The Joys and Sorrows of Variable Construction

Ayala Wineman

MwAPATA Seminar

July 29, 2021

MICHIGAN STATE
UNIVERSITY

Some common variables

Fertilizer use (binary indicator)

Rate of fertilizer use in the population (% of farms)

Fertilizer intensity (kg/hectare)

Farm size

Crop harvest (kg, farm level)

Crop harvest (kg, country level)

Crop yield (kg/hectare)

Value of crop production (MK)

Considerations in variable construction

Define the variable

What source of information will you trust?

Nonstandard units of measure

What belongs in the numerator and denominator?

Clean the data
- What counts as an outlier?
- How to identify outliers
- How to deal with outliers

Define the variable: Medium-scale farm

RESEARCH QUESTION: Are there any spillover effects from medium-scale farms, such that small-scale farms benefit from, or are harmed by, their medium-scale neighbors?

**For Discussion:
How would you define
a "medium-scale farm"?**

Define the variable: Medium-scale farm

- What criteria would you use to categorize a farm as being "medium-scale"?

- Do you want the definition to reflect the aggregate scale of cropping, livestock/fish production, or both activities?

- Do you want the definition to reflect the *potential* for production (the asset base held by a household) or the *actual* production (the economic scale of crops and livestock products produced)?

- What information do you actually have in the data set?

- Should the definition be consistent with the definition applied by other analysts, perhaps in other countries?

Define the variable: Medium-scale farm

LAND SIZE
5-20 hectares
Houssou et al. 2016 (Ghana);
Lay et al. 2018 (Zambia)

>10 hectares
Deininger and Xia 2016 (Mozambique);
Ali et al. 2017 (Ethiopia)

5-50 hectares
Anseeuw et al. 2016 (Malawi)

5-100 hectares
Jayne et al. 2016, 2019, 2021 (multiple countries)

100-1,000 hectares
Ango 2018 (Ethiopia)

Vegetable growers that rent in at least 0.5 hectares
Bachewe and Minten 2020 (Ethiopia)

Land held
Jayne et al. 2016 (multiple countries)

Land cultivated
Houssou et al. 2016 (Ghana)

Land cultivated, cultivation of high-value crops, livestock holdings
Wineman et al. 2020a (Tanzania)

For Discussion:
What are the implications for systematic reviews of evidence / meta-analyses?

5

What source of information to use?

RESEARCH QUESTION: What is the yield premium of improved maize seed?

G0A	G0B	GOB_1	GOB_2
What is the primary variety of [CROP] being cultivated on this [PLOT]? ENUMERATOR: THIS IS ONLY ASKED FOR SPECIFIC CROPS.	Is the variety local or improved ?	Is the variety recyclable?	When did you last buy the seed you planted on this [PLOT]?
VARIETY CODE	LOCAL......1 >>G0B_2 IMPROVED...2	YES...1 NO....2	**YEAR (4-DIG.)**

How accurately do farmers report the type of maize seed they've planted?

DNA results

Farmer report
↓

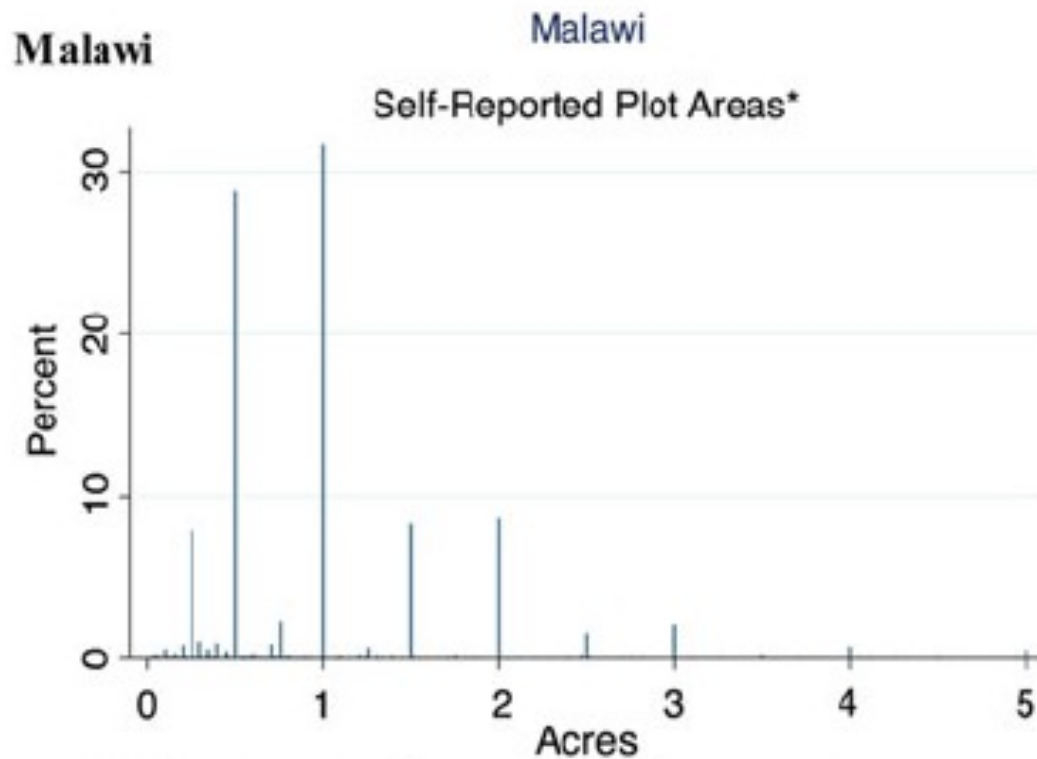|	Local	Improved
Local	13%	14%
Improved	16%	57%

The yield premium in Tanzania...

