Your Eyes in the Sky

Using Remote Sensing to Improve Planning, Data Collection and Impact Evaluation

Seth Morgan

[1]The Department of Agricultural & Consumer Economics
The University of Illinois at Urbana-Champaign
The World Agroforestry Centre (ICRAF)

ECHO Best Practices in Highland Areas Symposium, November 2016

Table of Contents

Background

Vi Agroforestry:

- Swedish NGO
- Reforestation and agroforestry promotion in Western Kenya since 1980s
- Expanded to Bungoma and Kakamega in 2008
- Site chosen for long-term impact evaluation

Background

ICRAF:

- CGIAR Standing Panel on Impact Assessment grant for underevaluated areas
- ICRAF engaged in partnership with Vi through 90s and 2000s
- Extent of ICRAF knowledge transfer is empirical question
- Opportunity to examine long-term household-level impacts of agroforestry

Research Questions

Primary Research Question

What are the downstream socio-economic and land health effects of agroforestry adoption in Kenya?

Sub-Question

Can we improve the rigor of a pseudo-experimental research design by choosing a comparison group using matching over geospatial variables?

Why Use Maps & Remote Sensing?

Why We Needed Spatial Data

1. No baseline data for non-program participants
2. Treatment (the agroforestry extension program) assigned non-randomly
3. Reason to believe geography and landscape play a part in treatment assignment and impacts

Why Use Maps & Remote Sensing?

Why We Needed Spatial Data

1. No baseline data for non-program participants
2. Treatment (the agroforestry extension program) assigned non-randomly
3. Reason to believe geography and landscape play a part in treatment assignment and impacts

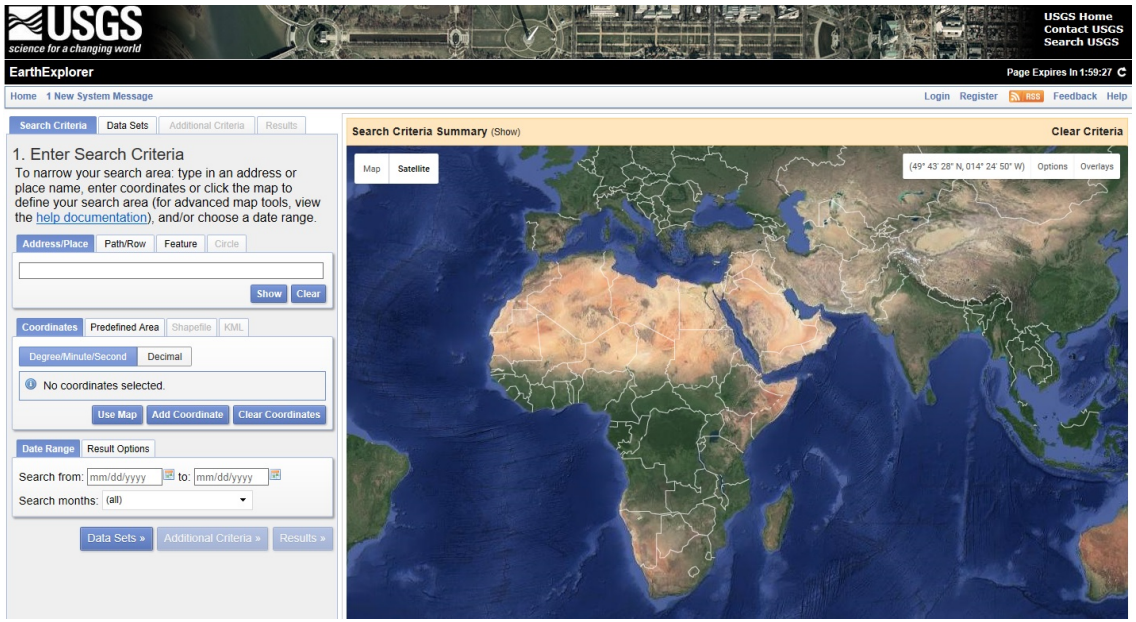
Other Reasons to Look at Spatial Data

- Space matters
- People are visual
- The available data is improving quickly
- Available datasets include information on soil, rainfall, tree cover, population, elevation and more
- Freely available tools

What is Available?

- Geography
  - Global Administrative Areas: gadm.org
  - ILRI GIS database
  - FAO Geonetwork: fao.org/geonetwork
  - OpenStreetMap
  - USGS Earth Explorer: earthexplorer.usgs.gov
  - WorldPop Population Density: worldpop.org.uk
- Forests and Tree Cover
  - Global Land Cover Facility: landcover.org
  - Global Forest Change: earthenginepartners.appspot.com
  - USGS Land Cover: landcover.usgs.gov
- Soil
  - ICRAF Landscapes Portal: landscapeportal.org
  - Africa Soil Information System: africasoils.net
- Climate
  - CHIRPS Precipitation: chg.geog.ucsb.edu/data/chirps
  - NOAA Climate Data Online: ncdc.noaa.gov/cdo-web

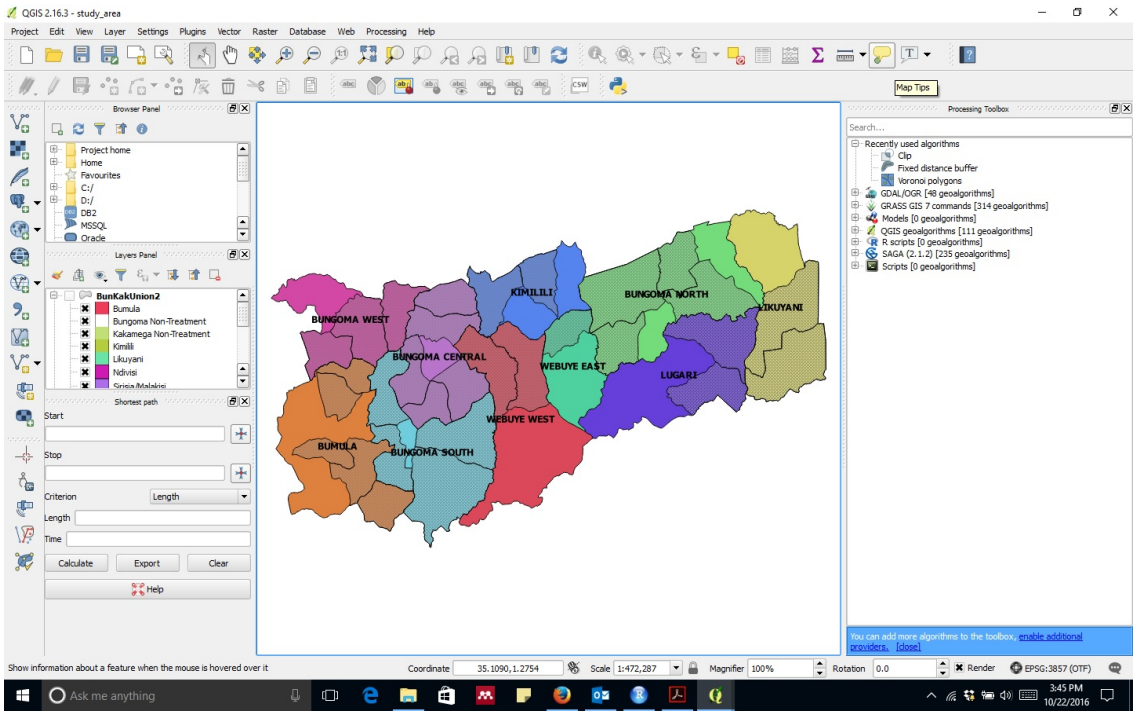
How can I analyse this stuff?



