Visualizing the question and the answer

Christopher B. Yenkey

Univ of South Carolina

Sonoco International Business Dept.

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Why visualize?

You probably have one or both of these problems, both originating from trying to distill a complex phenomena into a clear, concise message:

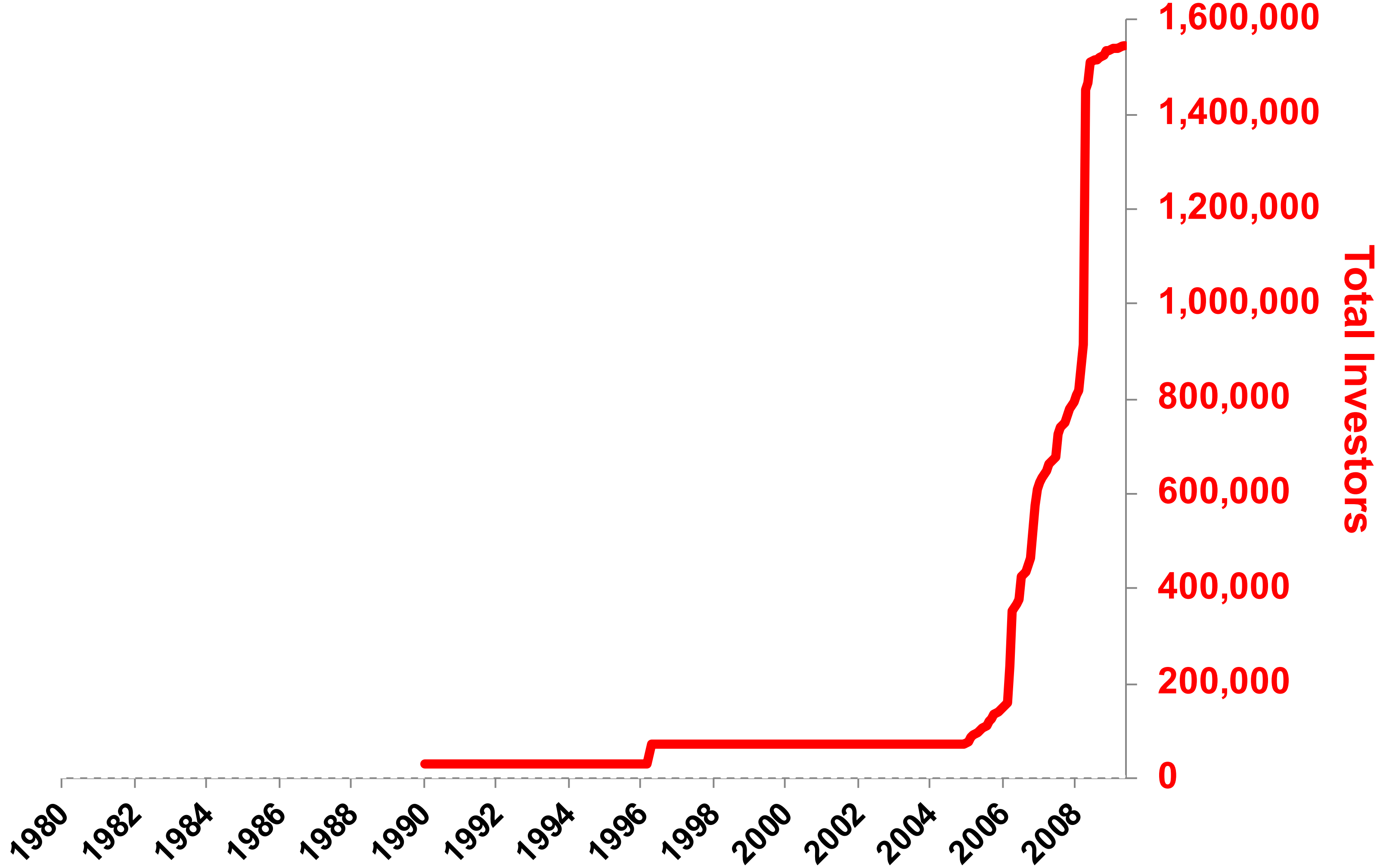
1. **I’m working in a new context.** ***How do I identify the research question and an appropriate theoretical approach?***
2. **I have great data.** ***How do I present a high volume of results clearly?***

Goal #1: set up the question

Yenkey (2015): *Mobilizing a Market*

I'm telling the story of investor recruitment into Kenya's emerging stock market. > 1 million poor, financially illiterate Kenyans decide to buy shares. WHY? Who or what convinced them that it's a good idea they should adopt, OR that it's a bad idea they should avoid?

