect could appear less severe. This change in study area can alter the overall conclusions about temperature patterns in the city.

Problem: Framing the question

- Spatial Scale
- **Statistical Scale**
- Spatial Arrangement
- Practicality
- Ecological Fallacy
- Atomistic Fallacy

How will you group your data for analysis and reporting?

If you're analyzing pollution levels, will you report them by neighborhood, city, or state? Reporting by neighborhood might show detailed patterns, while reporting by state could hide local variations.



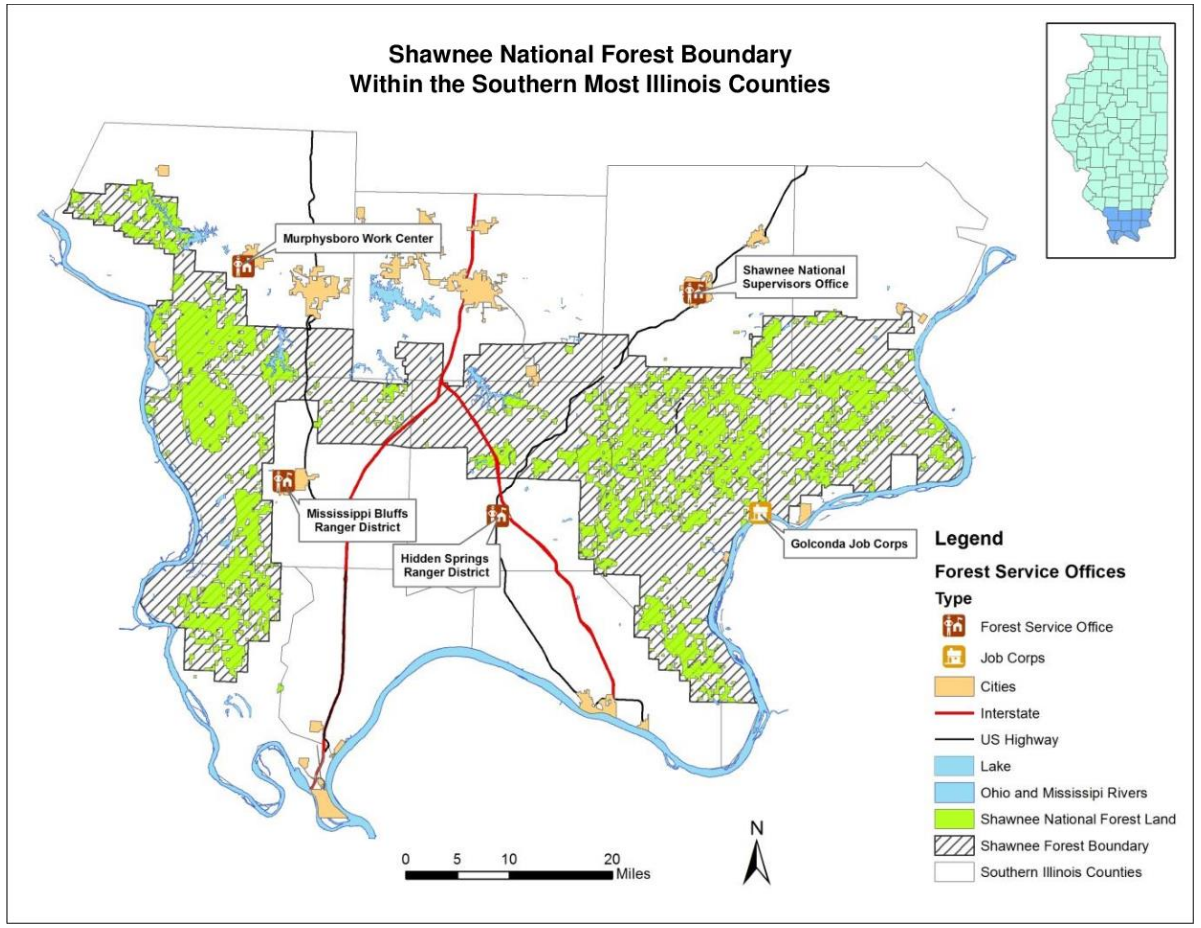
Source

Problem: Framing the question

- Spatial Scale
- Statistical Scale
- **Spatial Arrangement**
- Practicality
- Ecological Fallacy
- Atomistic Fallacy

Will changing the way you arrange or divide your study area affect your analysis?

