



Evapotranspiration
sensor on drone

Smart phone

Decision support

Weather station

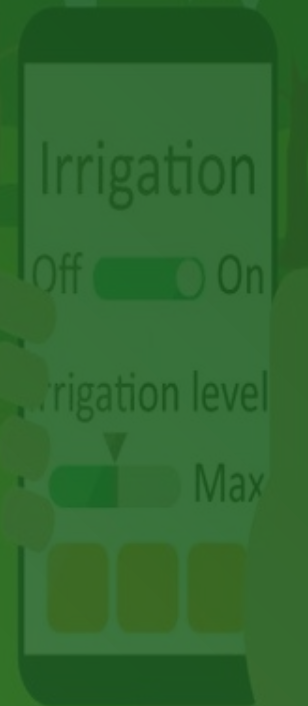
Soil moisture
sensor station

Data Collection and Processing Operations

Irrigation system

Dr. KEHILI HE

Academic Year 2025-2026



Fundamental Definitions



Data

Raw, unorganized facts collected through field observations, measurements, or experiments. In plant production, data includes values such as plant height, soil pH, or rainfall amounts without context or structure.



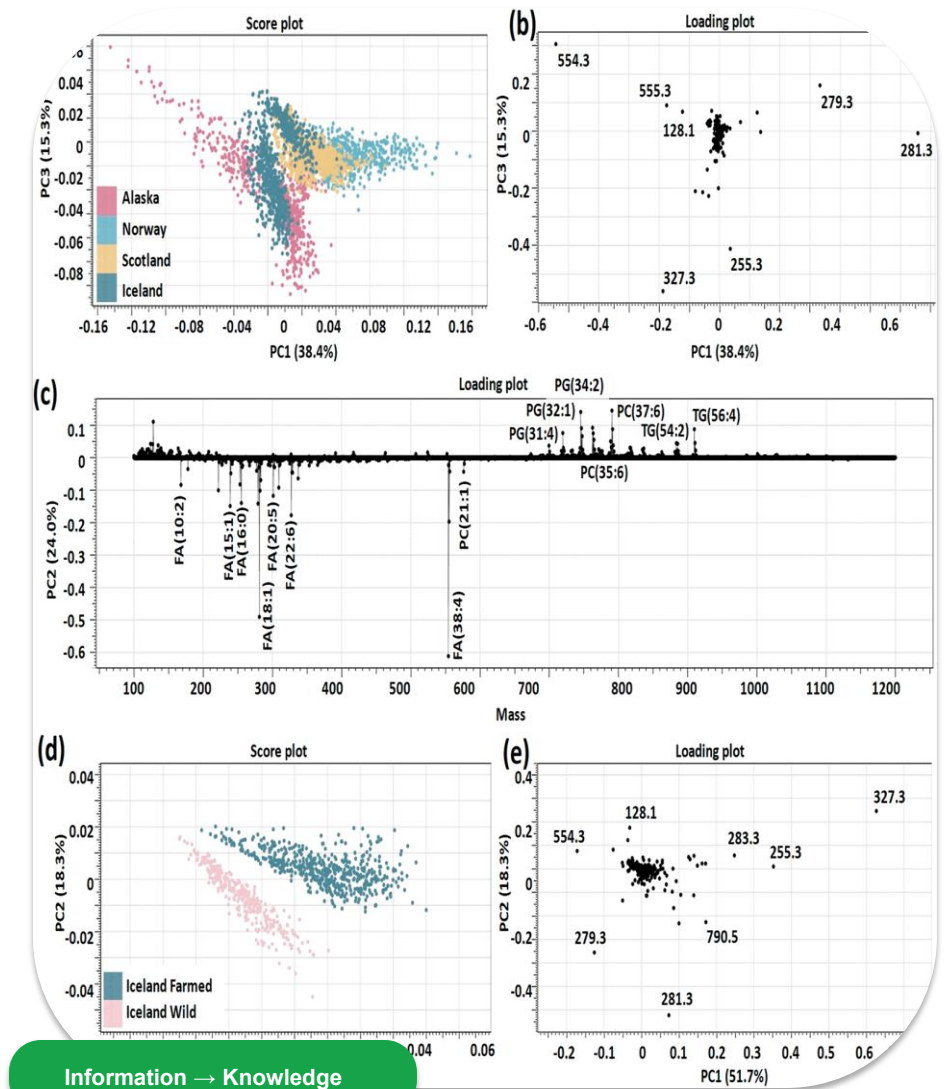
Information

Data that has been processed, organized, and structured to provide meaningful context. For instance, calculating the average yield of a wheat variety transforms raw measurements into actionable information for comparison.



