Day 3: Capstone Research Project Guide

Independent GIS and AI-Assisted Spatial Health Analysis

Research Framework | Spatial Analysis | AI Integration | Professional Presentation

Welcome to Your Capstone Research Project

Today represents the culmination of your workshop experience. You'll apply everything you've learned across the previous labs to conduct an independent research project that addresses real-world public health challenges using cutting-edge geospatial technologies.

Project Objectives

- Apply scientific methodology to spatial health problems
- Integrate QGIS, Google Earth Engine, and ChatGPT effectively
- Develop professional research presentation skills
- Create portfolio-worthy research outputs

Spatial Analysis Workflow

1. Research Project Framework

Scientific Method in Spatial Health

1. **Observation & Problem Identification** – Identify spatial patterns in health data that require explanation
2. **Research Question Formulation** – Develop testable questions about spatial relationships
3. **Hypothesis Development** – Propose explanations based on spatial theory
4. **Data Collection & Analysis** – Gather and analyze spatial data using GIS/remote sensing
5. **Interpretation & Conclusions** – Draw evidence-based conclusions about spatial relationships

Research Question Development

Strong Research Questions:

- Are spatially specific and measurable
- Address public health significance
- Can be answered with available data
- Consider spatial scale appropriately

Example Questions:

- "How does rainfall variation affect malaria prevalence across Ugandan districts?"
- "Which areas have high disease burden but poor facility access?"
- "What environmental conditions cluster together to create disease risk?"

Ethical Considerations in Health Research

Data Privacy

- Use aggregated data only
- Protect individual privacy
- Follow data use agreements

Representation

- Avoid stigmatization
- Present balanced findings
- Consider social context

Interpretation

- Acknowledge limitations
- Avoid causal overstatement
- Consider alternative explanations

2. Project Selection and Planning

Project 1: Malaria-Rainfall Correlation

Objective:
Analyze relationship between precipitation patterns and malaria prevalence across Uganda

Key Datasets:

- CHIRPS daily rainfall data
- District malaria prevalence
- Uganda administrative boundaries

Tools Required: GEE, QGIS, ChatGPT

Project 2: Healthcare Access Gaps

Objective:
Identify high-burden areas with inadequate healthcare facility access

Key Datasets:

- Health facility coordinates
- District health indicators
- Population density data

Tools Required: QGIS, GEE, ChatGPT

Project 3: Environmental Risk Clustering

Objective:
Use machine learning to identify environmental risk zones for disease transmission

Key Datasets:

- MODIS NDVI time series
- CHIRPS precipitation
- Elevation data (SRTM)

Tools Required: GEE, ChatGPT, QGIS

Project Design Template

1. Project Title
2. Research Question
3. Hypothesis
4. Required Datasets
5. Main Tools
6. Expected Output
7. Intended Use
8. Timeline & Milestones

3. Technical Integration Workflow

Google Earth Engine

- **Data Access & Loading** – Access satellite data collections and administrative boundaries
- **Temporal Filtering** – Filter data by date ranges relevant to your research question
- **Spatial Processing** – Apply reductions, clipping, and zonal statistics
- **Visualization** – Create appropriate visualizations with color palettes
- **Export** – Export processed data to Google Drive for further analysis

QGIS Desktop

- **Data Import** – Load vector and raster data from multiple sources
- **Table Joins** – Connect spatial and attribute data using common fields
- **Spatial Analysis** – Perform buffer analysis, spatial queries, and geoprocessing
- **Visualization** – Apply symbology and create professional map layouts
- **Output** – Export maps and analysis results in various formats

ChatGPT Assistance

- **Code Generation** – Generate initial scripts for data processing tasks
- **Debugging Support** – Troubleshoot errors and optimize code performance
- **Method Explanation** – Understand complex functions and analytical approaches
- **Literature Review** – Research background information and best practices
- **Documentation** – Create clear documentation and explanations

Data Management & Quality Control

File Organization

```
Project_Name/
  01_Data/
    raw_data.csv
    processed_data.gpkg
  02_Scripts/
    gee_analysis.js
    qgis_project.qgs
  03_Output/
    final_map.pdf
    results_table.csv
  04_Documentation/
    methodology.md
    data_sources.txt
```

Quality Control Checklist

- [ ] Data sources documented with full citations
- [ ] Coordinate reference systems consistent
- [ ] Processing steps recorded and reproducible
- [ ] Intermediate outputs validated
- [ ] Error handling implemented
- [ ] Results cross-checked with alternative methods
- [ ] Backup copies created