Ok, it's an s-shaped adoption curve, so social influence matters. But….how? <u>I DON'T KNOW!</u>

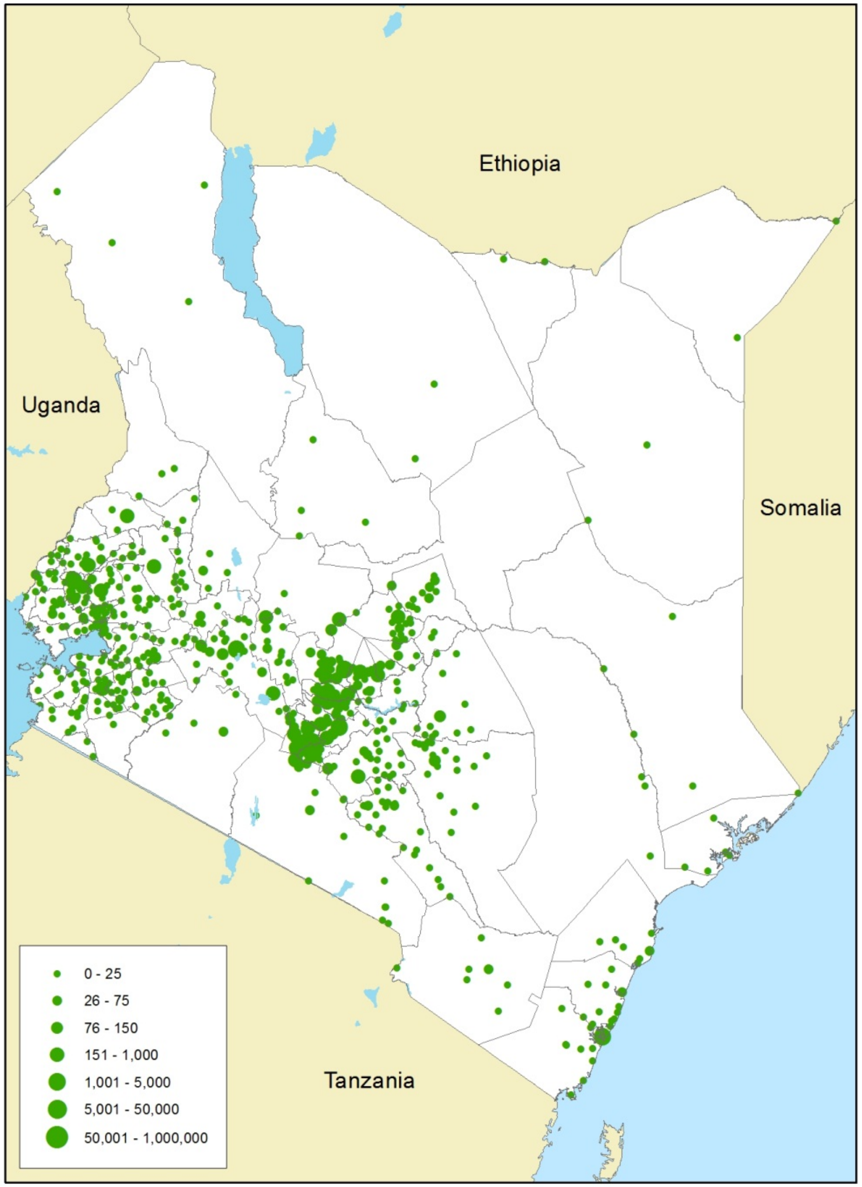


Where do the ~140,000 "early adopters" live?

If you're like me (and 100% of Kenyans and most sociologists), you expect that this is a city-based "cosmopolitan" practice that later spreads across the country.

I'll use geographic location to learn about the places where investment decisions are made.

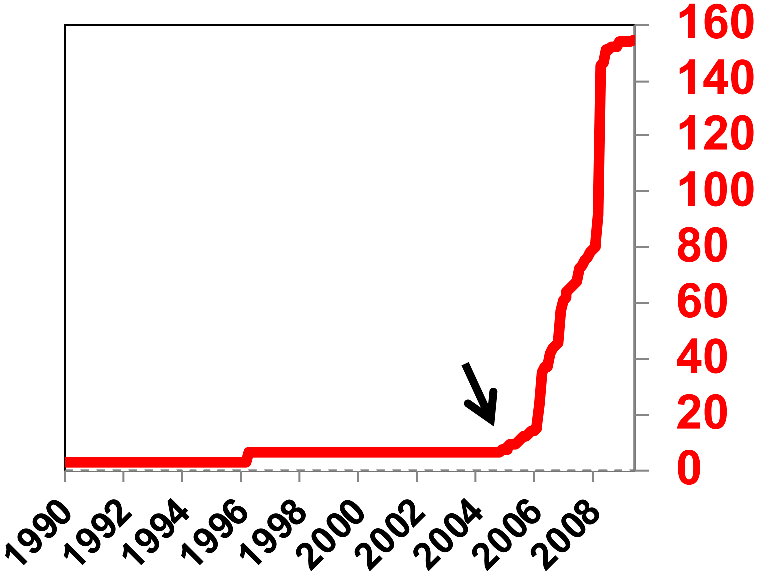
Geographic Distribution of Kenyan Shareholders: December 2005