If you're studying deforestation, dividing the area by natural forest boundaries might give different results compared to using political boundaries like districts, potentially affecting the accuracy of your findings.

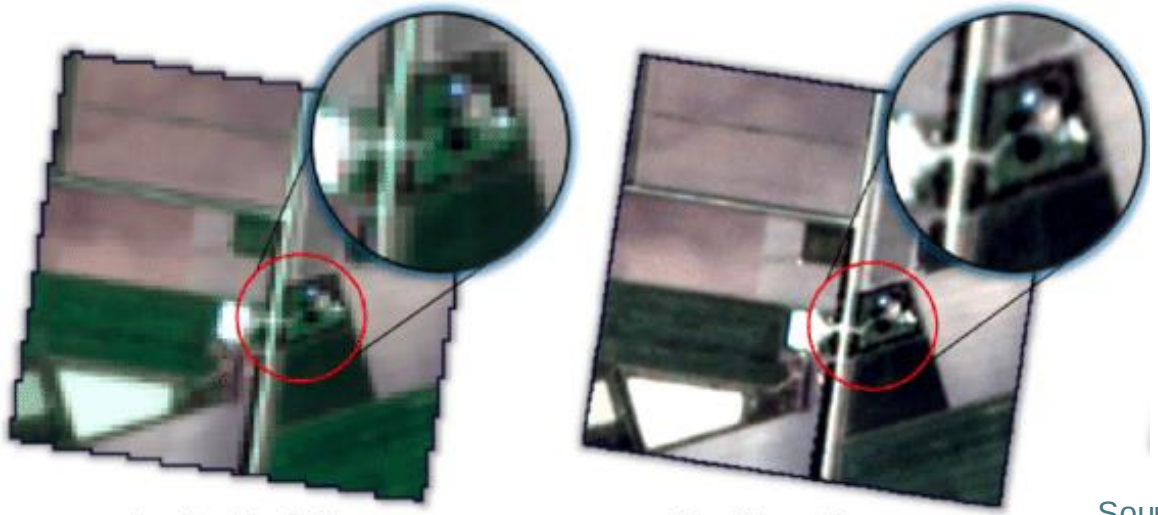


Source

Problem: Framing the question

- Spatial Scale
- Statistical Scale
- Spatial Arrangement
- **Practicality**
- Ecological Fallacy
- Atomistic Fallacy

Can you get the data you need within your time, budget, and resources?



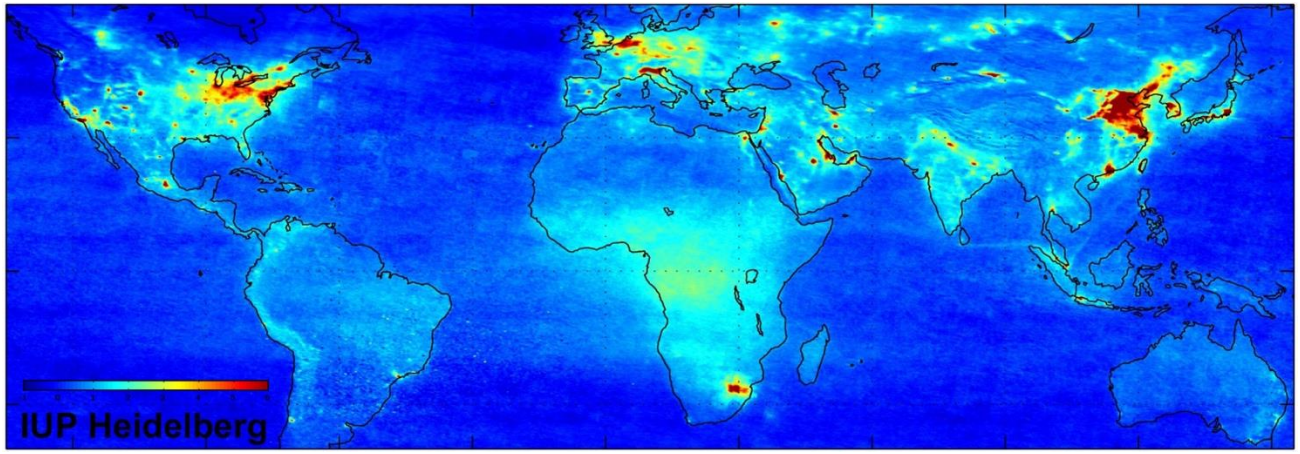
Source

If you need high-quality satellite images but only have a small budget and limited time, you might need to use lower-quality images instead.