4. Research Methodology

Spatial Analysis Design Principles

Scale Considerations

- Match analysis scale to research question
- Consider ecological fallacy risks
- Account for modifiable areal unit problem

Temporal Considerations

- Align temporal scales of different datasets
- Consider lag effects and seasonality
- Account for data collection timing

Spatial Dependencies

- Test for spatial autocorrelation
- Consider spatial spillover effects
- Account for boundary effects

Statistical Approaches

Descriptive Analysis

- Calculate summary statistics by region
- Create distribution maps and histograms
- Identify outliers and data quality issues

Correlation Analysis

- Pearson/Spearman correlation coefficients
- Scatter plots with spatial context
- Consider non-linear relationships

Spatial Statistics

- Moran's I for spatial autocorrelation
- Local indicators of spatial association
- Hot spot analysis (Getis-Ord)

Validation & Uncertainty Assessment

Data Validation

- Cross-reference with ground truth data
- Compare with published literature
- Check for consistency across sources
- Assess data completeness and coverage

Method Validation

- Test alternative analytical approaches
- Sensitivity analysis on parameters
- Cross-validation techniques
- Bootstrap confidence intervals

Uncertainty Quantification

- Document data limitations
- Propagate measurement errors
- Report confidence intervals
- Discuss interpretation caveats

Literature Review & Background Research

Key Search Strategies

Academic Databases:
PubMed, Web of Science, Google Scholar, PLoS ONE

Search Terms:
"spatial epidemiology", "disease mapping", "GIS health", "remote sensing malaria"

Grey Literature:
WHO reports, government health statistics, NGO publications

Critical Evaluation Framework

- [ ] Study design appropriate for research question
- [ ] Sample size and geographic scope adequate
- [ ] Data sources clearly documented
- [ ] Analytical methods properly described
- [ ] Limitations and biases acknowledged
- [ ] Results relevant to your context

5. Implementation Guidelines

Step-by-Step Technical Execution

A. Google Earth Engine Workflow

1. Environment Setup

- Open `https://code.earthengine.google.com/`
- Create new script with descriptive name
- Define study area boundary

2. Data Loading

- Load required image collections
- Apply date and spatial filters
- Select relevant bands/variables

3. Processing

- Apply temporal reductions (mean, sum, median)
- Clip to study area
- Calculate zonal statistics if needed

4. Visualization

- Choose appropriate color palettes
- Set min/max values for scaling
- Add layers to map with descriptive names

5. Export

- Use `Export.image.toDrive()`
- Set appropriate resolution and projection
- Monitor task completion

6. Quality Check

- Verify export completed successfully
- Check data ranges and distributions
- Document any processing issues

B. QGIS Desktop Workflow

1. Project Setup

- Create new project and save immediately
- Set appropriate CRS for region
- Organize layers in logical groups

2. Data Import

- Load vector data (shapefiles, GeoPackages)
- Import raster data from GEE exports
- Add CSV data as delimited text layers

3. Data Integration

- Join tables using common identifiers
- Ensure data types are compatible
- Handle missing or mismatched records

4. Spatial Analysis

- Perform buffer analysis for proximity
- Use spatial selection tools
- Calculate spatial statistics

5. Visualization

- Apply graduated symbology
- Choose appropriate classification methods
- Add labels and annotations

6. Layout Design

- Create print layout with map frame
- Add legend, scale bar, north arrow
- Export as PDF or high-resolution image

C. AI-Assisted Problem Solving

Effective ChatGPT Prompting for GIS Projects

Code Generation Prompts:

"Write a GEE script to calculate annual mean NDVI for Uganda using MODIS data"
"Create QGIS Processing model for buffer analysis around health facilities"
"Generate Python script to export attribute table to CSV with specific columns"

Debugging & Explanation:

"Explain this GEE error: 'Image.filter is not a function'"
"Why is my QGIS join not working? Here's my data structure..."
"How can I optimize this code for better performance?"