Using farmer reports   503 kg / hectare
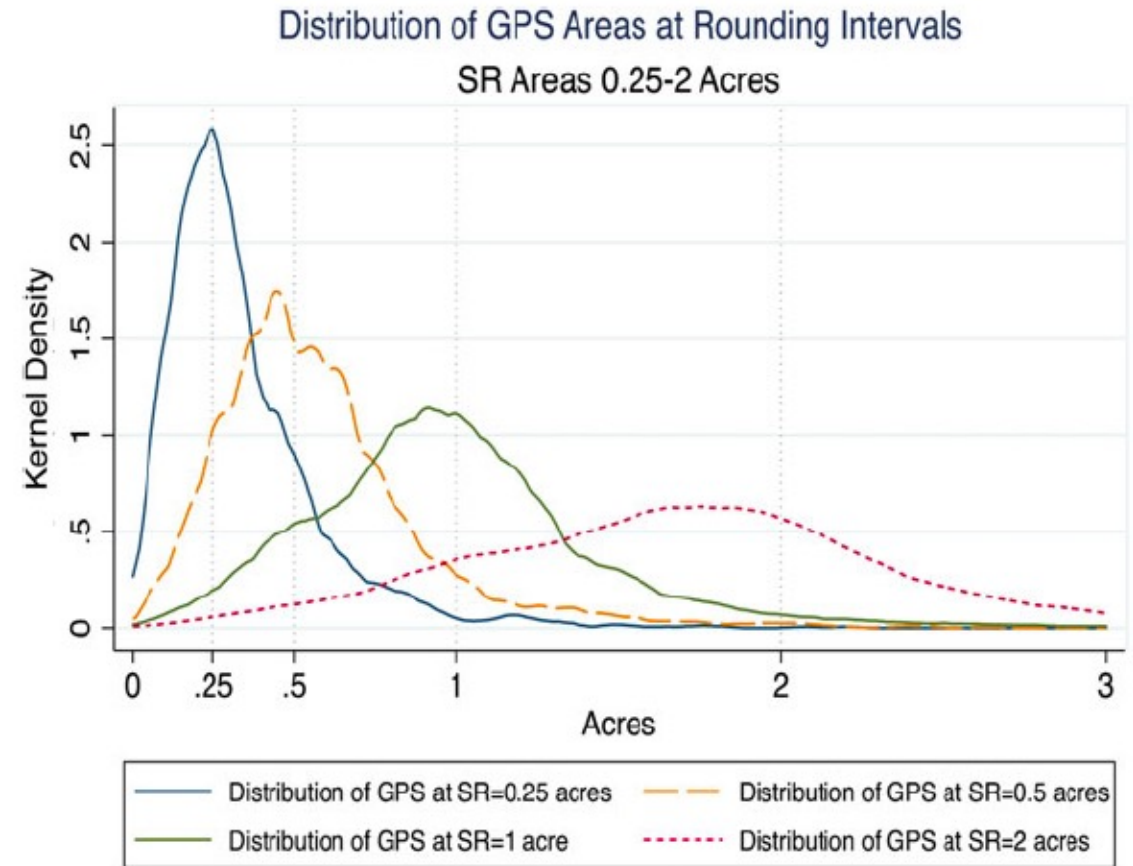
**For Discussion:
Does recycled seed ever stop being considered as "improved" in the minds of farmers?**

Source: Wineman et al. 2020b

What source of information to use?

TOPIC: Field size (Garden, plot, farm size)

Source: Carletto et al. 2015

What source of information to use?

Self-Reported Area Measurement
Bias as % of GPS Area
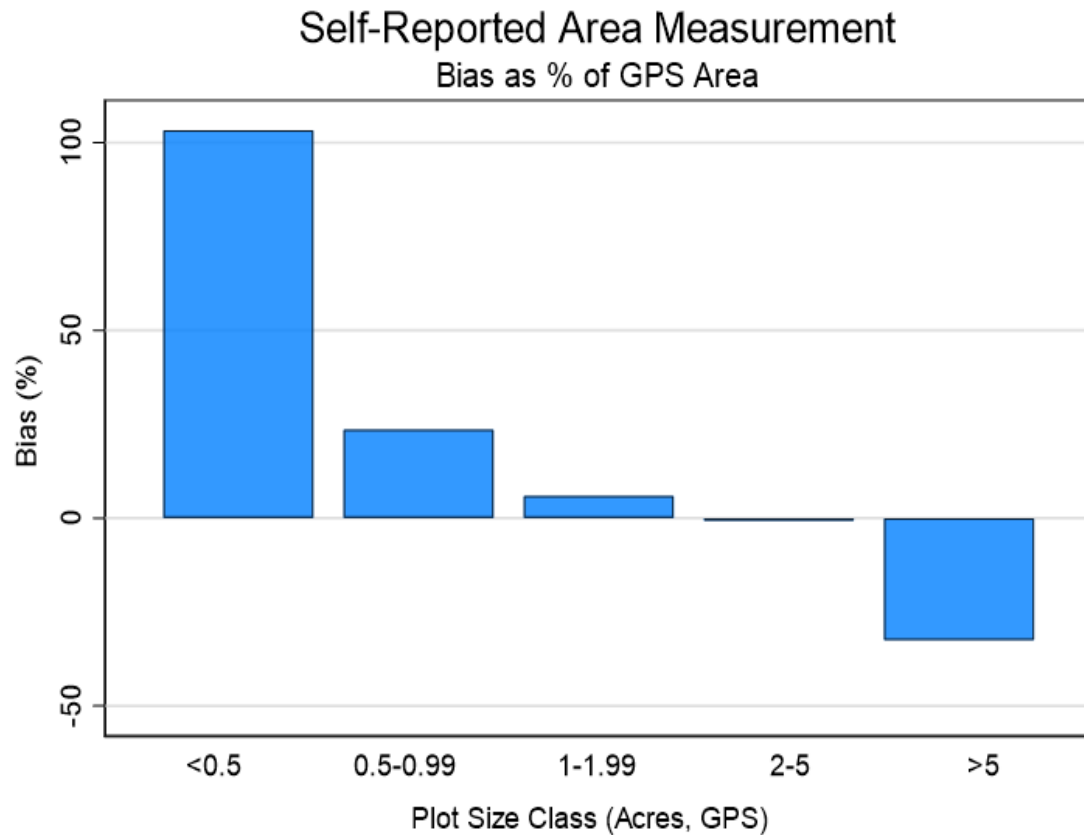
For Discussion:
Which source of information on field size would you use?