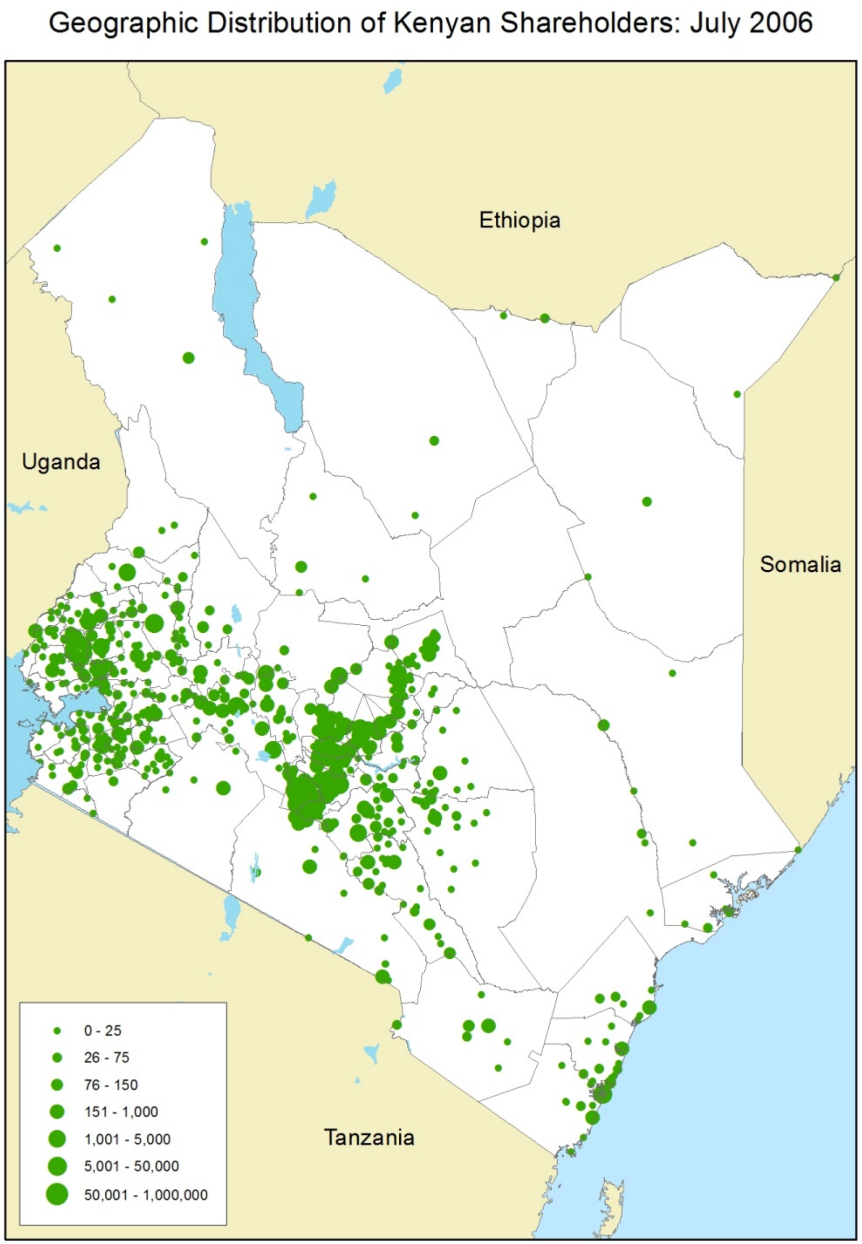


Early adopters

140,000 Investors

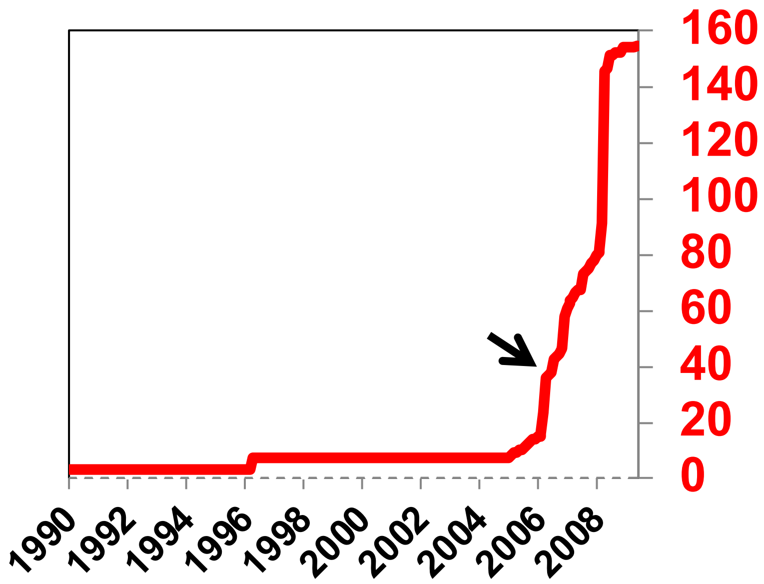
366 towns



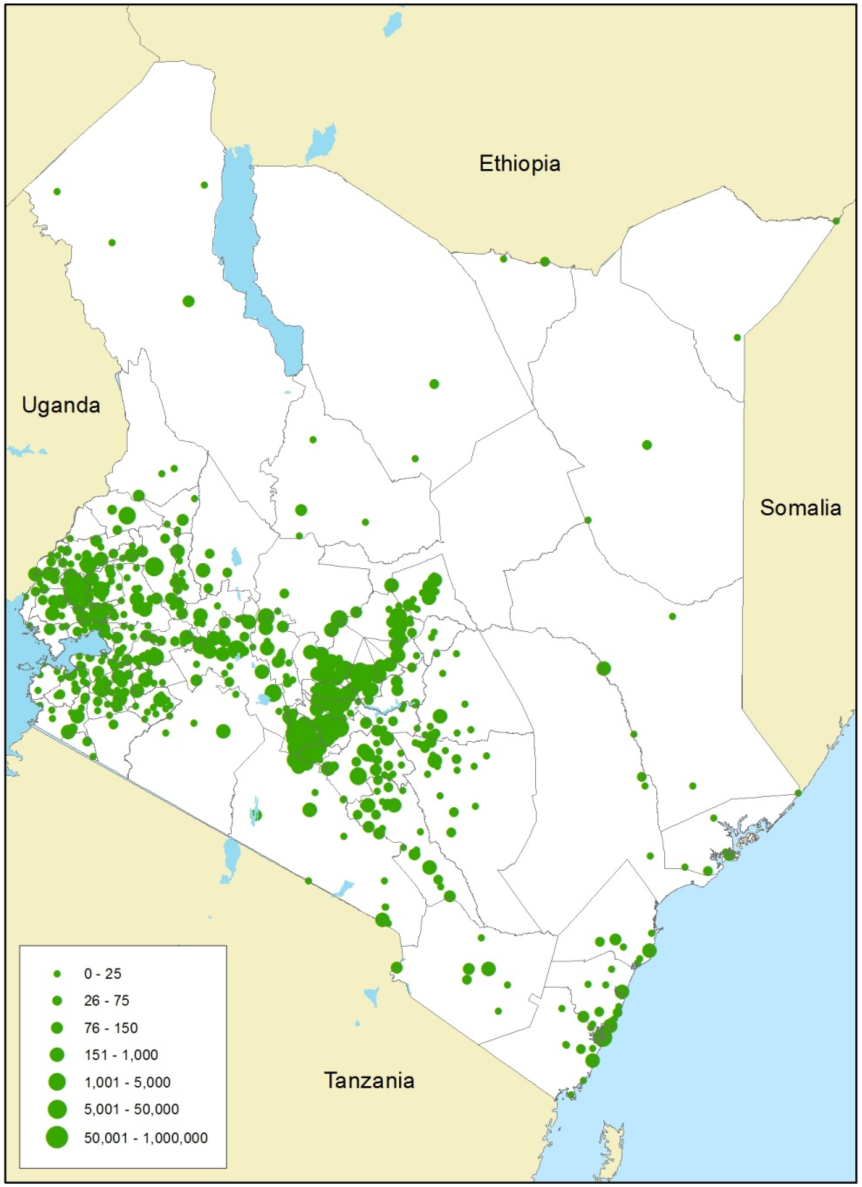