Common Integration Challenges & Solutions

Data Format Issues

Problem: CRS mismatch between datasets
Solution: Reproject all data to common CRS (e.g., WGS84)

Problem: Table join fails
Solution: Check field data types and clean text fields

Problem: Large file export times out
Solution: Reduce resolution or split into smaller regions

Analysis Errors

Problem: Unexpected analysis results
Solution: Validate with subset data and check calculations

Problem: Missing data in results
Solution: Check spatial and temporal filters

Problem: Script runs slowly
Solution: Optimize filters and reduce computational complexity

6. Professional Presentation

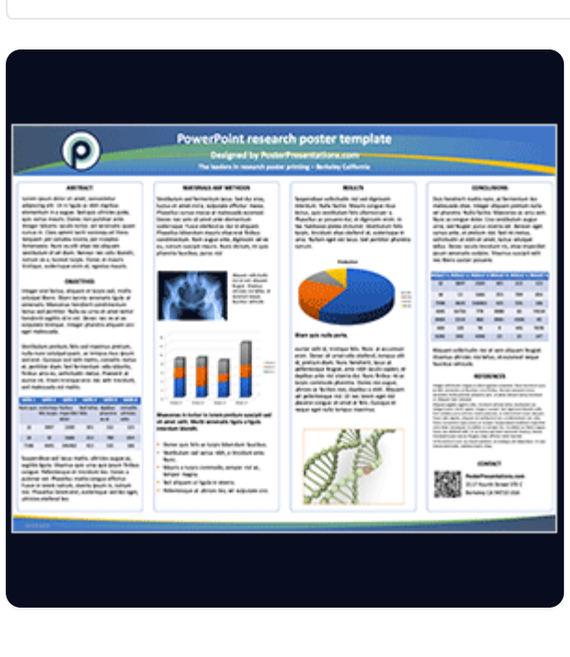
Research Poster Design Principles

Layout Structure

- **Title:** Clear, descriptive, institution/author info
- **Introduction:** Background, objectives, research questions
- **Methods:** Data sources, analytical approach, tools used
- **Results:** Key findings with maps and visualizations
- **Discussion:** Interpretation, limitations, implications
- **Conclusion:** Main takeaways and future directions

Visual Design Guidelines

- Use consistent fonts (max 3 different sizes)
- Maintain white space for readability
- Apply institution branding appropriately
- Ensure text is readable from 3 feet away



Effective Data Visualization

Map Design

- Choose appropriate classification method
- Use colorblind-friendly color schemes
- Include essential map elements (legend, scale, north arrow)
- Ensure sufficient contrast for printing

Charts & Graphs

- Match chart type to data type and message
- Label axes clearly with units
- Highlight key patterns or outliers
- Keep design simple and uncluttered

Color Psychology

- Red/Orange: High values, risk, heat
- Blue: Water, cold, low values
- Green: Vegetation, growth, positive
- Gray: Neutral, no data, uncertainty

Storytelling with Spatial Data

1. **Context Setting** – Establish the public health problem and why spatial analysis matters
2. **Pattern Revelation** – Show spatial patterns clearly and guide viewer attention
3. **Insight Generation** – Explain what patterns mean and their implications
4. **Action Orientation** – Connect findings to concrete actions or recommendations

Presentation Delivery Techniques

Preparation

- [ ] Practice 2-minute elevator pitch
- [ ] Prepare for common questions
- [ ] Test all technology beforehand
- [ ] Bring backup copies of materials

During Presentation

- [ ] Maintain eye contact with audience
- [ ] Use pointer/gestures effectively
- [ ] Speak clearly and at appropriate pace
- [ ] Welcome questions and interaction

Key Messages

- [ ] State research question clearly
- [ ] Highlight novel findings
- [ ] Acknowledge limitations honestly
- [ ] Connect to broader implications

7. Project Assessment

Evaluation Criteria & Rubric

Criteria	Excellent (4)	Good (3)	Satisfactory (2)	Needs Improvement (1)	Weight
Research Question	Clear, specific, answerable, significant	Clear and answerable	Somewhat clear	Vague or overly broad	20%
Technical Implementation	Sophisticated use of tools, error-free	Competent use, minor errors	Basic use, some errors	Limited use, major errors	25%
Data Analysis	Appropriate methods, well-executed	Mostly appropriate	Basic analysis	Inappropriate or flawed	20%
Visualization	Professional, clear, effective	Clear and readable	Basic but functional	Poor or confusing	15%
Interpretation	Insightful, well-supported conclusions	Sound interpretation	Basic interpretation	Unsupported or incorrect	20%

Peer Review Guidelines

Review Structure

- Summarize the project's main objective
- Identify 2-3 strengths of the work
- Suggest 2-3 areas for improvement
- Provide constructive, specific feedback

Focus Areas

- Clarity of research question and methods
- Appropriateness of analytical approach
- Quality of visualizations and maps
- Validity of conclusions drawn

Self-Reflection Framework

Learning Assessment

- What new skills did you develop?
- What challenges did you overcome?
- How did AI tools help your workflow?
- What would you do differently next time?