Three kinds of tools:

- Geographic Information System (GIS)
  - QGIS
  - ArcGIS
  - R
- Data Collection Tool
  - ODK Collect
  - ona.io
- Data Analysis Tool
  - R
  - Stata
  - SPSS
  - Excel
  - LibreOffice Calc

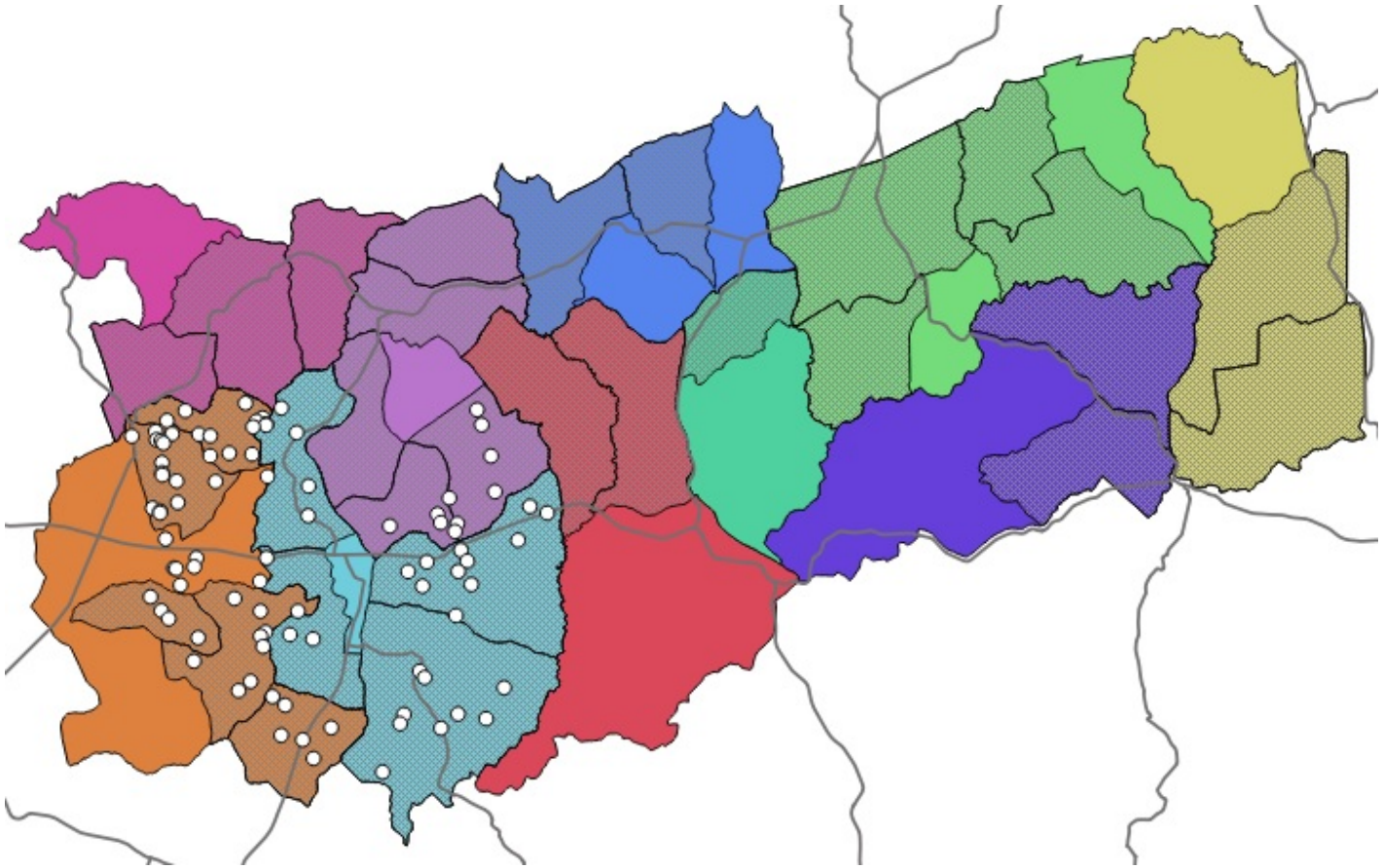
Intro to QGIS



- Free GIS system
- Access to plugins for OpenStreetMap, Google Maps
- Can utilize algorithms from R, GRASS SAGA for geoprocessing
- Can build map projects and assemble map layouts for publication and sharing