What if just 80% of the fields are GPS'd?

Source: Carletto et al. 2015

Nonstandard units of measure

Nonstandard units of measure

For Discussion:
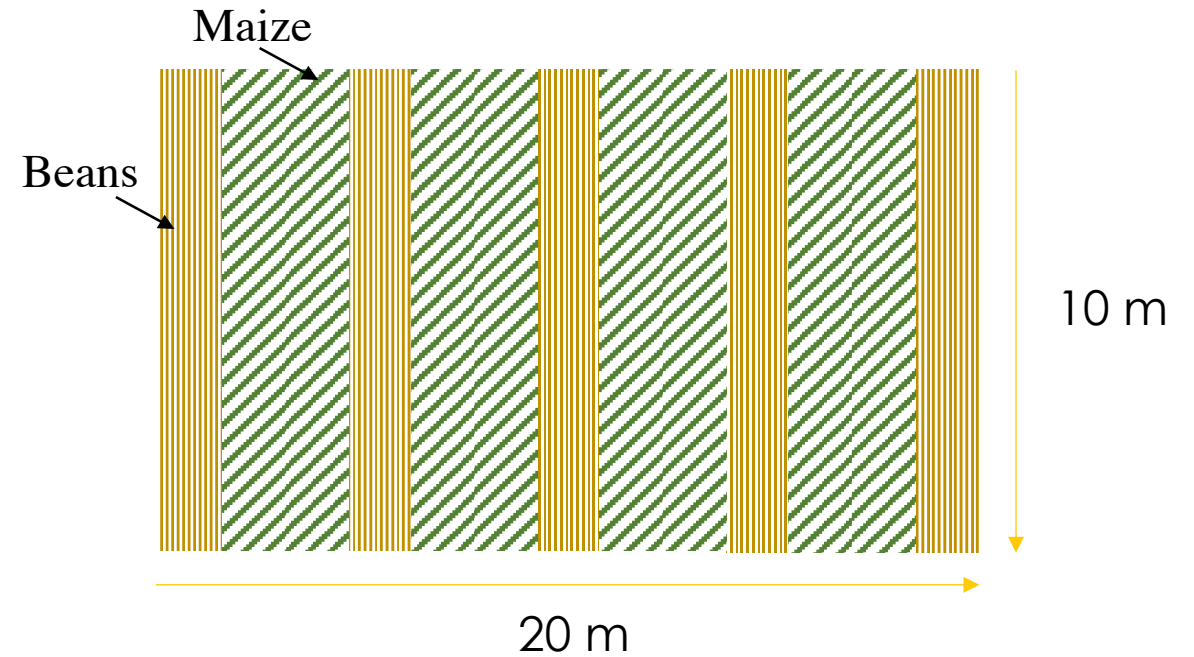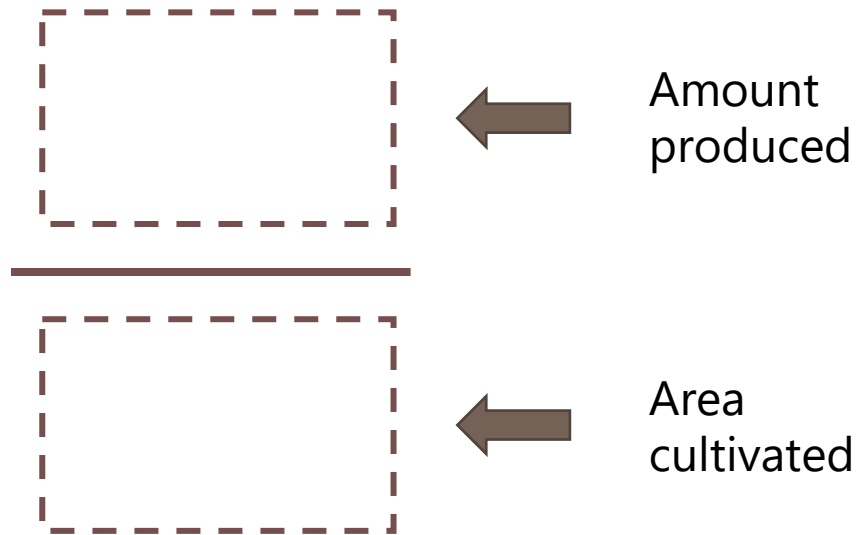What do you do when a conversion factor is missing?