410,000 Investors

468 towns

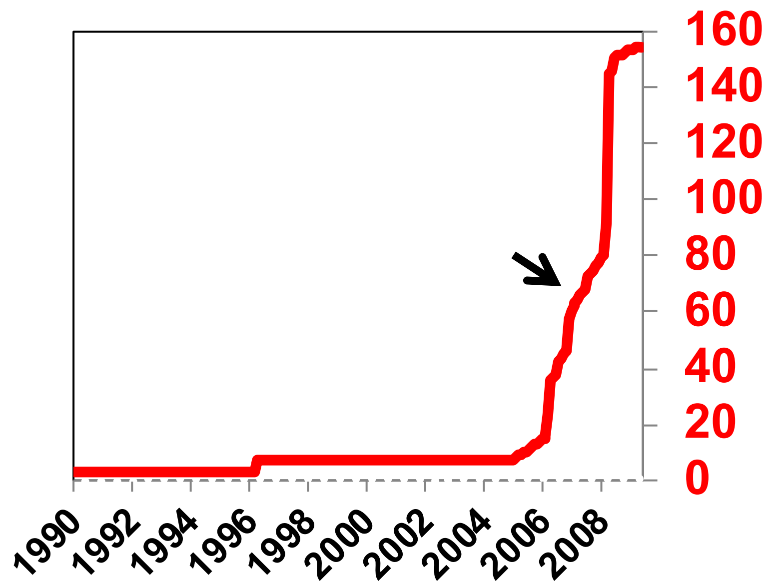


Geographic Distribution of Kenyan Shareholders: January 2007