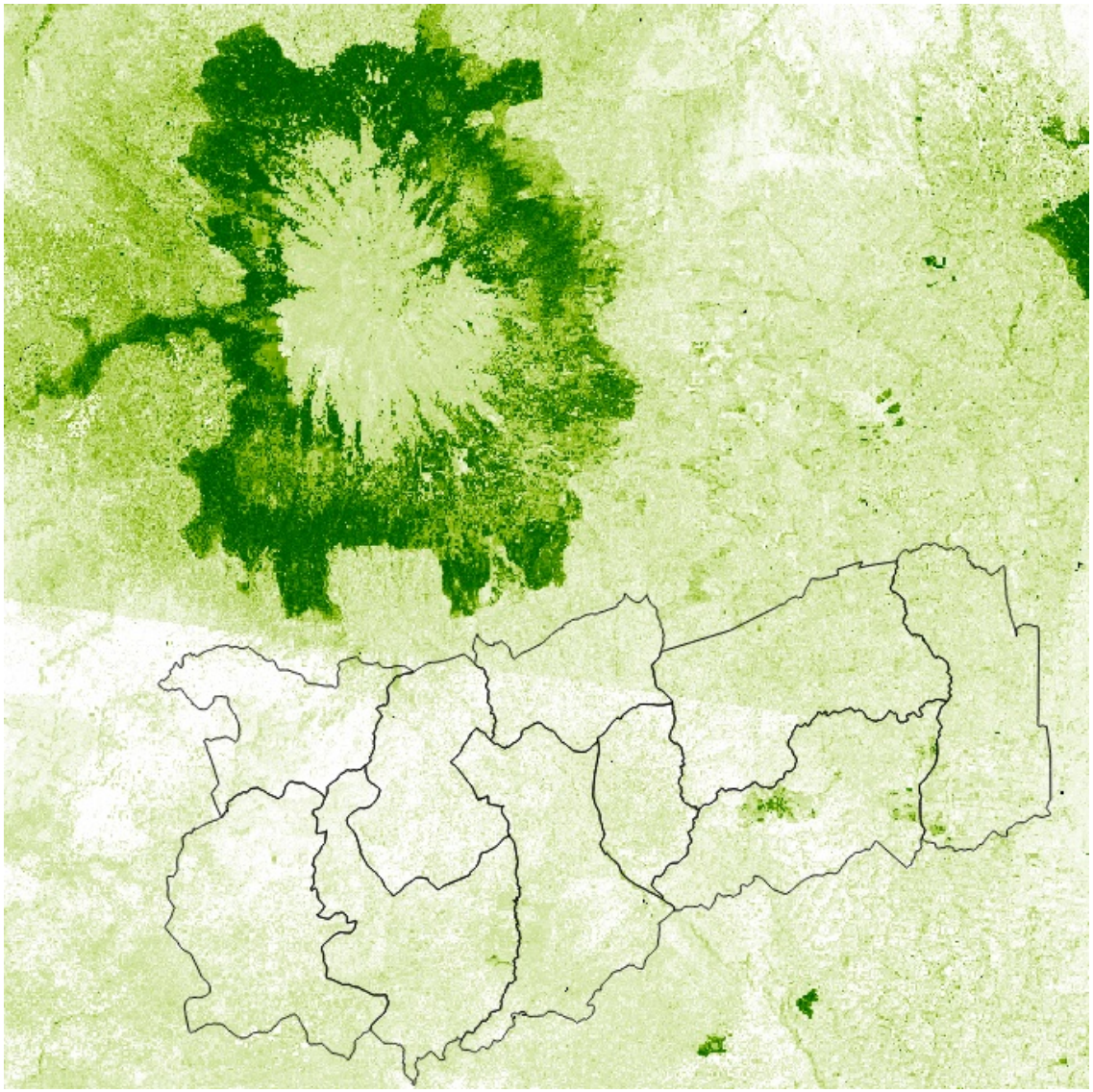
Types of Data

Vector Data: points, lines and ploygons



- Used for depicting roads, borders and centers of things.
- **File Types**: .shp (plus .dbf .shx), .rds

Types of Data

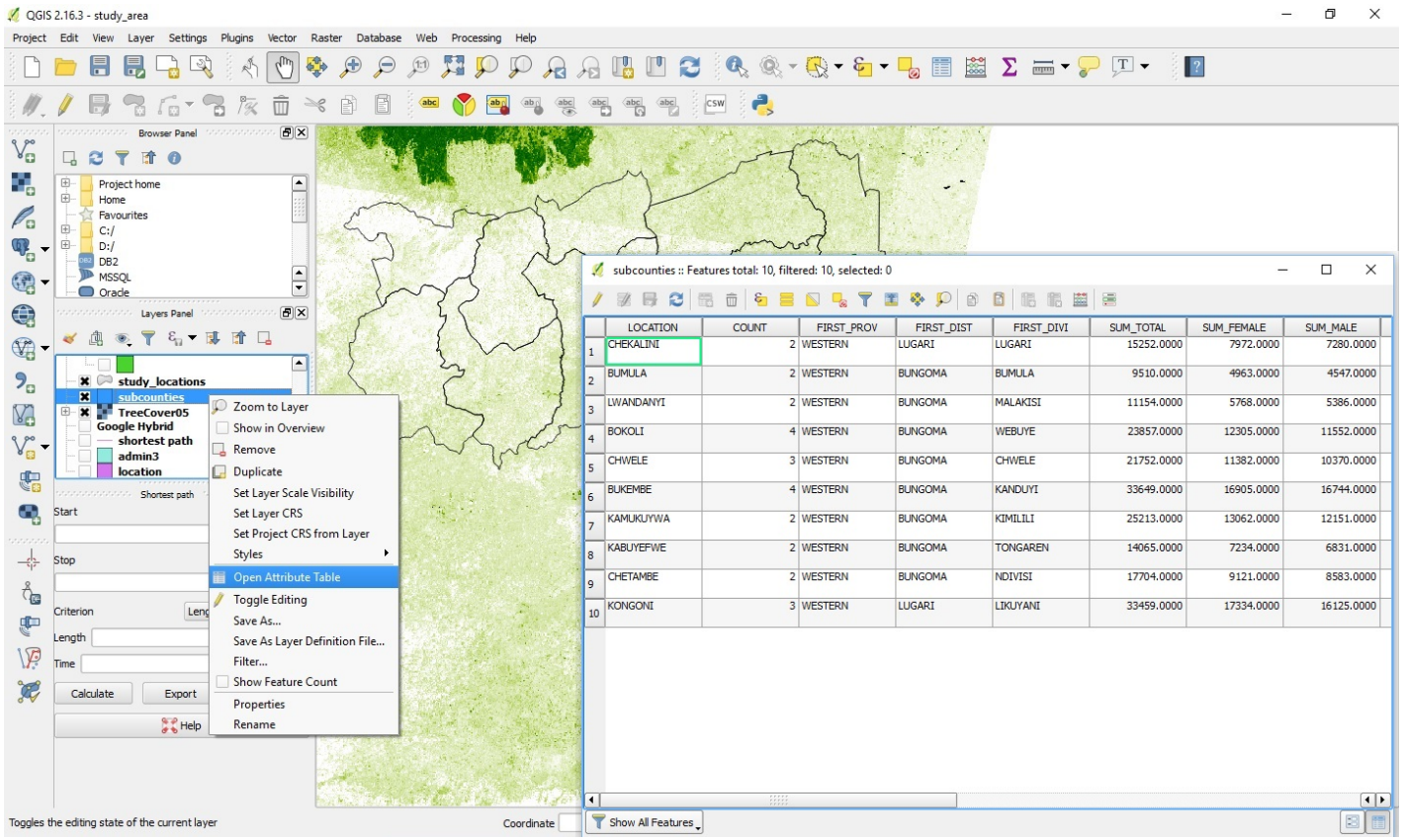


Raster Data: Grid cells with numbered values

- Used for terrain and representing the value of variables like rainfall across space.
- Can overlay with vectors and perform calculations on values
- **File Types**: .tif, .asc, .jpg, .png