Knowledge

The application of information and experience to understand patterns and make informed decisions. In agriculture, this means knowing that specific conditions lead to higher yields, enabling effective farm management strategies.



Types of Data in Plant Production

📍 Nature of Data

- **Quantitative Data**

Numerical measurements that can be counted or measured (e.g., yield in tons/ha, stem diameter, temperature in °C).

- **Qualitative Data**

Descriptive characteristics that cannot be measured numerically (e.g., disease severity, leaf color, soil texture).

📁 Source of Data

- **Primary Data**

Fresh information collected firsthand specifically for the current investigation (e.g., field trial measurements).

- **Secondary Data**

Existing data collected by others for different purposes (e.g., historical weather records, agricultural census).



Field-Level Collection

Why Data Collection is Crucial



Data collection forms the backbone of scientific investigation in plant production, enabling researchers and agronomists to base decisions on empirical evidence rather than assumptions.



Foundation of Scientific Method

Essential for testing hypotheses, validating experimental treatments, and establishing causality between agricultural practices and outcomes.



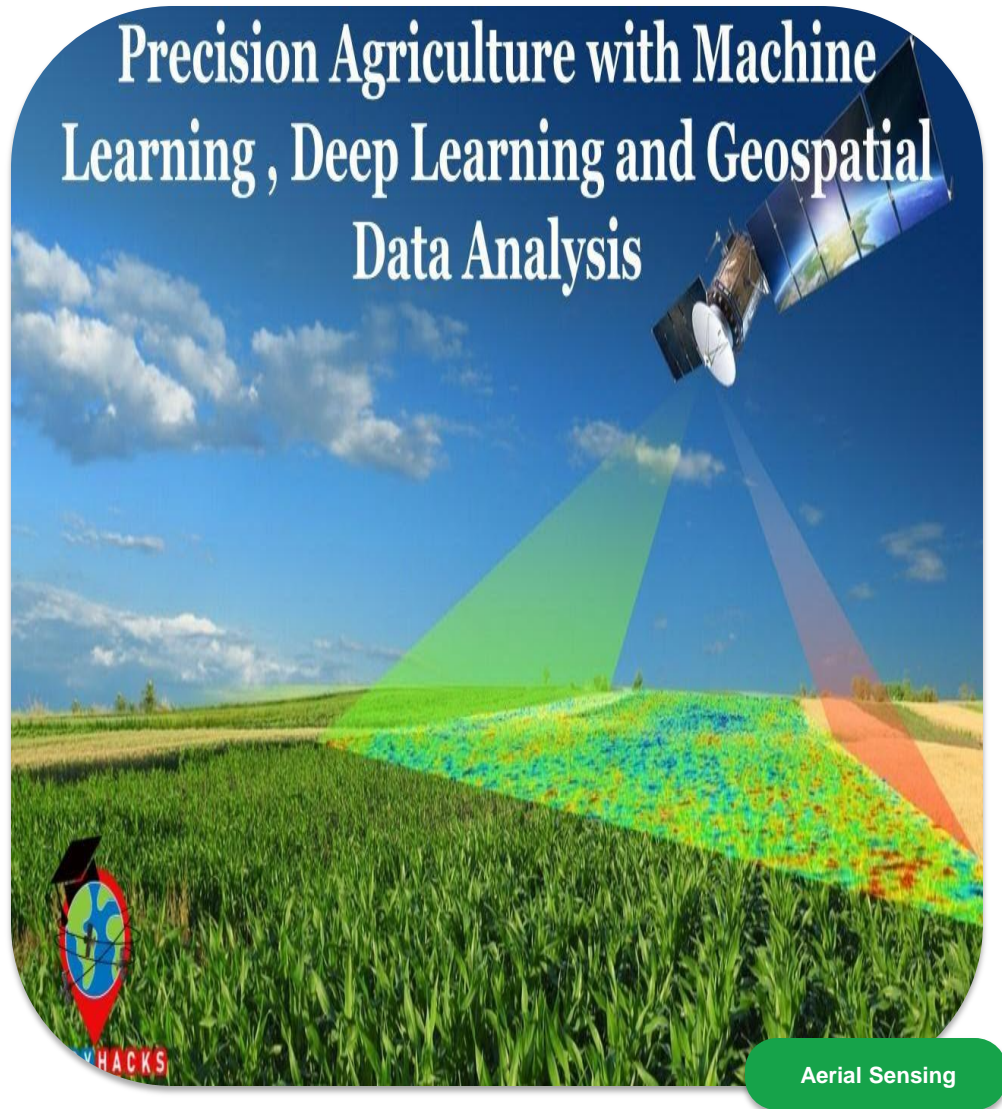
Managing Spatial Variability

Agricultural fields are inherently heterogeneous. Data collection captures this variability to enable site-specific management strategies.



Optimizing Inputs & Sustainability

Facilitates precise application of water, fertilizers, and pesticides, reducing waste, minimizing environmental impact, and maximizing economic returns.



Modern investigation methods leverage **digital technologies** to collect high-volume, high-accuracy data in real-time, transforming qualitative observations into quantitative metrics.



Precision Agriculture

A management strategy that gathers, processes, and analyzes spatial and temporal data to support site-specific decisions, optimizing inputs and maximizing yield.