...How many kilograms correlate to a pickup truck?

Conversion factors from the World Bank

tuber of yam: small (3 kg)
tuber of yam: medium (5 kg)
tuber of yam: big/large (8 kg)
bundle of millet, g/corn, sugarcane, vegetable etc: small (1
bundle of millet, g/corn, sugarcane, vegetable etc: medium
bundle of millet, g/corn, sugarcane, vegetable etc: big (40
wheel barrow: small (60 kg)
wheel barrow: medium (85 kg)
wheel barrow: big/large (110 kg)
wheel barrow: extra large (150 kg)
pick-up van: small (1,500 kg)
pick-up van: medium (2,000 kg)
pick-up van: big (2,500 kg)

What belongs in the denominator of crop yield?


Amount produced

Area cultivated

Maize

Beans

10 m

20 m

For Discussion:
What is the yield of maize?

Total maize production: 400 kg
Total bean production: 100 kg

What belongs in the denominator of crop yield?

G01	G02	G03
What type of crop stand was on the [PLOT]?	Was the [CROP] planted in the entire area of the plot?	Approximately, how much of the [PLOT] is under [CROP]?
READ RESPONSES		READ RESPONSES
Pure Stand/ Sole ......1>>4		
		Less than 1/4.......1
Strip Intercrop..2		
		1/4.......2
Row Intercrop..3		
		1/2.......3
Relay Intercrop..4		
		3/4.......4
Mixed Intercrop..5	YES...1>>4 NO....2	More than 3/4.......5
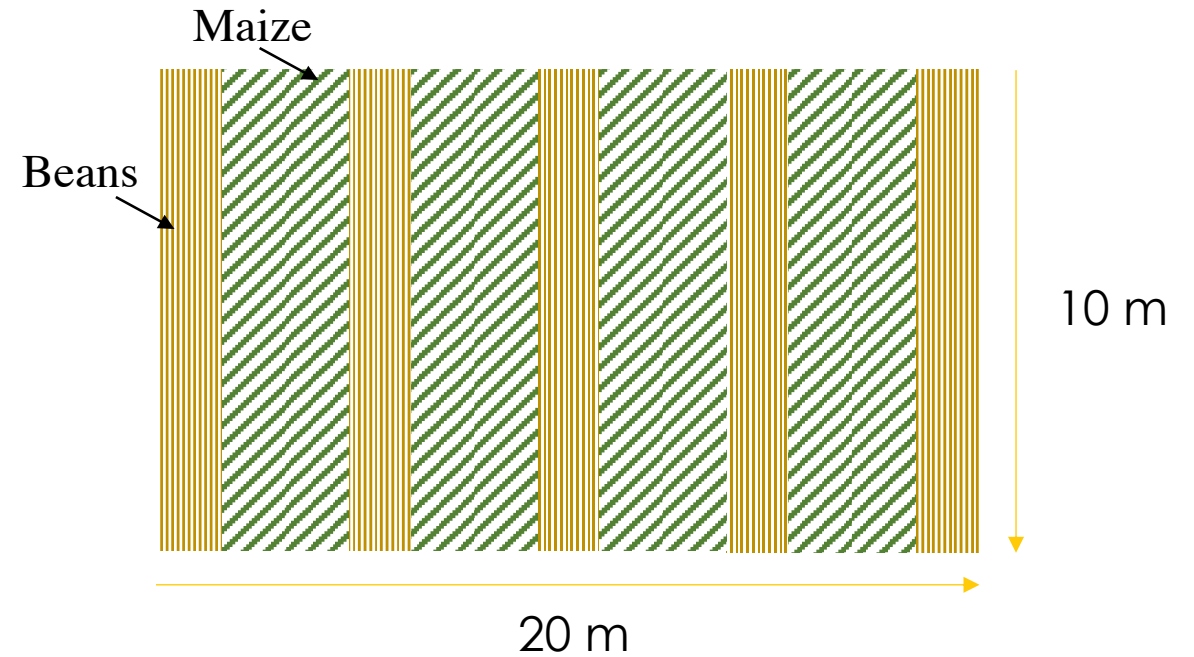
Maize

Beans

10 m

20 m

Total maize production: 400 kg
Total bean production: 100 kg

What belongs in the denominator of crop yield?

Methods used to estimate the area under crops

Area 1: Use the entire plot size

Area 2: Plot size * proportion cultivated with crop i

Area 3: Plot size divided by number of crops

Area 4: Area under monocrops estimated as in Area 2
For intercropped crops, if their summed areas exceed the residual plot area that is not monocropped, these areas are scaled down proportionally

Maize

Beans
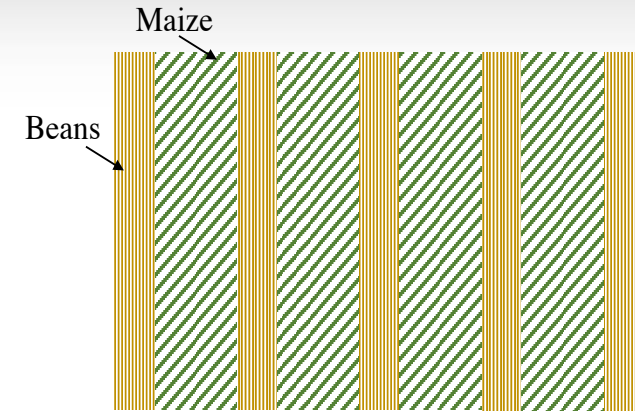
Yields in Tanzania (kg/ha, mean values)

	Maize	Rice
Yield 1	1,067	1,554
Yield 2	1,292	1,756
Yield 3	1,992	1,803
Yield 4	1,865	1,810

### Calorie productivity across crops
(calories/ha = mean yield * calories/kg)

	Maize	Rice	Most calorie-productive crop
Yield 1	3.81 million	5.55 million	Rice
Yield 2	4.61 million	6.27 million	Rice
Yield 3	7.11 million	6.44 million	Maize
Yield 4	6.66 million	6.46 million	Maize

Source: Wineman et al. 2019

What belongs in the denominator of land productivity?