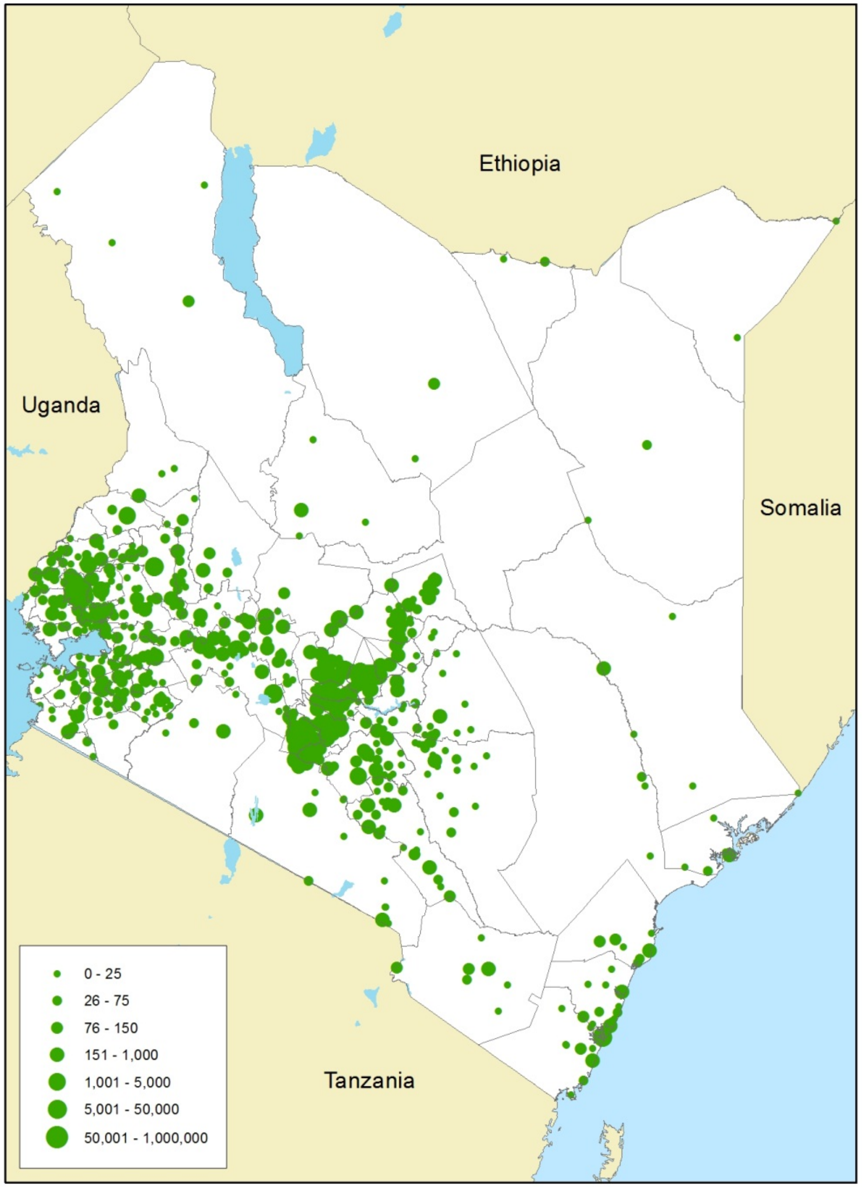


625,000 investors

486 towns



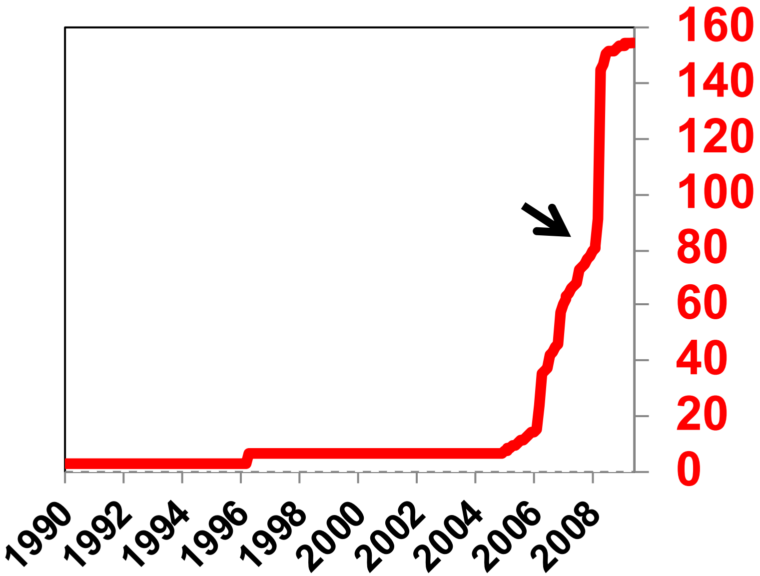
Geographic Distribution of Kenyan Shareholders: September 2007