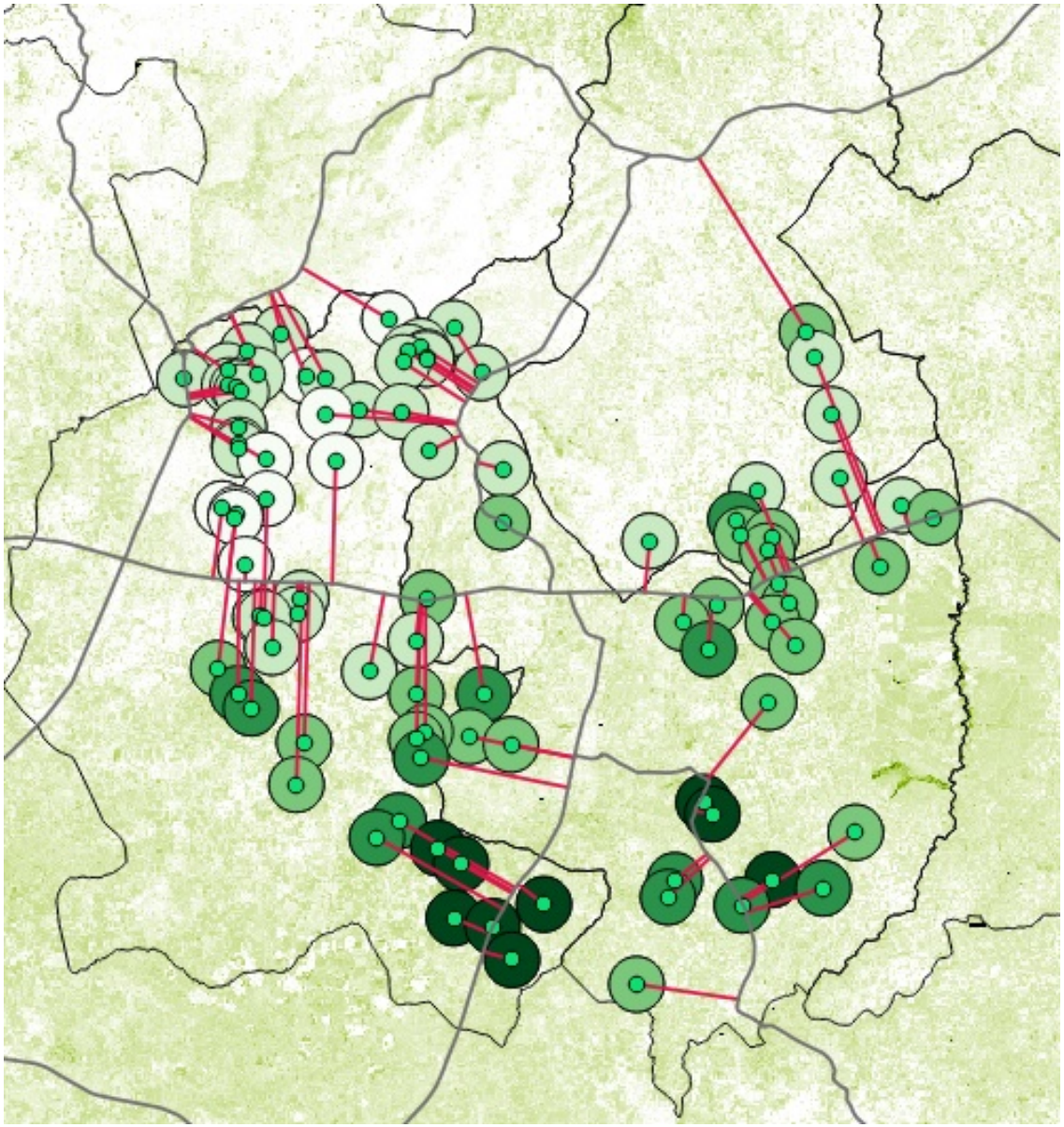
Examining Data

Attribute Table: Spreadsheet data connected to your shapefile



- Data is saved in .dbf file
- Limited manipulation within QGIS
- Can be analysed or edited in Excel or other program

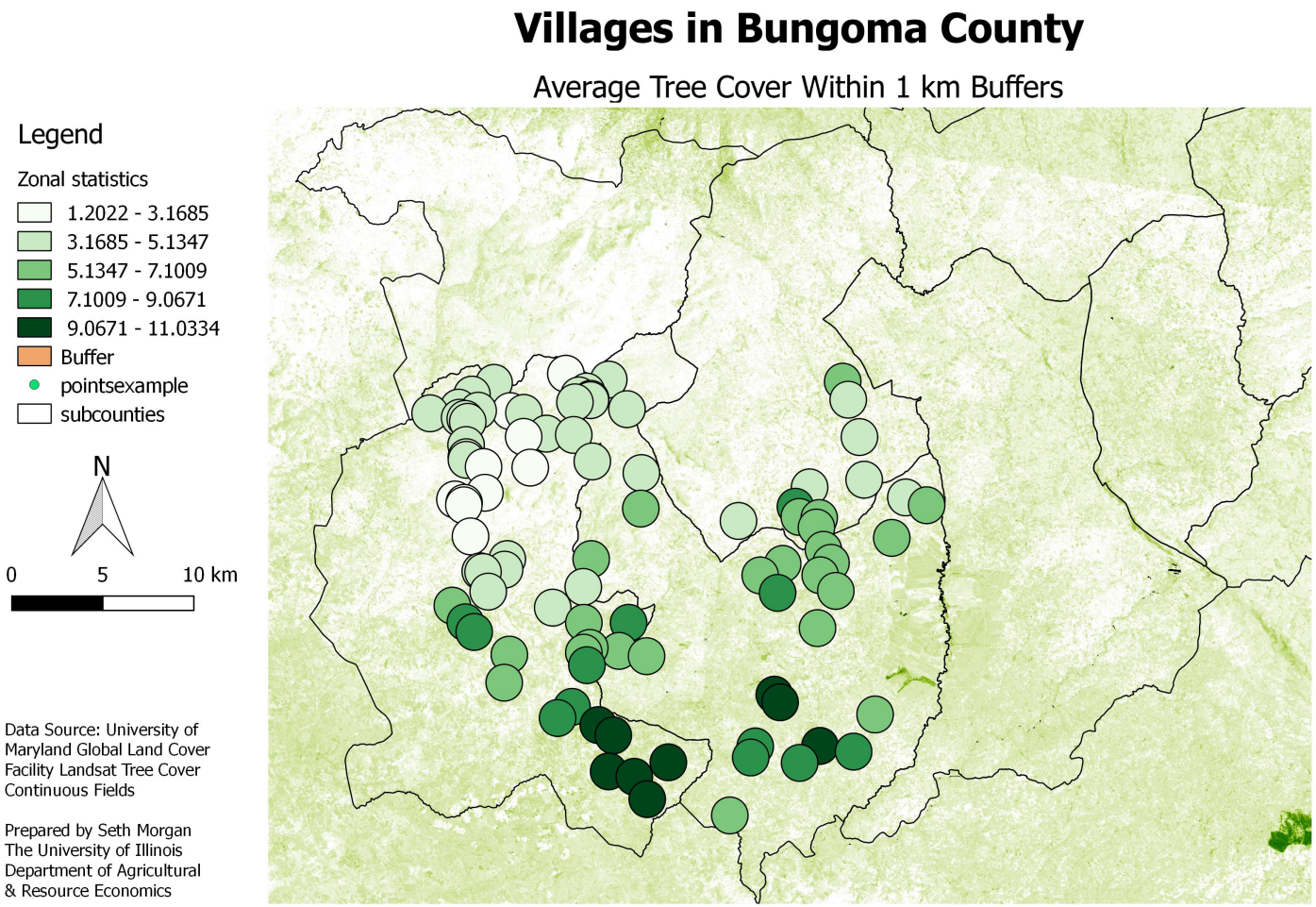
Geoprocessing Functions



Useful Functions:

- Clip
- Buffer
- v.distance (GRASS)
- Zonal Statistics
- v.what.rast.points (GRASS) or Add raster values to points (SAGA)