Remote Sensing

Acquiring information without physical contact using sensors mounted on drones, satellites, or aircraft to capture multispectral imagery of crop health.



Internet of Things (IoT)

Network of physical devices embedded with sensors and connectivity, enabling real-time monitoring of soil moisture, temperature, and crop conditions.

Remote Sensing in Agriculture

Remote sensing involves acquiring information about objects or areas from a distance, without physical contact. In plant production, it is essential for monitoring crop health, biomass, and stress patterns over large areas.



Satellites vs. Drones (UAVs)

Satellites: Provide broad coverage (hectares to regions) with regular revisit times but lower spatial resolution.

Drones: Offer high-resolution imagery (cm-level) for detailed field analysis but limited coverage area.



NDVI: Vegetation Index

Normalized Difference Vegetation Index (NDVI) quantifies plant health by measuring the difference between near-infrared (reflected by healthy plants) and red light (absorbed).

Application: Early detection of nutrient deficiencies, water stress, and disease infestations before they are visible to the human eye.



Spatial Variability Mapping

IoT Devices for Agricultural Data

Internet of Things (IoT) refers to a network of interconnected sensors and devices that collect, exchange, and process data without human intervention, enabling real-time field monitoring.



Soil Sensors

Measure moisture content, temperature, and nutrient levels (NPK). Critical data for optimizing irrigation scheduling and fertilizer application timing.



Weather Stations

Monitor micro-climate conditions including humidity, wind speed, solar radiation, and precipitation for localized weather pattern analysis.



Automation & Alerts

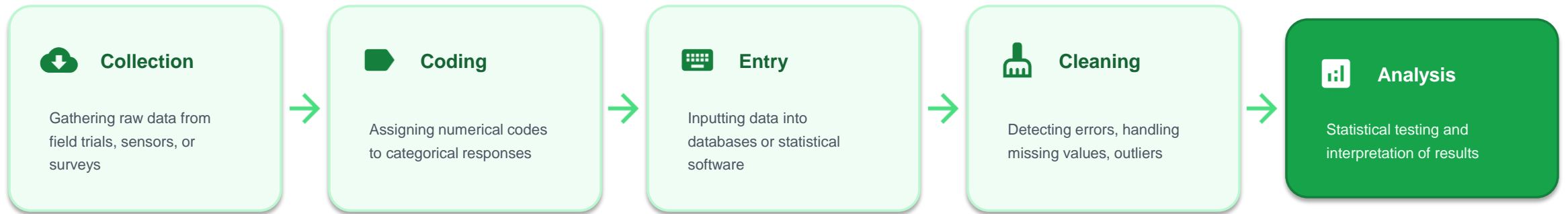
Enable automated irrigation systems and real-time alerts for frost, heat stress, or pest pressure, allowing immediate agronomic response.