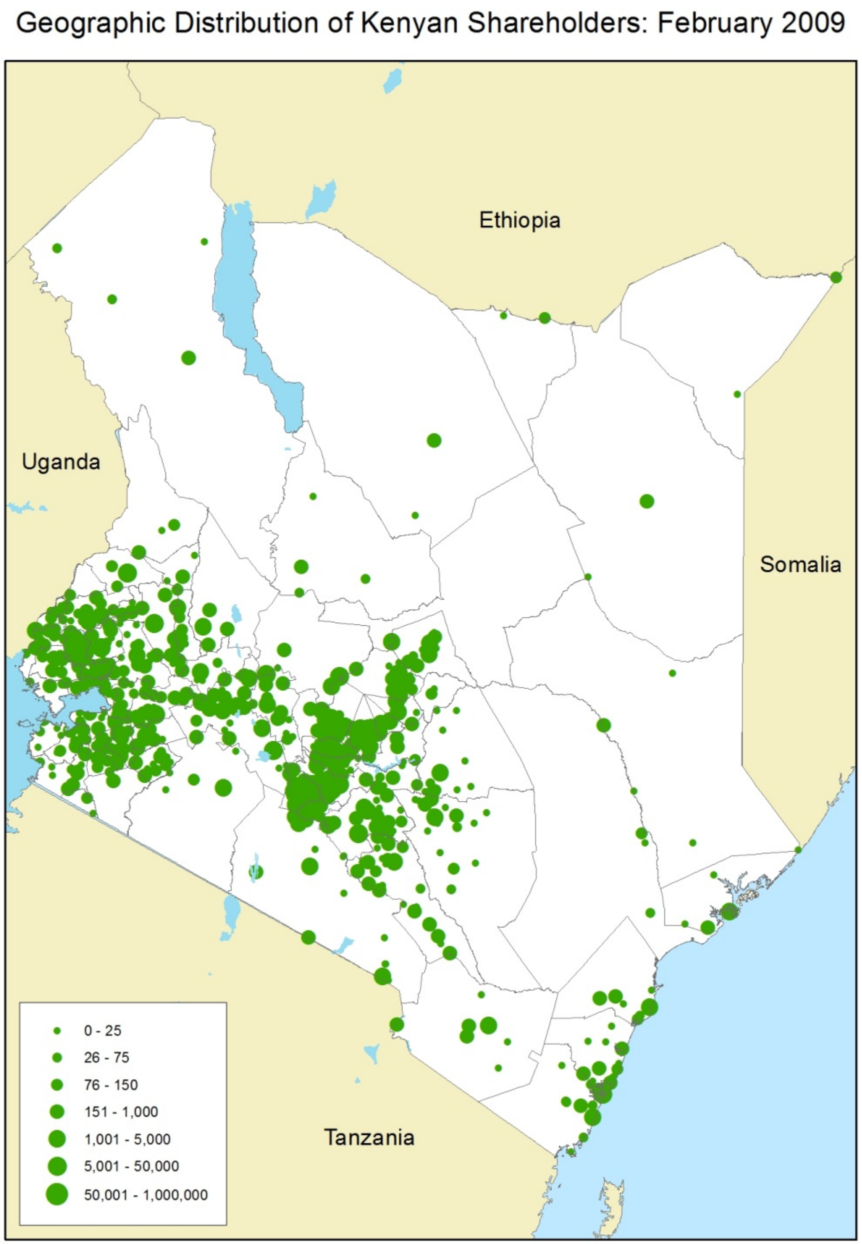


After 5th IPO

~ 700,000 investors

499 towns