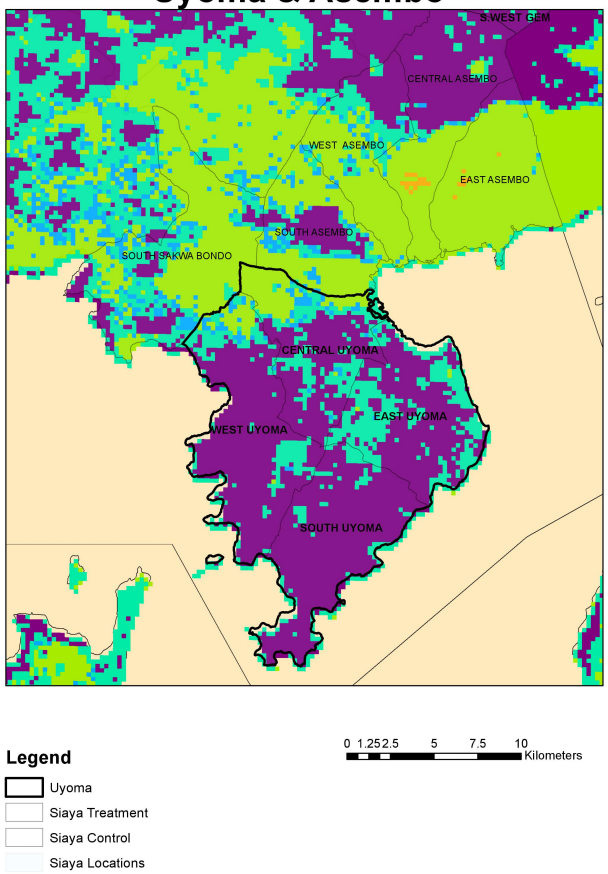
Map Layouts

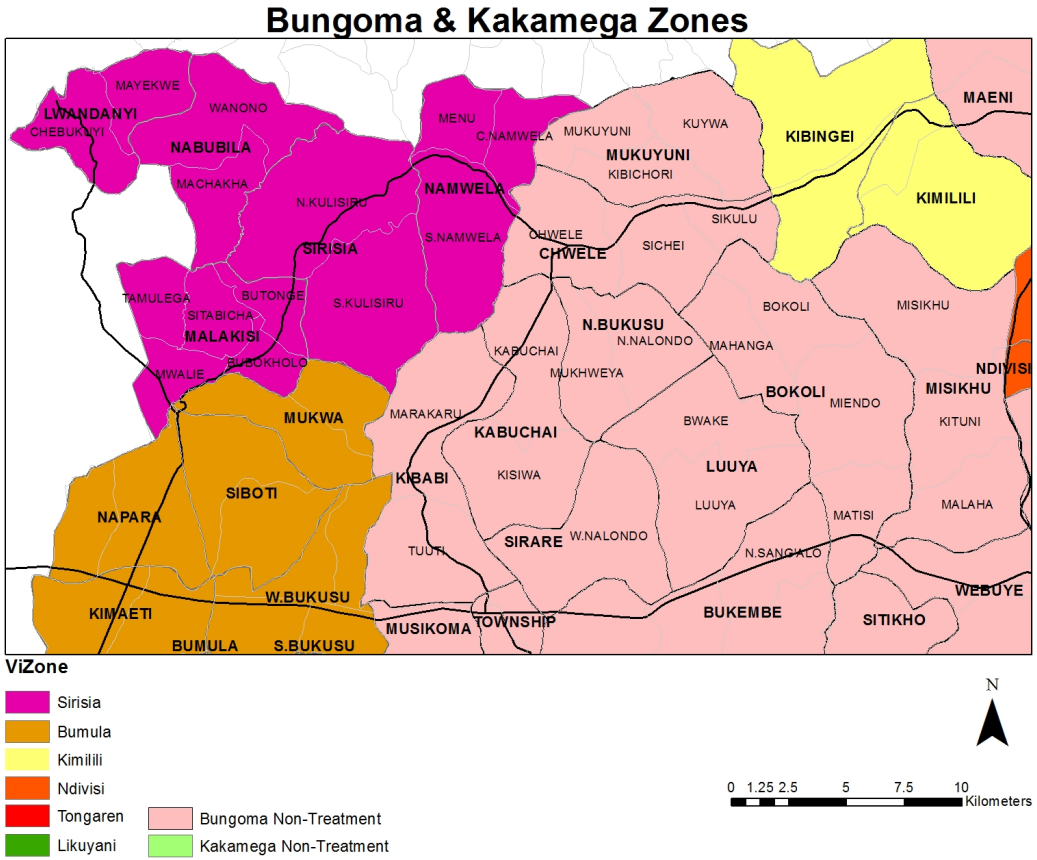


TOSSLAD:

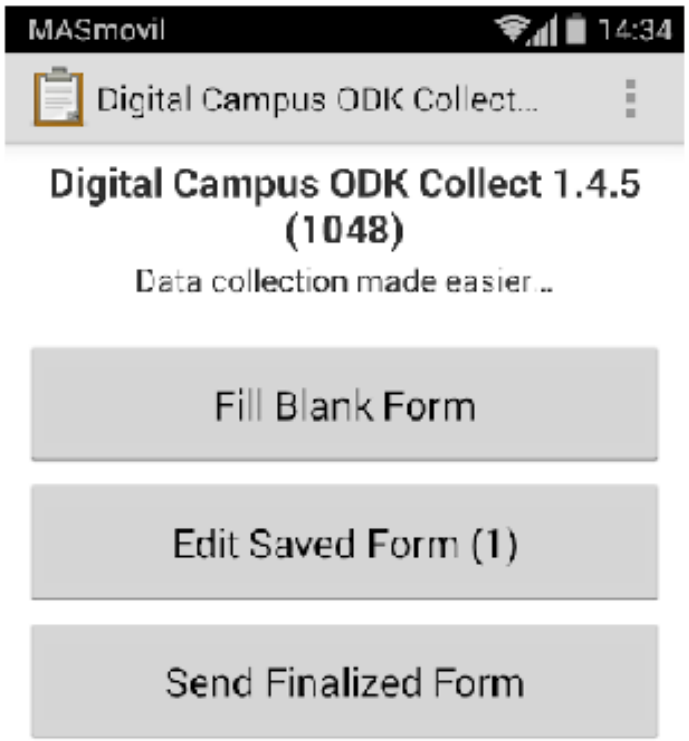
- Title
- Origin
- Source
- Scale
- Legend
- Author
- Date