Real-Time Monitoring

The Data Processing Chain

Data processing is the systematic transformation of raw observations into a usable, understandable format. It follows a strict sequence to ensure accuracy and reliability in research outcomes.



Error Reduction

Systematic processing minimizes human errors and biases in the final dataset, ensuring research validity.



Data Integration

Standardized coding and entry allow for merging datasets from multiple sources or experimental years.



Reproducibility

Documented processing steps ensure that research findings can be verified and replicated by other scientists.

Data Cleaning & Validation

Data cleaning is the critical process of detecting, correcting, or removing corrupt, inaccurate, or irrelevant records from a dataset. It ensures that subsequent analyses yield valid results.



Outlier Detection

Identifying extreme values that deviate significantly from the rest of the dataset (e.g., plant height of 500cm). Requires verification or removal to prevent skewed analysis.



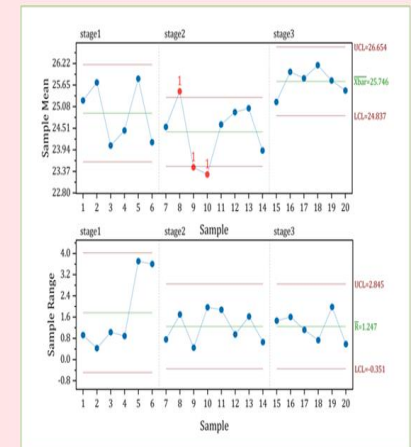
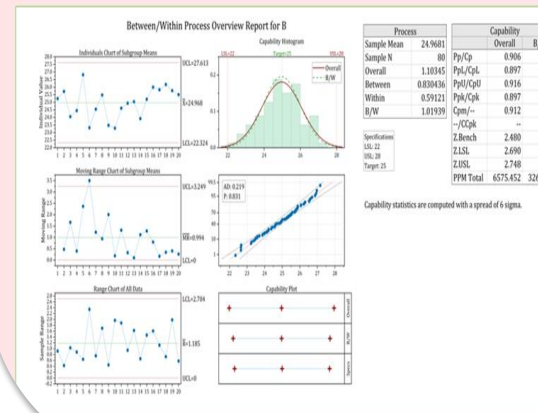
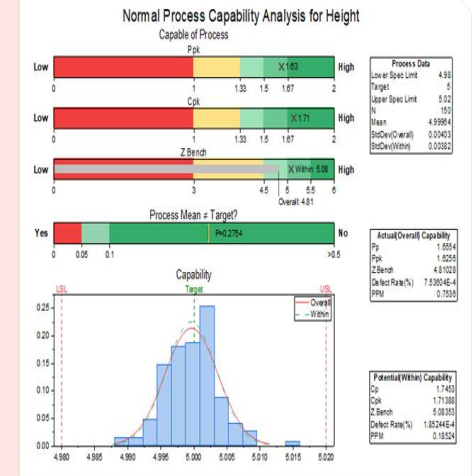
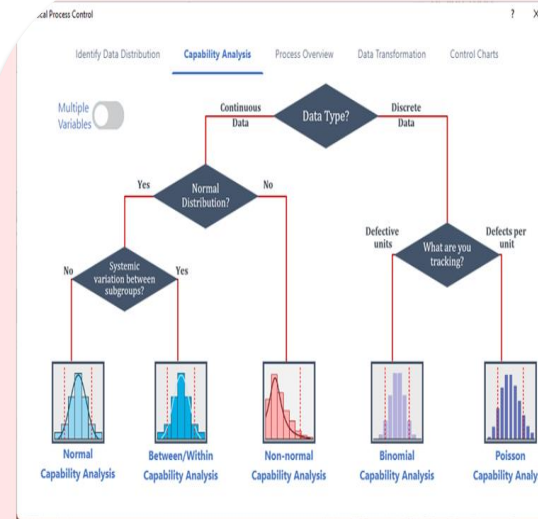
Handling Missing Data

Addressing gaps in the dataset through techniques such as listwise deletion or imputation (estimating missing values based on other available data).