Problem: Framing the question

- Spatial Scale
- Statistical Scale
- Spatial Arrangement
- Practicality
- **Ecological Fallacy**
- Atomistic Fallacy

Are you assuming that what is true for an entire group is also true for each individual within that group?



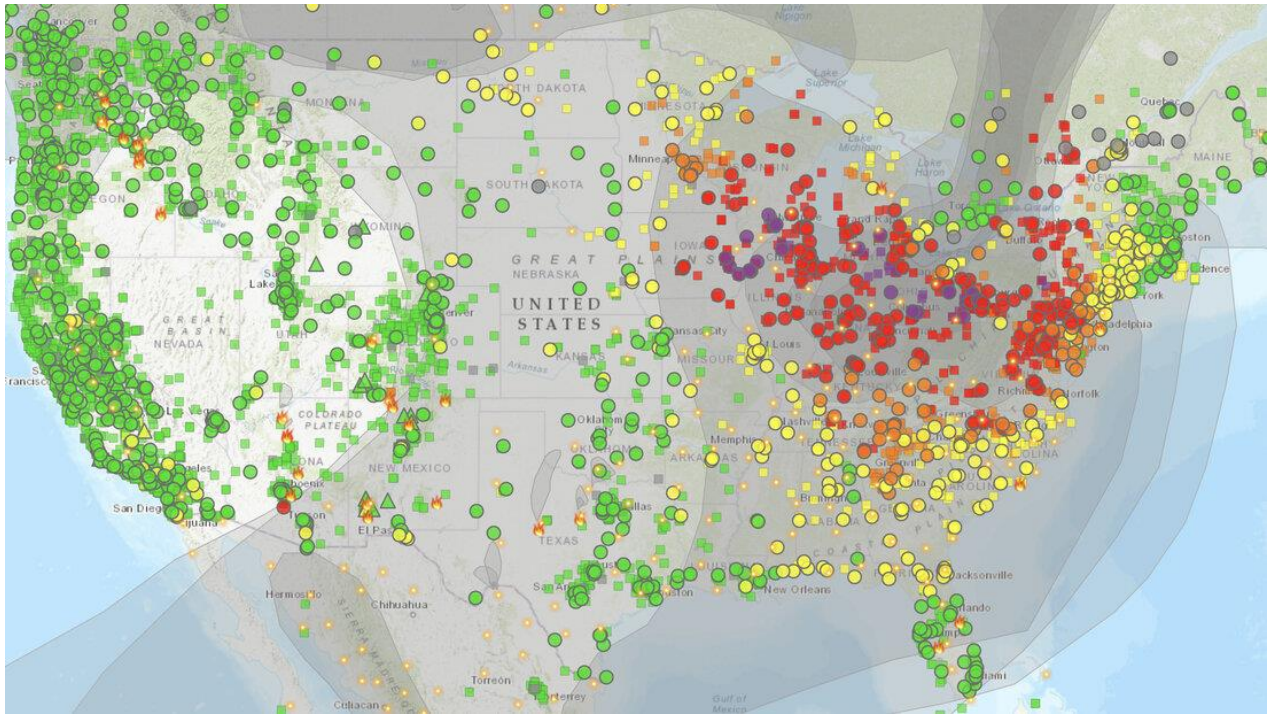
Source

If a region has high average air pollution, assuming every neighborhood has high pollution could be incorrect.

Problem: Framing the question

- Spatial Scale
- Statistical Scale
- Spatial Arrangement
- Practicality
- Ecological Fallacy
- **Atomistic Fallacy**

Are you assuming that what is true for a few sampled individuals is also true for the entire group they represent?



Source

If a few neighborhoods have low pollution, assuming the entire region has low pollution might also be wrong.

Thought Experiment

- Think about a major problem area in your domain/field/major/research
- You have 5 minutes to write down 2-3 sentences about your problem
- You will then tell us about the problem you have defined and what could be the potential solutions
- Consider your audience (your peers in this class) so that they can understand
- This is part of your class participation grade (Attendance)

Plan: Formulating the approach

Is it focused on exploring patterns or understanding causes and predicting outcomes?