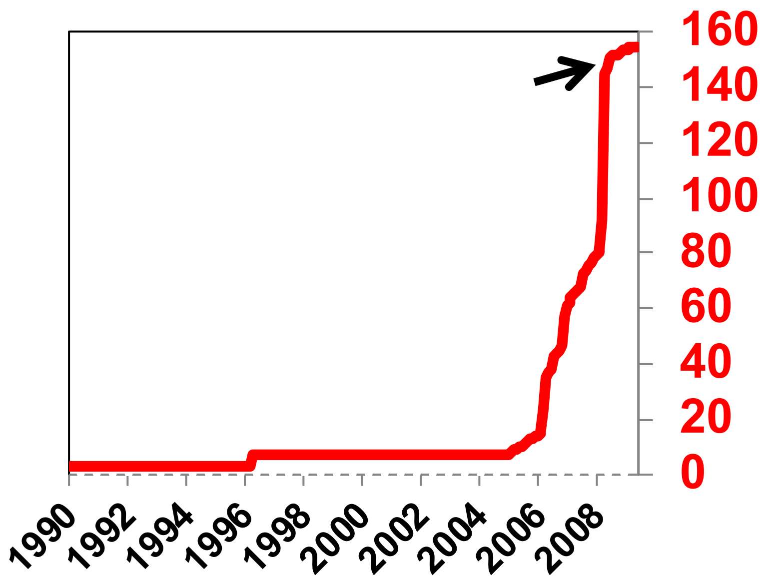




Saturation:

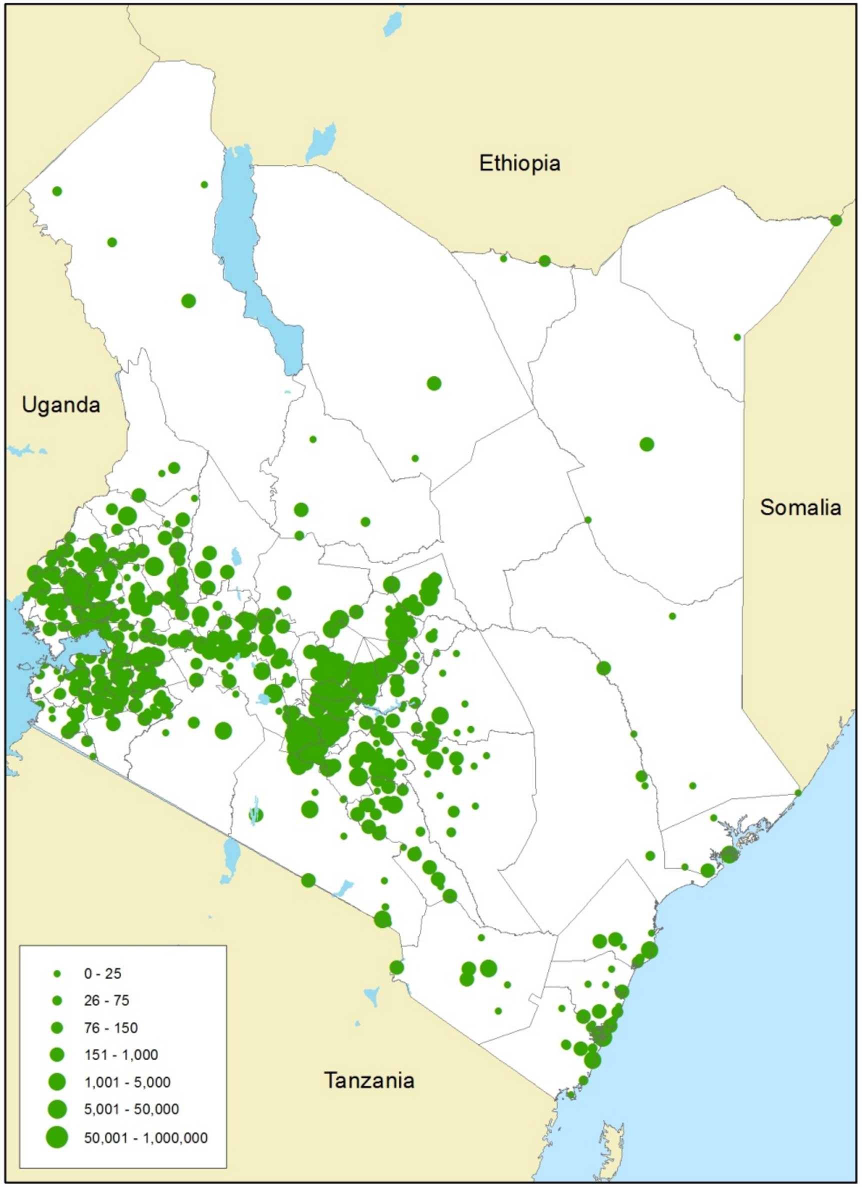
~ 1.4 M investors
(+ 900%)