RESEARCH QUESTION: Is there a trend over time in land productivity in Tanzania?

Value of crops produced
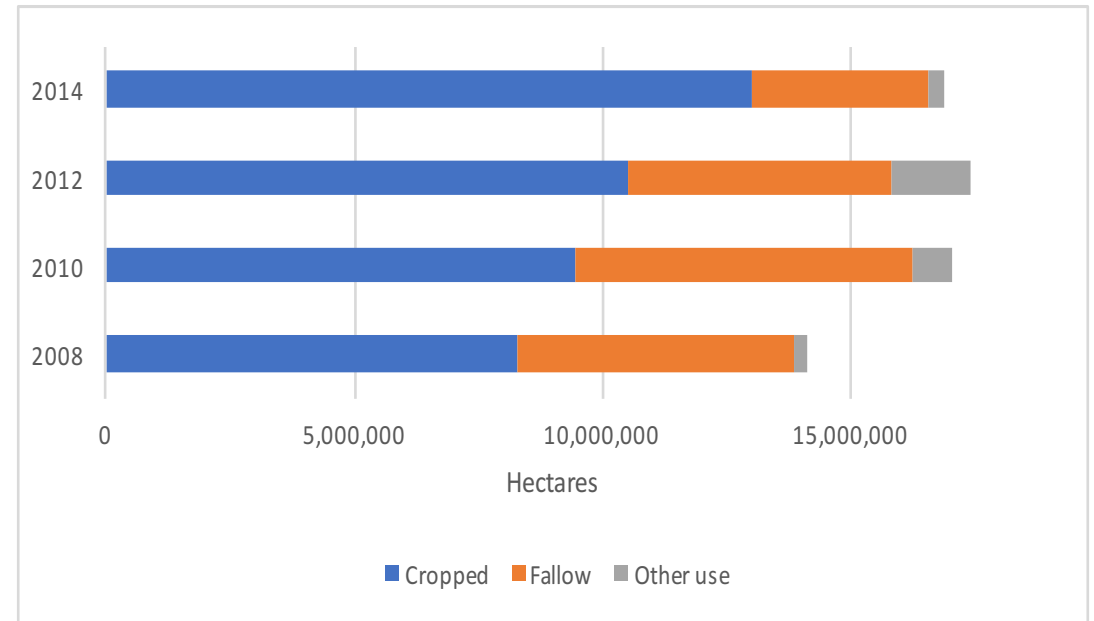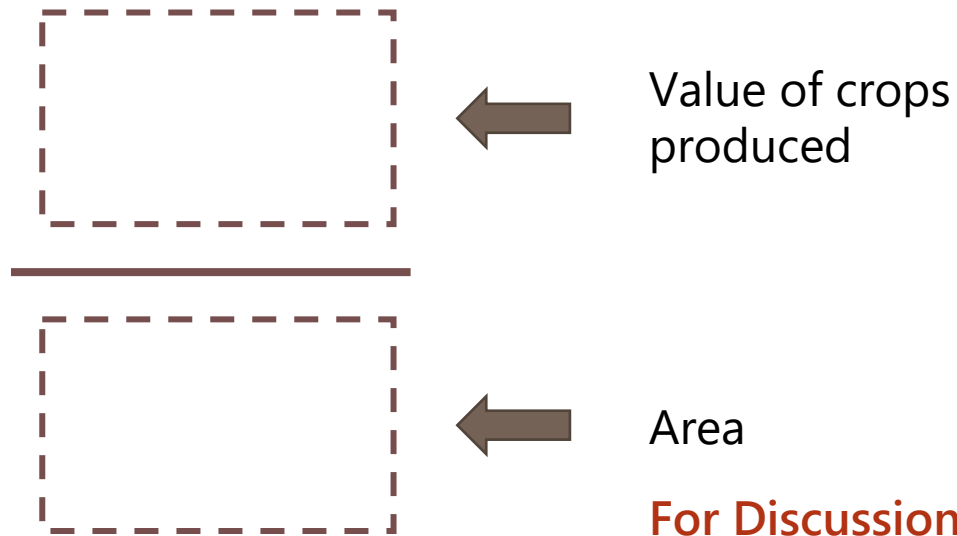
Area

What belongs in the denominator of land productivity?