What funds and other resources are available?

Does it require commercial costings and/or cost-benefit analysis?

Is the project technically feasible, and what are the risks of failure based on its complexity?

Are particular decision support tools and procedures needed?

What are the client (commercial, government, academic) expectations?

What level of public involvement and public awareness is involved?

Are there specifications, standards, quality parameters and/or procedures that must be used?

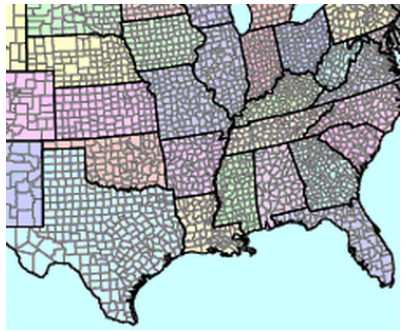
What particular operational needs and conditions area associated with the exercise?

How does the research relate to other studies on the same or similar problems?

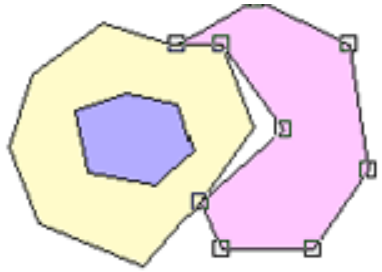
What time is available to conduct the research and are there any critical deadline?

What data components are needed and how will they be obtained?

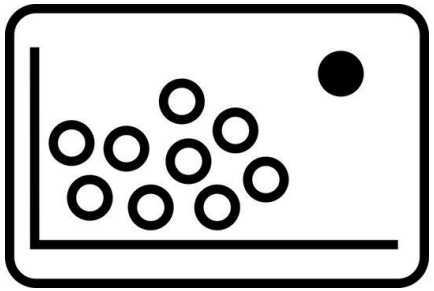
Data: Data acquisition



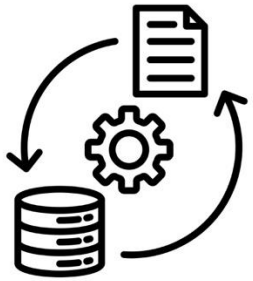
Boundary definition and density estimation



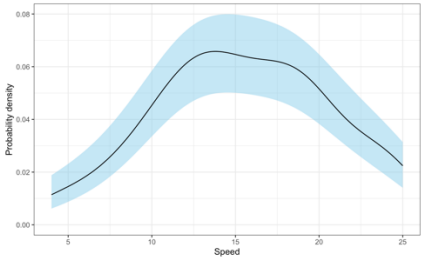
Modeling procedures that automatically adjust faulty topologies, datasets, and/or projections



Fill in missing data and/or detect outliers

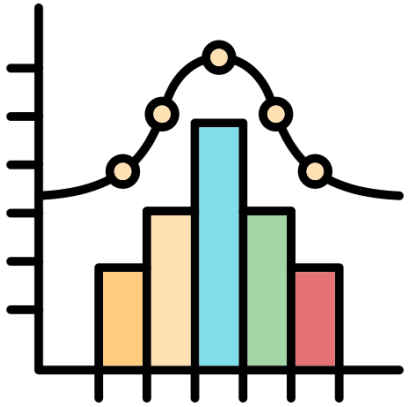


Data transformation, weighing, smoothing, normalization can be necessary



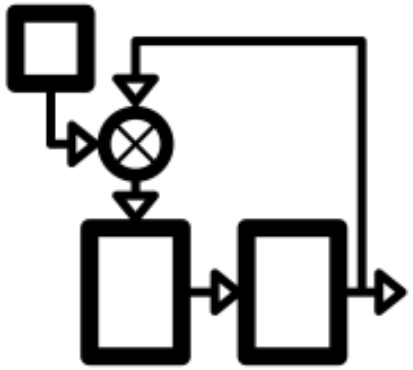
A range of techniques exist for modeling data problems and generating error bounds, confidence envelopes