Example: Site Scoping and Sublocation Matching



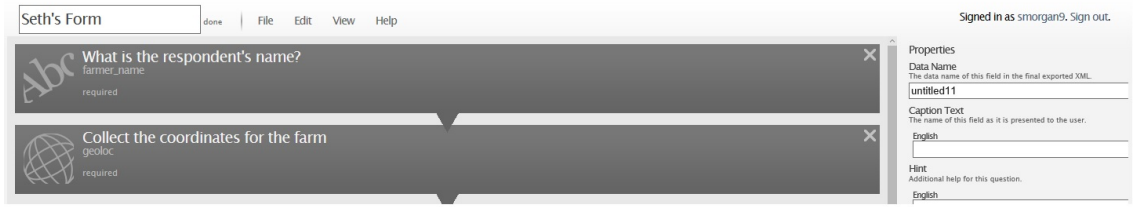


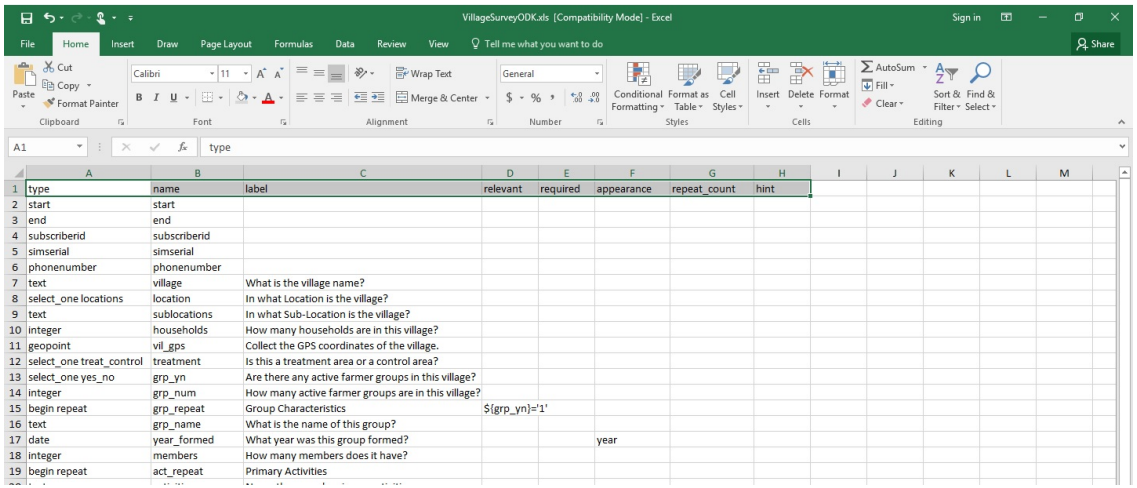
ODK Collect



- Mobile platform for data collection
- Takes geocodes, photos, text, multiple choice
- Works offline, then uploads with network
- Requires Android smartphone or table with SD card

Creating Forms for ODK

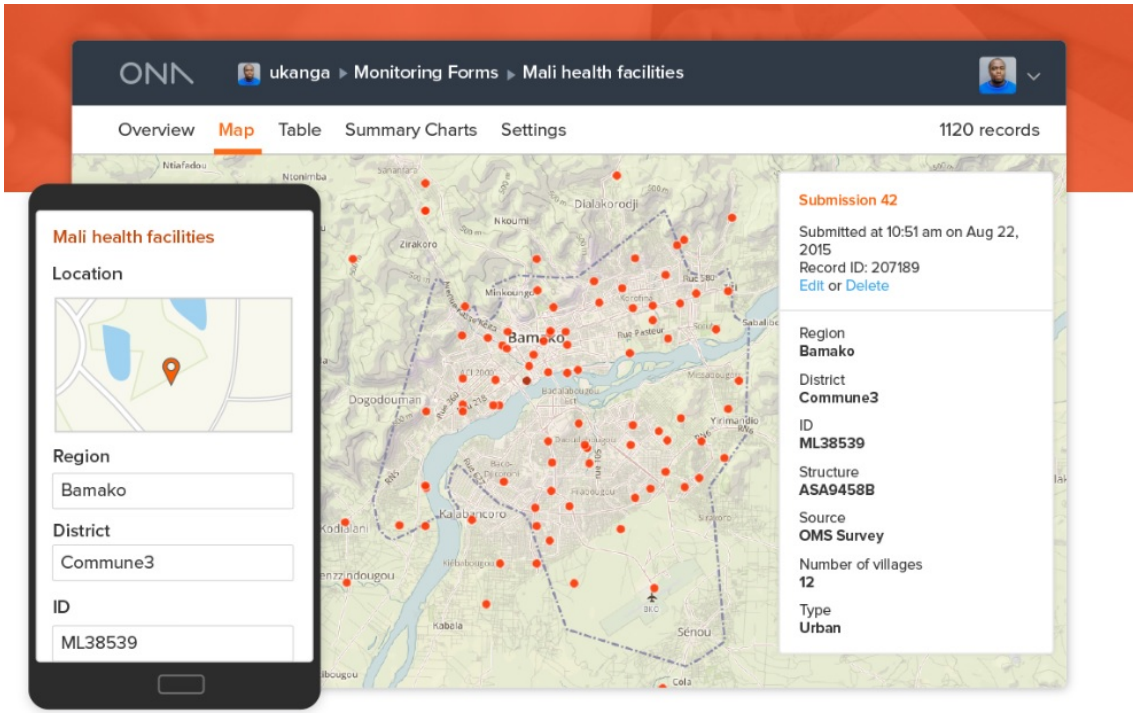




	A	B	C	D	E	F	G	H
1	type	name	label	relevant	required	appearance	repeat_count	hint
2	start	start						
3	end	end						
4	subscriberid	subscriberid						
5	simserial	simserial						
6	phonenumber	phonenumber						
7	text	village	What is the village name?					
8	select_one locations	location	In what Location is the village?					
9	text	sublocations	In what Sub-Location is the village?					
10	integer	households	How many households are in this village?					
11	geopoint	vil_gps	Collect the GPS coordinates of the village.					
12	select_one treat_control	treatment	Is this a treatment area or a control area?					
13	select_one yes_no	grp_yn	Are there any active farmer groups in this village?					
14	integer	grp_num	How many active farmer groups are in this village?					
15	begin repeat	grp_repeat	Group Characteristics	${grp_yn}='1'				
16	text	grp_name	What is the name of this group?					
17	date	year_formed	What year was this group formed?			year		
18	integer	members	How many members does it have?					
19	begin repeat	act_repeat	Primary Activities					

- build.opendatakit.org provides free form builder app
- For longer or more complex forms, create in Excel or LibreOffice
- Instructions available at xlsform.org
- Upload to ona.io or opendatakit.org/xiframe to convert for use in ODK Collect

Aggregating data on ona.io



- Free account for public data
- Collect and store data in the cloud
- Display geocoded data on a map
- Export survey data as a spreadsheet

Alternatives:

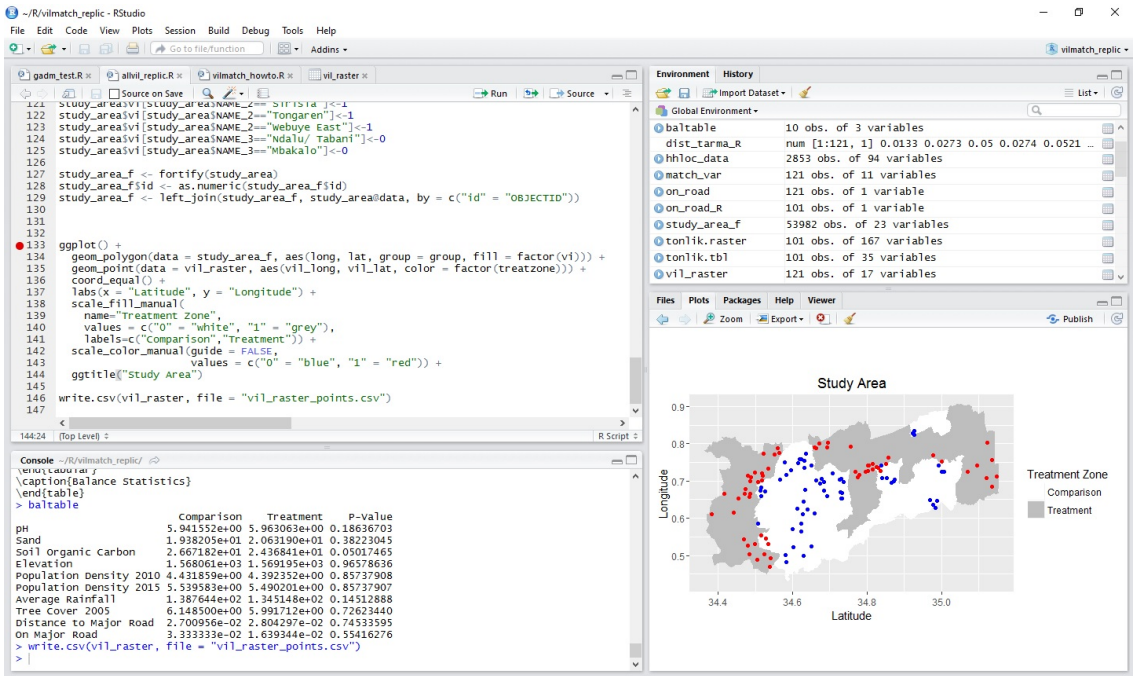
- KoBo Collect
- CommCare
- ODK Briefcase

Example: Village Data Collection

Village Survey Questions

- How many households in the village?
- How many active farmer groups?
- When were the groups formed?
- What activities do they pursue?
- Geocode collection

What is R?



"A language and environment for statistical computing and graphics"

Essential Packages:

- rstudio
- dplyr
- readr
- tidyr
- tidyverse
- ggplot2

R Pros and Cons

Why Use R

- Extremely adaptable
- Widely available
- Results are reproducible
- Enthusiastic user community
- Growing library of packages and documentation

Why Not to Use R

- Not intuitive for beginners
- Minimal GUI: have to learn some code
- Difficult to examine data
- No centralized help system

Example: Village Matching

Objective: assemble a sampling frame made up of villages in which treatment and comparison villages are statistically indistinguishable across the following variables:

- Number of Households
- Average Soil Sand Content (Vågen et al., 2016)
- Average Soil pH (ibid.)
- Average Soil Organic Carbon (ibid.)
- Average Tree Cover in 2005 (Sexton et al., 2013)
- Elevation (Jarvis et al., 2008; Kruska and Kariuki, 2016)
- Average Population Density in 2010 and 2015 (Stevens et al., 2015)
- Average Rainfall (Funk et al., 2015)
- Distance to Tarmac Road
- Binary for Villages 0.25 m from Tarmac Road ("on road")
- Binary for presence of microfinance activities

Village Matching Model

Propensity Score: The probability of receiving treatment, conditional on the covariates.

$$e(X_i) = Pr(T_i = 1|X_i)$$

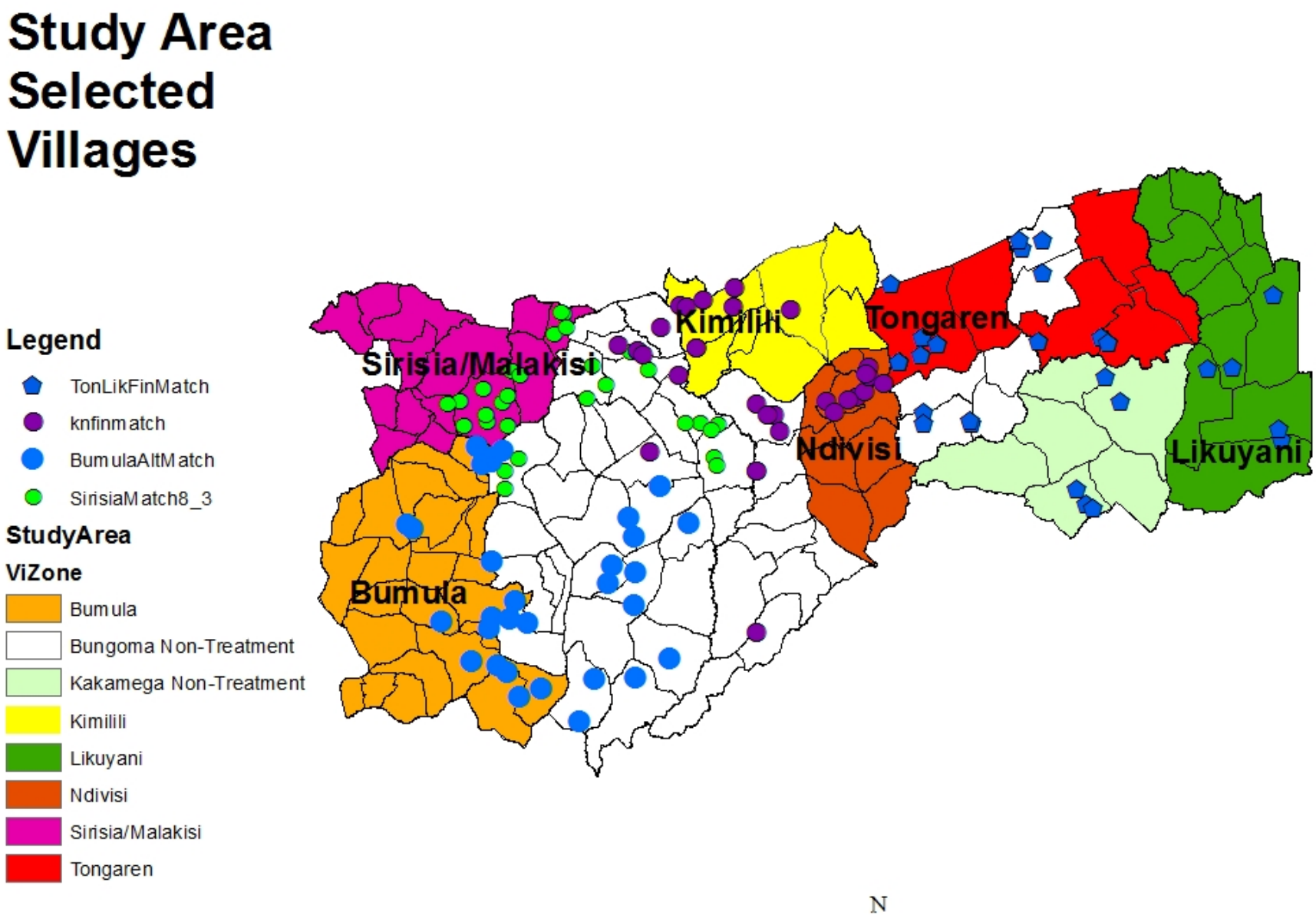
The score is generated with a probit regression, a regression model in which the outcome is binary (0/1, Treatment/Control):

$$z = X\beta + \epsilon$$

Where z is an unobserved variable and y is the observed binary corresponding to treatment assignment such that:

$$y_i = \begin{cases} 1 \text{ if } z_i \geq 0 \\ 0 \text{ if } z_i \leq 0 \end{cases}$$