RESEARCH QUESTION: Is there a trend over time in land productivity in Tanzania?

|	2008	2010	2012	2014
Value crop production (billions shillings, real 2015 values)	3,211	3,890	4,506	5,135
Area cropland (millions hectares)	13.59	15.68	15.24	16.30
Value crop production / hectare (millions shillings)	0.24	0.25	0.30	0.31
Area cropped in main season (millions hectares)	8.31	9.46	10.51	13.00
Value crop production / hectare (millions shillings)	0.39	0.41	0.43	0.40

Source: Wineman et al. 2019
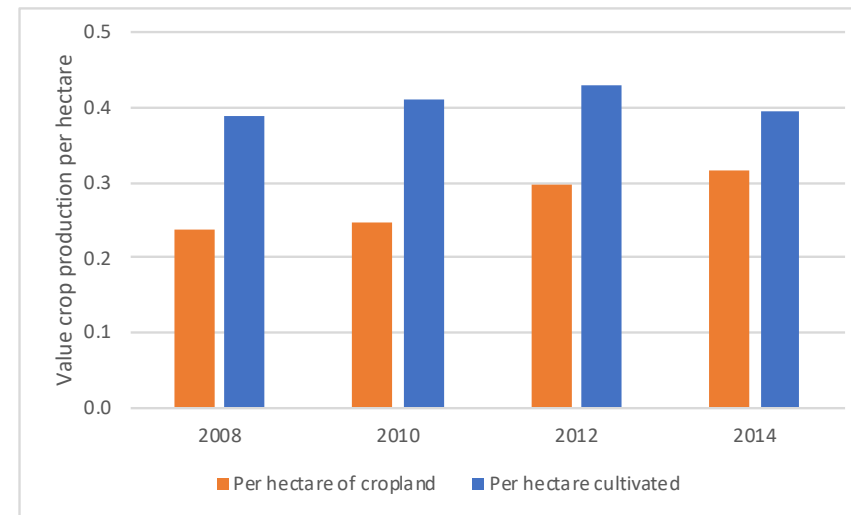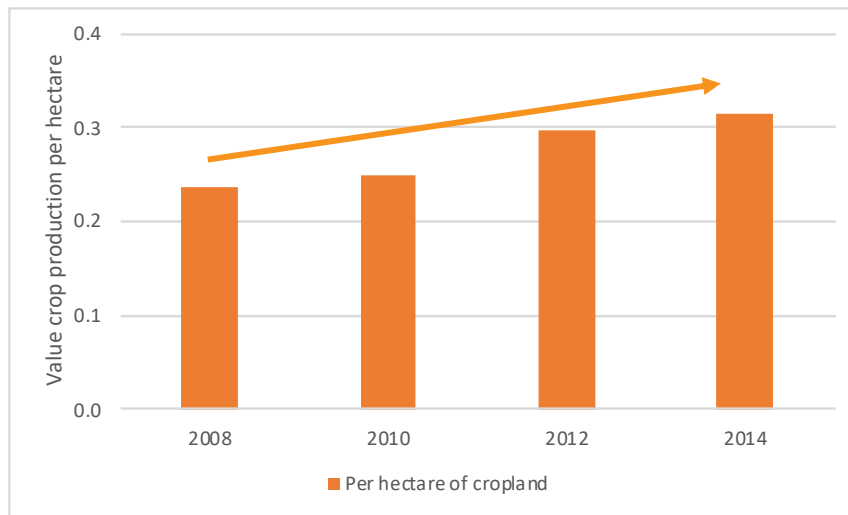
Data cleaning

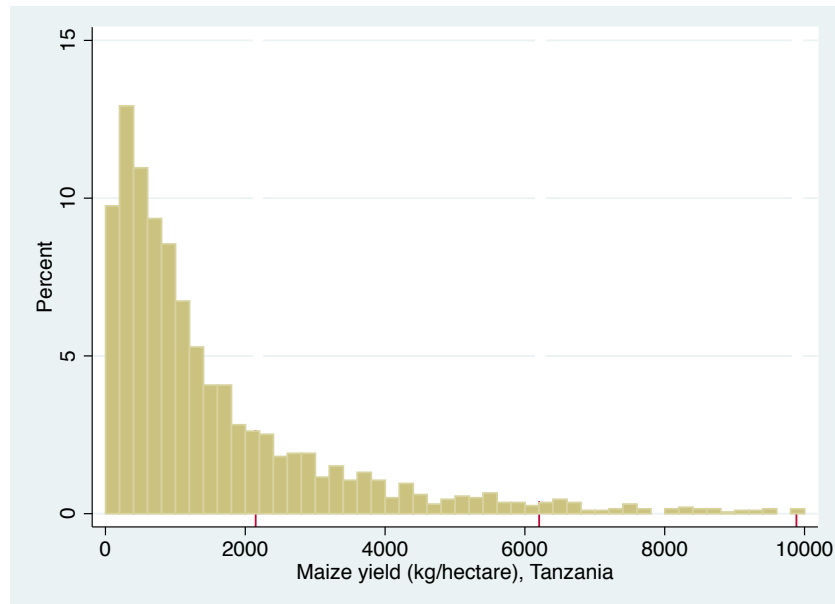Thom Mpinganjira

Bill Gates

Tea estate

For Discussion:
What counts as an outlier?
- Unreasonable (unrealistic) value
- Extreme (albeit reasonable) value