Validation Procedures

Cross-checking data entries against physical reality or established standards to ensure consistency, accuracy, and reliability of the final dataset.



Quality Control Monitoring

Statistical Analysis Methods



Descriptive Statistics

Provides simple summaries about the sample and measures. Essential for understanding basic features of the data before further analysis.

Key measures: Mean, Median, Standard Deviation, Variance



Inferential Statistics

Allows researchers to make inferences about the population based on sample data. Used for hypothesis testing and determining significance of treatment effects.

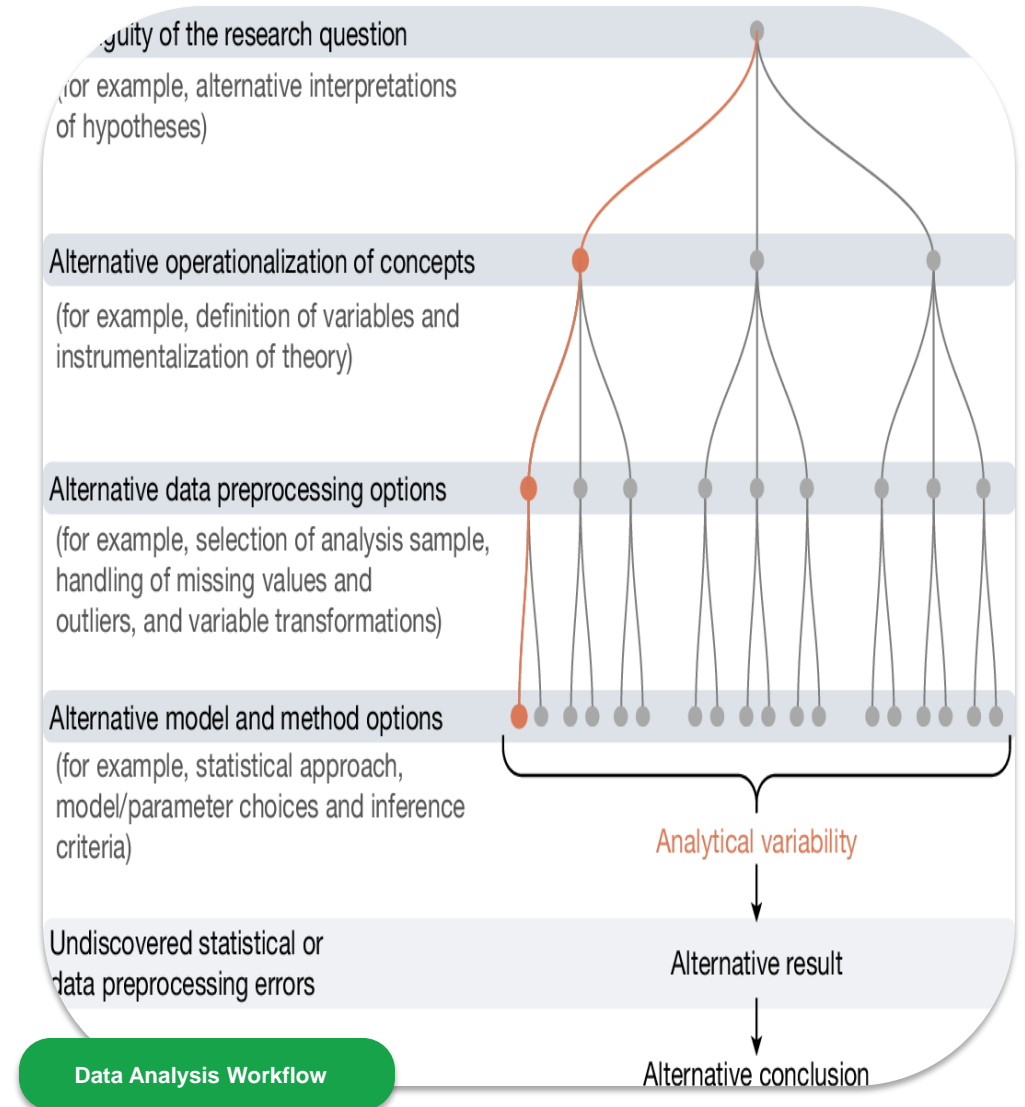
Key methods: ANOVA, t-test, Chi-square test