Matching Results: Selected Village Points

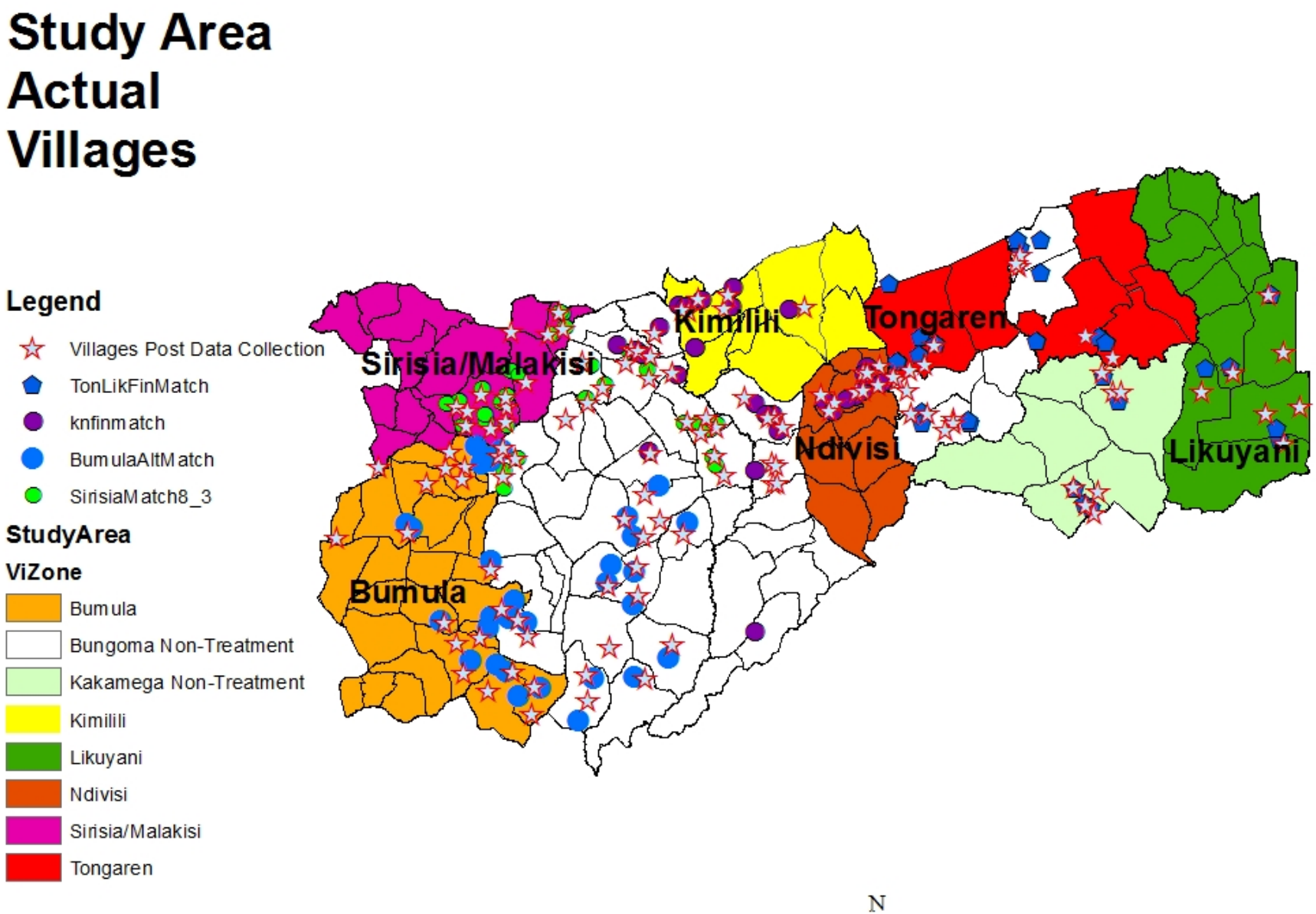


Matching Results: Selected Village Points

Table: Balance Statistics for Selected Villages

	Mean Difference	P-Value
Households	16.83	(0.82)
Sand	-176.0	(-1.25)
pH	-1.950	(-1.14)
Tree Cover 2005	-0.134	(-0.30)
Elevation	-12.90	(-0.49)
Population 2010	0.0789	(0.36)
Soil Organic Carbon	27.20	(0.17)
Average Rainfall	3.614	(0.99)
Distance too Major Road	-0.000186	(-0.06)
On Major Road	0.0323	(0.58)

Matching Results: Actual Village Points



Matching Results: Actual Village Points

Table: Balance Statistics Post Data Collection

	Comparison	Treatment	P-Value
pH	5.94	5.96	0.19
Sand	19.38	20.63	0.38
Soil Organic Carbon	26.67	24.37	0.05
Elevation	1568.06	1569.20	0.97
Population Density 2010	4.43	4.39	0.86
Population Density 2015	5.54	5.49	0.86
Average Rainfall	138.76	134.51	0.15
Tree Cover 2005	6.15	5.99	0.73
Distance to Major Road	0.03	0.03	0.75
On Major Road	0.03	0.02	0.55

Lessons Learned

Lessons for Impact Evaluations

- Scoping takes time
- Take the time to get all stakeholders on board
- Recruiting comparisons takes 2-3 times the investment as treatment
- Respect respondents' wishes and expectations, even when you can't fulfil them
- Remote sensing data is useful, but not perfect
- Use remote sensing to complement local information, not supplant it
- Take note of opportunities for capacity building with local staff

Thanks

The University of Illinois at Urbana Champaign Department of Agricultural and Consumer Economics

The World Agroforestry Centre

Vi Agroforestry Kenya

Borlaug Fellowship in Food Security

Dr. Katherine Baylis (UIUC)

Dr. Karl Hughes (ICRAF)

Questions and Follow-up

Contact

smorgan9@illinois.edu
sethhenrymorgan@gmail.com

Bibliography

- Funk, Chris et al. (2015). "The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes". In: *Scientific Data* 2. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.66.
- Jarvis, A et al. (2008). *Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database*.
- Kruska, Russell and Patrick Chege Kariuki (2016). *ILRI - GIS Services Digital Elevation Model*.
- Sexton, J.O. et al. (2013). *Landsat Tree Cover Continuous Fields*. DOI: 10.1080/17538947.2013.786146.
- Stevens, Forrest R. et al. (2015). "Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data". In: *PLoS ONE* 10.2, pp. 1–22. ISSN: 19326203. DOI: 10.1371/journal.pone.0107042.
- Vågen, Tor-G. et al. (2016). "Mapping of soil properties and land degradation risk in Africa using MODIS reflectance". In: *Geoderma* 263, pp. 216–225. ISSN: 0016-7061. DOI: http://dx.doi.org/10.1016/j.geoderma.2015.06.023.