Data cleaning

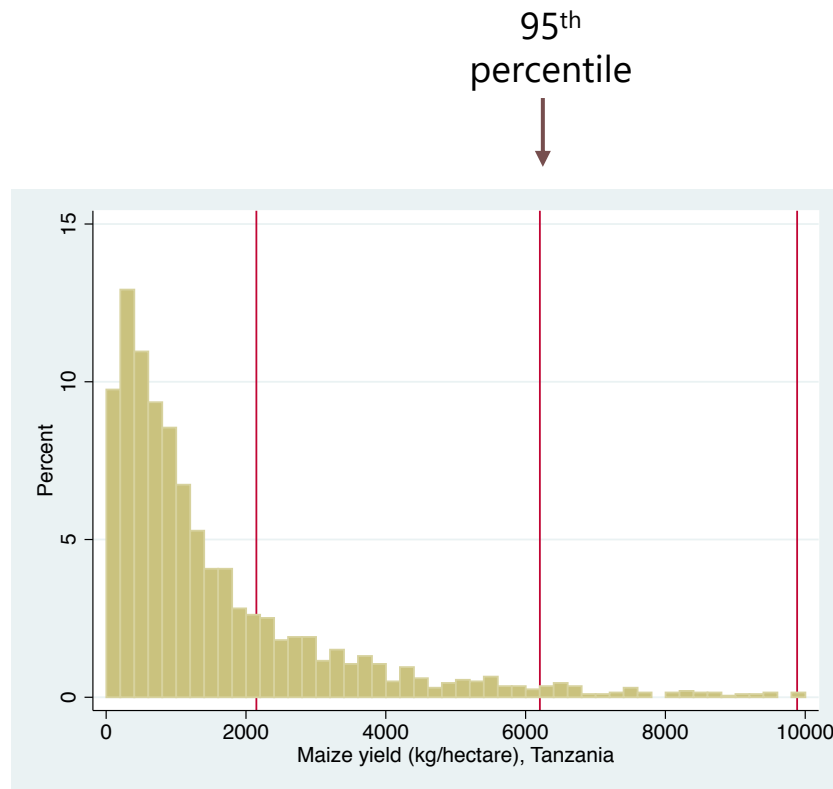2.5 times the median absolute deviation (~500) above the median yield of ~900 kg per hectare

95th and 98th percentile

Percent

Maize yield (kg/hectare), Tanzania

How to identify outliers?
- Expert opinion
- Percentile of the distribution
- Median absolute deviation (MAD method)

Data cleaning

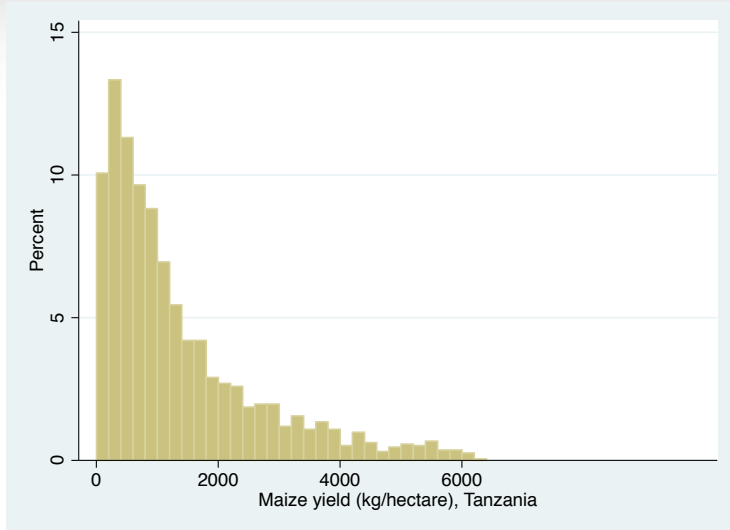95th percentile

How to deal with outliers?
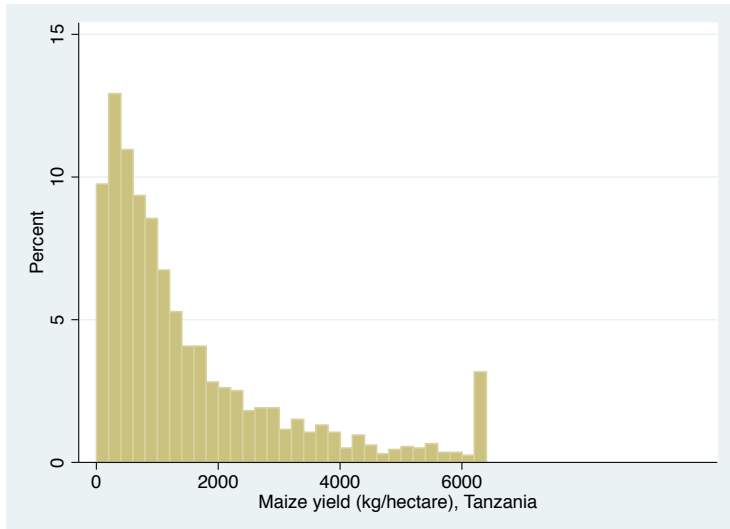- Censor / Trim / Truncate

Data cleaning

Censor

Winsorize

How to deal with outliers?
- Censor / Trim / Truncate
- Winsorize
- Replace outliers with the median

Data cleaning

RESEARCH QUESTION: How big is the gender gap in agricultural productivity in Ethiopia?