Analysis: Analytical methods and tools



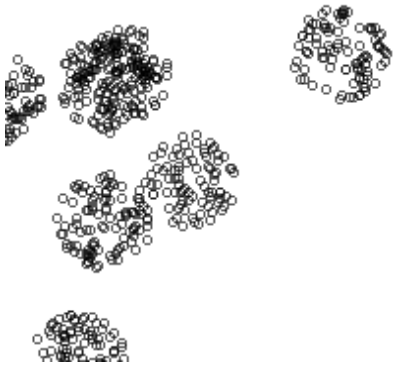
Exploratory Data Analysis

E.g., mean, median, mode, standard deviation, range, IQR, Outlier fixing



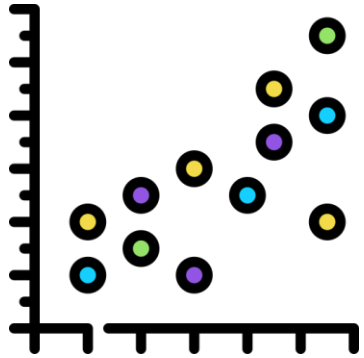
Mathematical modeling

Develop new methods or apply existing to interesting dataset



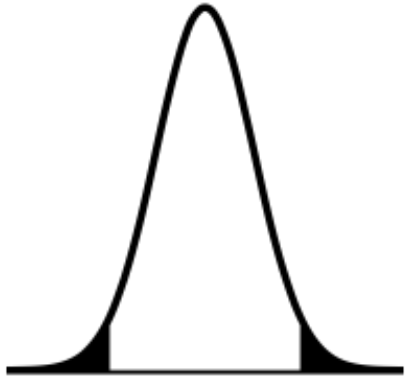
Spatial Patterns on interested variables

E.g., are variables spatially clustered? Are the clustered locally or globally?



Validation

Methods to validate findings, use data from other sources or set aside data from current study to validate; can be qualitative too.



Inference on Population

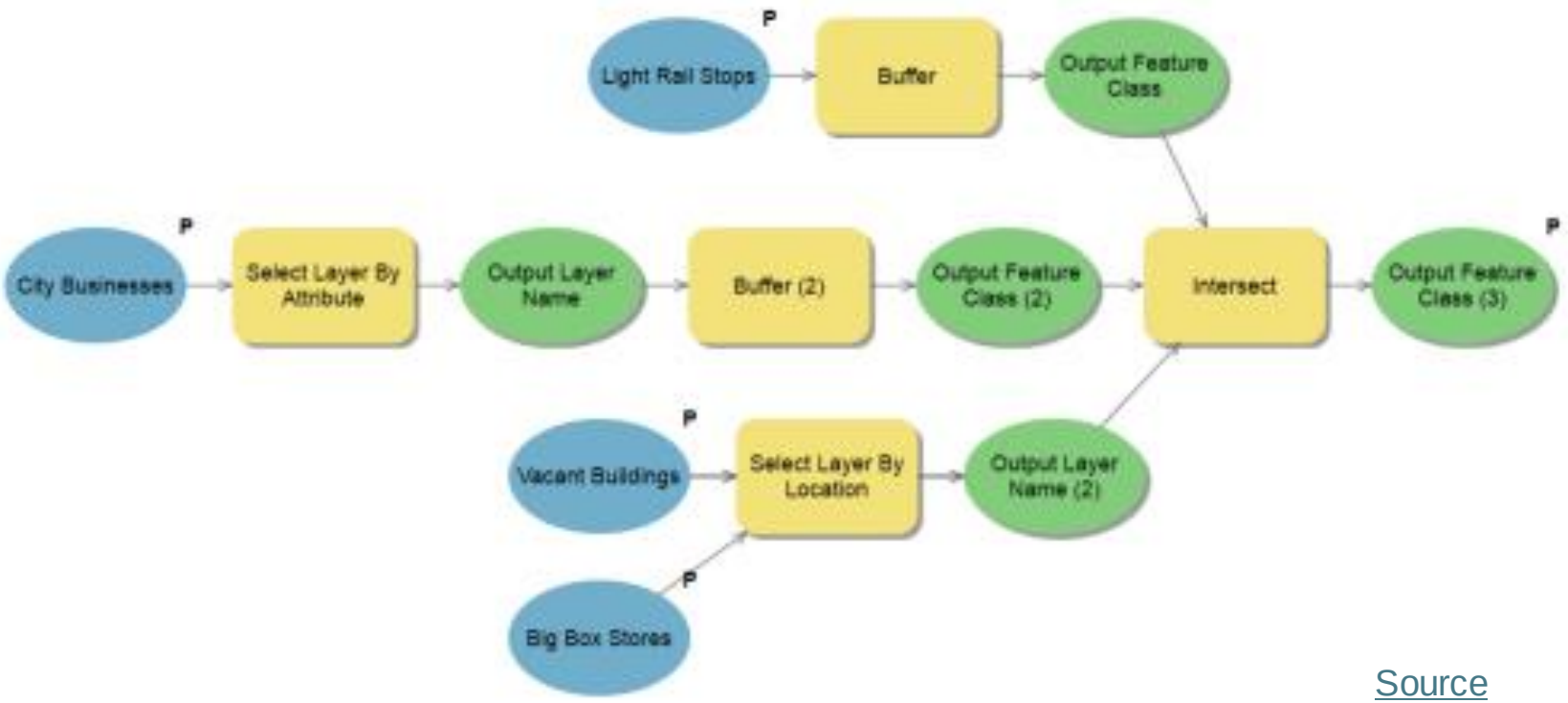
Calculate statistical significance of the results towards the population