Future Applications

- How might you extend this analysis?
- What additional data would be useful?
- How could results inform policy?
- What career opportunities connect to this work?

Building Your Professional Portfolio

Technical Artifacts

- [ ] Final project maps and visualizations
- [ ] Code scripts (GEE, Python, R)
- [ ] Technical documentation
- [ ] Data processing workflows

Communication Materials

- [ ] Research poster or presentation
- [ ] Project summary/abstract
- [ ] Video presentation (if created)
- [ ] Blog post or article draft

Reflection & Learning

- [ ] Learning reflection essay
- [ ] Skills inventory/checklist
- [ ] Certificate of completion
- [ ] Future learning plan

8. Real-World Applications

Public Health Practice Integration

Disease Surveillance Systems

Your skills directly apply to:

- Real-time outbreak detection and monitoring
- Environmental risk factor identification
- Resource allocation optimization
- Public health emergency response

Program Planning & Evaluation

Applications include:

- Intervention targeting and prioritization
- Health facility planning and placement
- Community health worker deployment
- Program impact assessment

Policy & Decision Making

Evidence-Based Policy

Your analysis contributes to:

- National health strategy development
- Resource allocation decisions
- Health system strengthening initiatives
- International aid targeting

Stakeholder Communication

Skills for engaging:

- Government health officials
- International development organizations
- NGOs and community groups
- Funding agencies and donors

Career Development Pathways

Public Health

- Epidemiologist
- Health Data Analyst
- Disease Surveillance Specialist
- Health Program Manager

Data Science

- Geospatial Data Scientist
- Health Informatics Specialist
- Research Analyst
- Business Intelligence Analyst

Academic Research

- Graduate Student (MPH/PhD)
- Research Assistant
- Post-doctoral Researcher
- Faculty Member

Consulting

- GIS Consultant
- Health Systems Consultant
- Environmental Health Specialist
- International Development Advisor

Future Research & Technology Directions

Emerging Technologies

Machine Learning Integration:
Deep learning for satellite image analysis, automated disease prediction models

Real-time Data Streams:
Social media health monitoring, mobile phone mobility data, IoT sensor networks

Data Integration

Multi-source Fusion:
Combining satellite, census, health, and social media data for comprehensive analysis

Temporal Dynamics:
Incorporating climate change projections, demographic transitions, urbanization trends

Global Health Applications

Climate-Health Nexus:
Disease risk under changing climate conditions, extreme weather health impacts

Health Equity:
Mapping health disparities, social determinants, environmental justice applications

Professional Networking & Continued Learning

Professional Organizations

International Association of Geographic Information Science (ICA-GIScience)
Global network of GIS researchers and practitioners

International Epidemiological Association (IEA)
Leading organization for epidemiologists worldwide

American Public Health Association (APHA)
Largest public health organization with GIS special interest groups

Continuing Education Resources

Online Courses:
Coursera, edX, ESRI Training, Google Earth Engine tutorials

Conferences & Workshops:
Annual GIS conferences, public health meetings, specialized workshops

Open Source Communities:
QGIS user groups, R spatial analysis community, Python geospatial developers

Final Project Checklist & Next Steps

Final Deliverable Requirements

- [ ] Clear project title and research question stated
- [ ] Professional map/visualization created
- [ ] 2-3 key findings clearly presented
- [ ] Data sources and methods documented
- [ ] Limitations and uncertainties acknowledged
- [ ] Public health implications discussed
- [ ] Export in presentable format (PDF/PNG)
- [ ] Prepared for 5-minute presentation

Beyond the Workshop

Immediate Actions (This Week)

- Complete and refine your project
- Share work with colleagues and mentors
- Connect with other workshop participants
- Save all materials to portfolio

Short-term Goals (Next Month)

- Explore advanced features of tools used
- Join relevant professional communities
- Apply skills to other projects or datasets
- Consider presenting at local meetings

Long-term Development (Next Year)

- Pursue formal training or certification
- Develop specialized expertise areas
- Contribute to open source projects
- Mentor others in these skills

Congratulations on Completing Your Capstone Project!

You've successfully integrated GIS, remote sensing, and AI tools to address real-world public health challenges. These skills will serve you well in your future academic and professional endeavors.

Technical Skills Mastered | Research Methods Applied | Career-Ready Portfolio Created