	Land productivity (USD per hectare)				
			Percentiles		
	Mean	25th	50th	75th	95th
Managed by men only	480	72	204	481	1,375
Managed by women only	336	46	174	401	980
Joint management	376	129	162	447	1,091

Mean land productivity by gender of field manager, Ethiopia LSMS-ISA
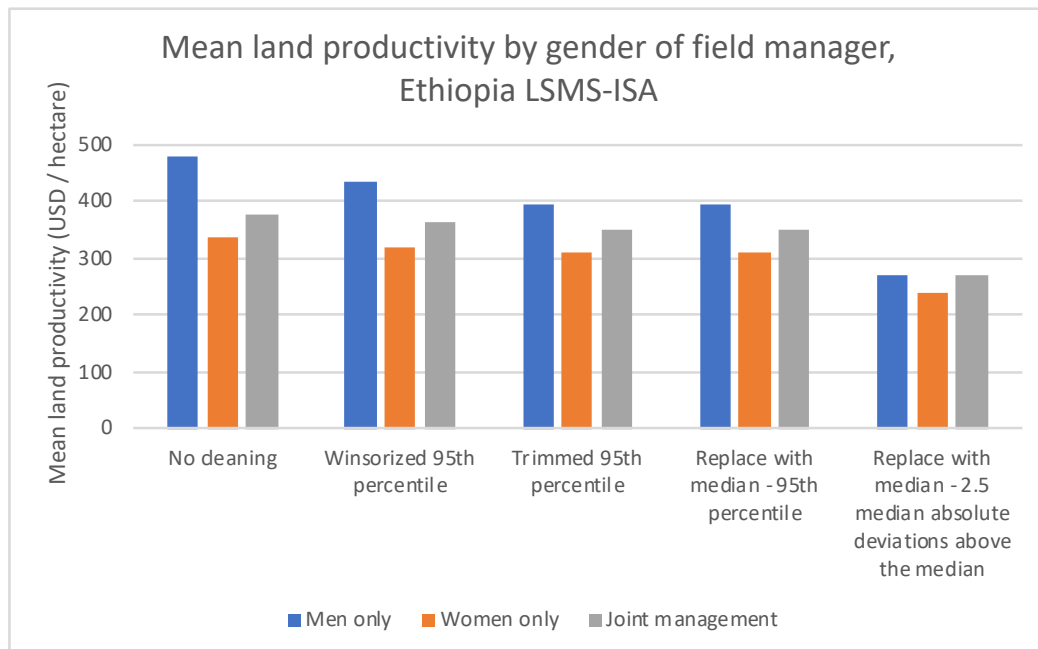
	Difference in mean land productivity between men-only and women-only
No cleaning	144
Winsorize at the 95th percentile	112
Trim at the 95th percentile	87
Replace observations above the 95th percentile with the median	87
Replace observations that are 2.5 median absolute deviations above the median with the median	27

20

Let's keep the conversation going...

info@mwapata.mw   www.mwapata.mw   ayala.wineman@gmail.com

References

Ali, D. A., K. Deininer, and A. Harris. 2017. Using national statistics to increase transparency of large land acquisition: Evidence from Ethiopia. *World Development*, 93: 62–74.

Ango, T. L. 2018. "Medium-scale" forestland grabbing in the Southwestern Highlands of Ethiopia: Impacts on local livelihoods and forest conservation. *Land*, 7 (1), DOI:10.3390/land7010024.

Anseeuw, W., Jayne, T., Kachule, R., and Kotsopoulos, J. 2016. The quiet rise of medium-scale farms in Malawi. *Land*, 5 (3): 19.

Bachewe, F. N., and B. Minten. 2020. Efficiency and profits of emerging medium-scale farms in Africa: Evidence from Ethiopia's commercial horticultural sector. Strategy Support Program Working Paper 156. International Food Policy Research Institute: Washington, D. C.

Carletto, C., S. Gourlay, and P. Winters. 2015. From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies,* 24 (5): 1-36.

Deininger, K., and F. Xia. 2016. Quantifying spillover effects from large land-based investment: The case of Mozambique. *World Development*, 87: 227–241.

Houssou, N., A. Chapoto, and C. Asante-Addo. 2016. Farm transition and indigenous growth: The rise to medium- and large-scale farming in Ghana. Discussion Paper No. 01499. International Food Policy Research Institute: Washington, D.C.

Jayne, T. S., J. Chamberlin, L. Traub, N. Sitko, M. Muyanga, F. K. Yeboah, W. Anseeuw, A. Chapoto, A. Wineman, and C. Nkonde. 2016. Africa's changing farmland ownership: The rise of medium-scale farms. *Agricultural Economics*, 47S: 197–214.

Jayne, T. S., M. Muyanga, A. Wineman, H. Ghebru, C. Stevens, M. Stickler, A. Chapoto, W. Anseeuw, D. Vanderwesthuisen, and D. Nyange. 2019. Are medium-scale farms driving agricultural transformation in sub-Saharan Africa? *Agricultural Economics*, 50 (S1): 75–95.

Jayne, T. S., M. Muyanga, J. Chamberlin, F. K. Yeboah, and A. Wineman. 2021. Changing farm size distributions and rural transformation in sub-Saharan Africa. *Mimeo*. Michigan State University: East Lansing.

Lay, J., K. Nolte, and K. Sipangule. 2018. Large-scale farms and smallholders: Evidence from Zambia. GIGA Working Papers, No. 310. German Institute of Global and Area Studies (GIGA): Hamburg.

Wineman, A., C. L. Anderson, T. Reynolds, and P. Biscaye. 2019. Methods of crop yield measurement on multi-cropped plots: Examples from Tanzania. *Food Security*, 11 (6): 1257–1273.

Wineman, A., T. S. Jayne, E. Isinika Modamba, and H. Kray. 2020. Characteristics and spillover effects of medium-scale farms in Tanzania. *European Journal of Development Research*, DOI: 10.1057/s41287-020-00323-7.

Wineman, A., T. Njagi, C. L. Anderson, T. Reynolds, D. Alia, P. Wainaina, E. Njue, P. Biscaye, and M. Ayieko. 2020. A case of mistaken identity? Measuring rates of improved seed adoption in Tanzania using DNA fingerprinting. *Journal of Agricultural Economics*, 71 (3): 719-741.