Conclusion: Delivering the results

- According to Mackay and Oldford (2000)
  - The purpose of the Conclusion is to report the results in the language of the Problem.
  - Concise **numerical summaries** and **presentation graphics** should be used to clarify the discussion.
  - **Statistical jargon** should be avoided.
  - Discuss the **strengths** and **weakness** of the Plan, Data and Analysis
- Think of Conclusions as not the “Conclusion” section, but the Results and Discussion sections together.
- The Conclusion section should be the direct answer to your research questions.

Geospatial Model Building

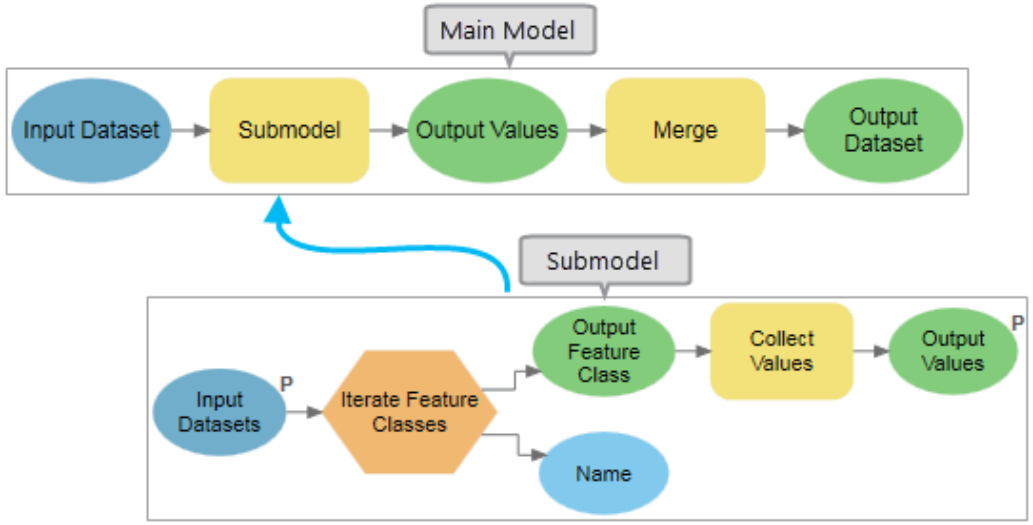
A collection of geoprocessing tools working together. It can be simple:

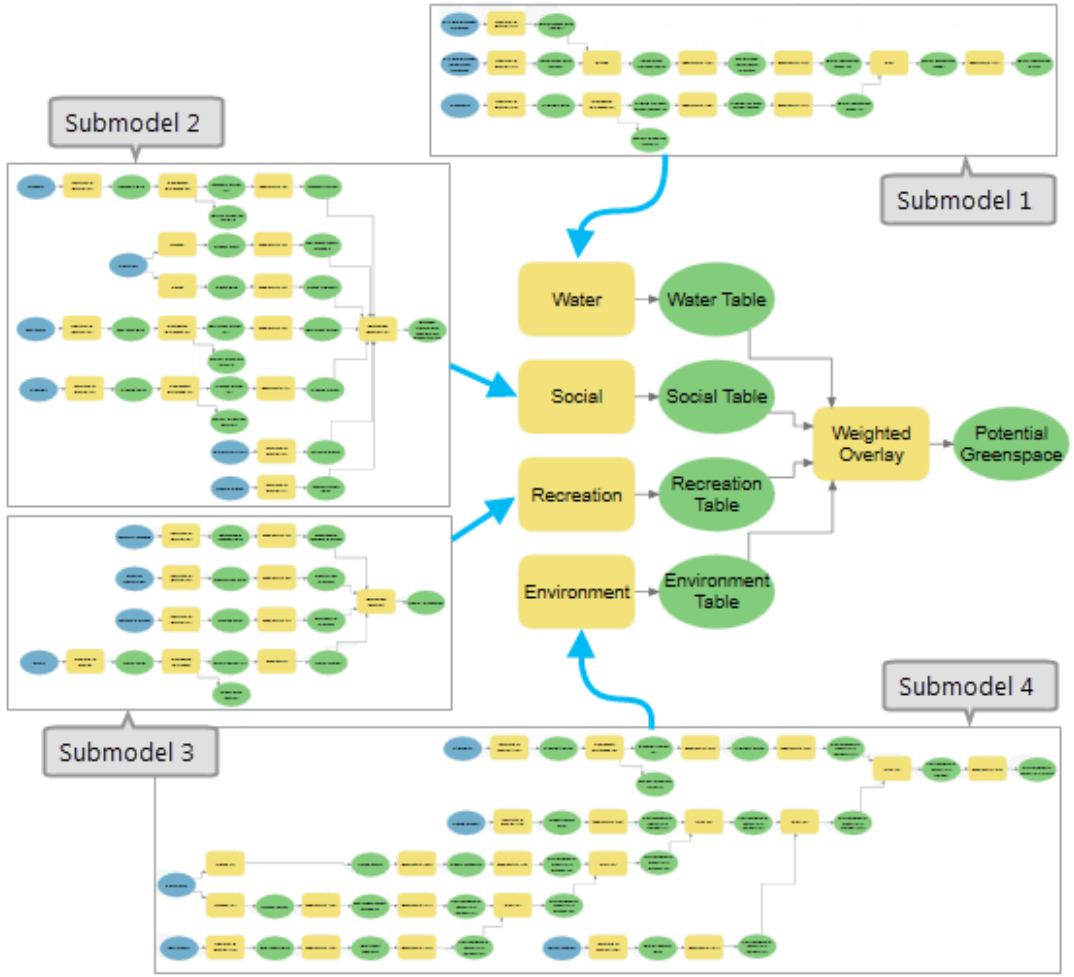


Source

Geospatial Model Building

Or it can be complex with sub models





Source

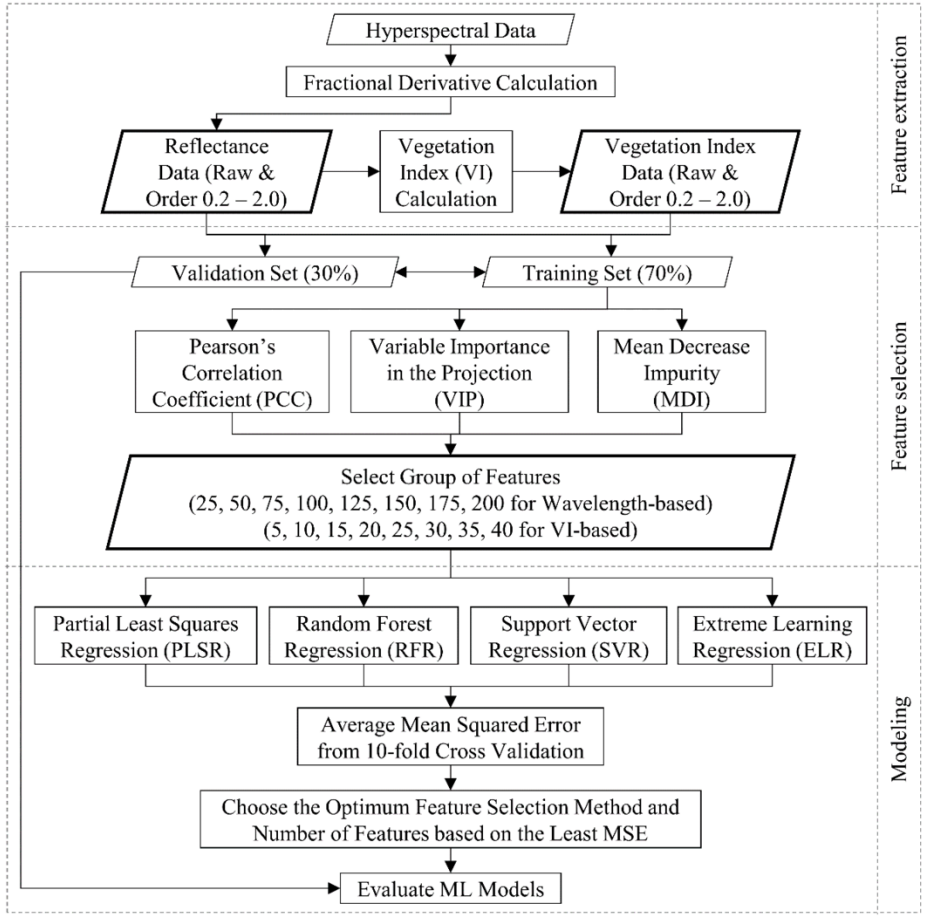
Geospatial Model Building

- Often, a geospatial model is not enough to squeeze the whole geospatial analytics methods.
- Rather, think of a model as part of the geospatial analytics process.
- At present, it is hard to rely on a a single geospatial software package for your project, which can result in multiple models in different platforms.
- It is always a good idea to define your methodology as part of workflow in your project.

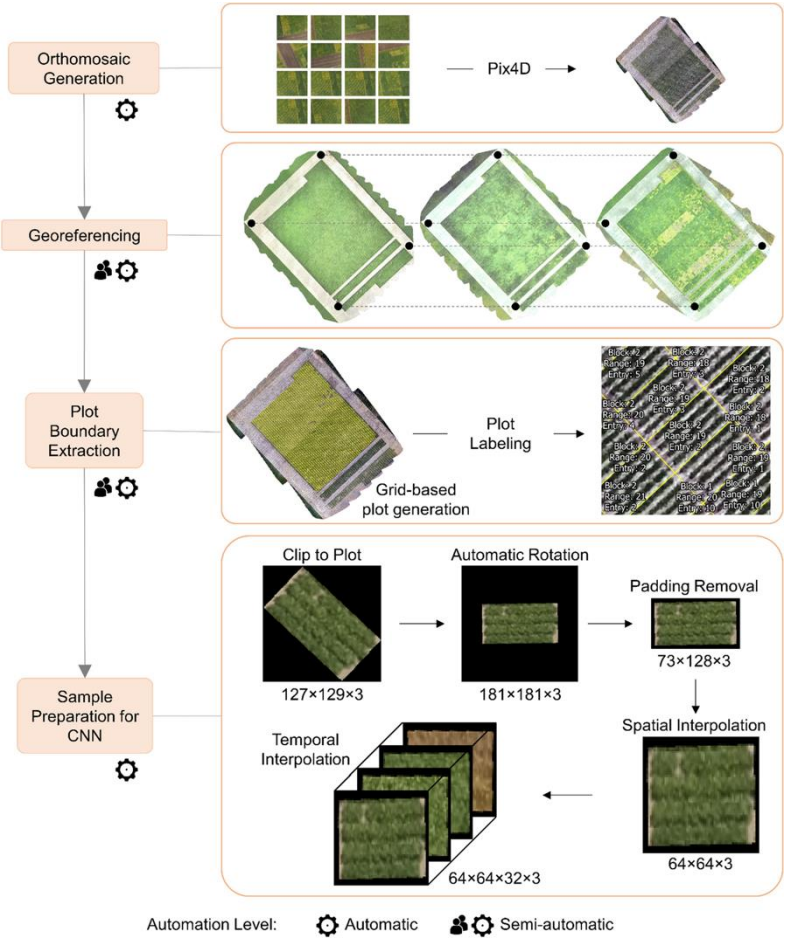
Source

Geospatial Model Building

Graphical workflows



Source



Source

In summary

- There is no cookie cutter approach when it comes to defining the methodological context for geospatial analytics.
- It is a collection of processes and often defined by the needs of the stakeholders.
- However, there are important questions to ask in every stages of the geospatial analytics.
- It is a good practice to ask the questions first rather later.
- Instead of focusing too much on data collection or methodological evaluation, similar importance should be given towards the problem definition.

SAINT LOUIS UNIVERSITY.

Thank You