Advanced & Multivariate

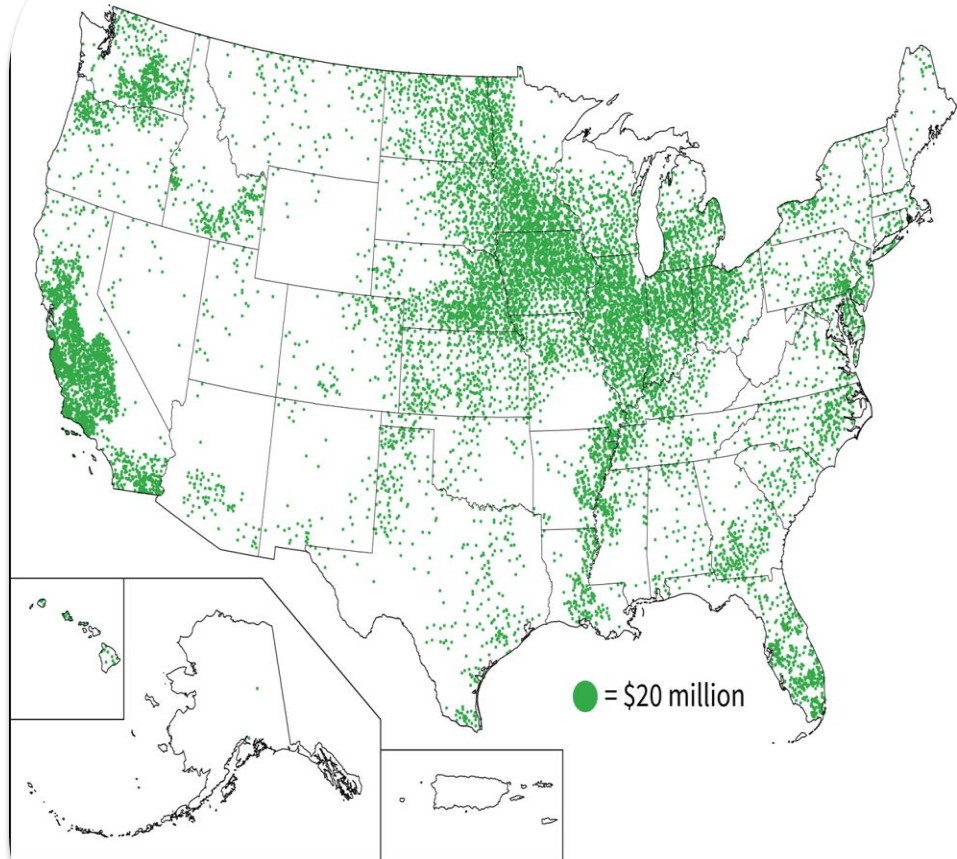
Analyzes complex datasets with multiple variables simultaneously. Crucial for modern precision agriculture and handling spectral data.

Key techniques: PCA, Regression Analysis, Cluster Analysis



Data Visualization & Interpretation

Market value of crops in 2022



Source: USDA, Economic Research Service using data from USDA, National Agricultural Statistics

Geospatial Data Representation

Data visualization is the graphical representation of information and data. Using visual elements like charts, graphs, and maps provides an accessible way to understand trends, outliers, and patterns in agricultural research.



Graphs & Charts

Bar charts, histograms, and scatter plots summarize treatment effects and relationships between variables (e.g., yield vs. nitrogen application).



GIS Mapping

Essential for displaying spatial variability in fields. Yield maps, nutrient distribution, and soil type maps guide precision agriculture decisions.



Interpretation

The critical final step where researchers explain what visual patterns mean in the context of their hypothesis, transforming data into actionable agricultural knowledge.