563 towns
(+ 54%)

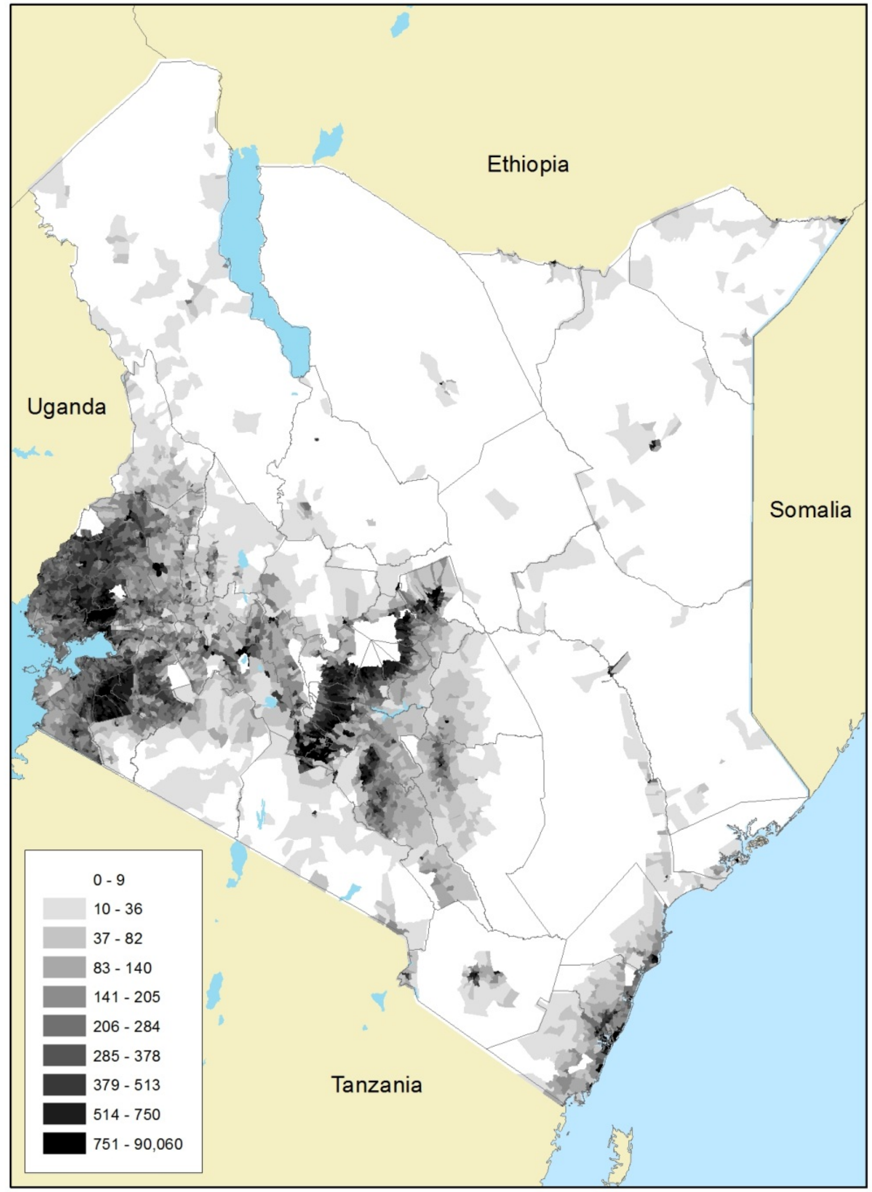


Investor pop distributed similarly to the general pop

Geographic Distribution of Kenyan Shareholders: February 2009

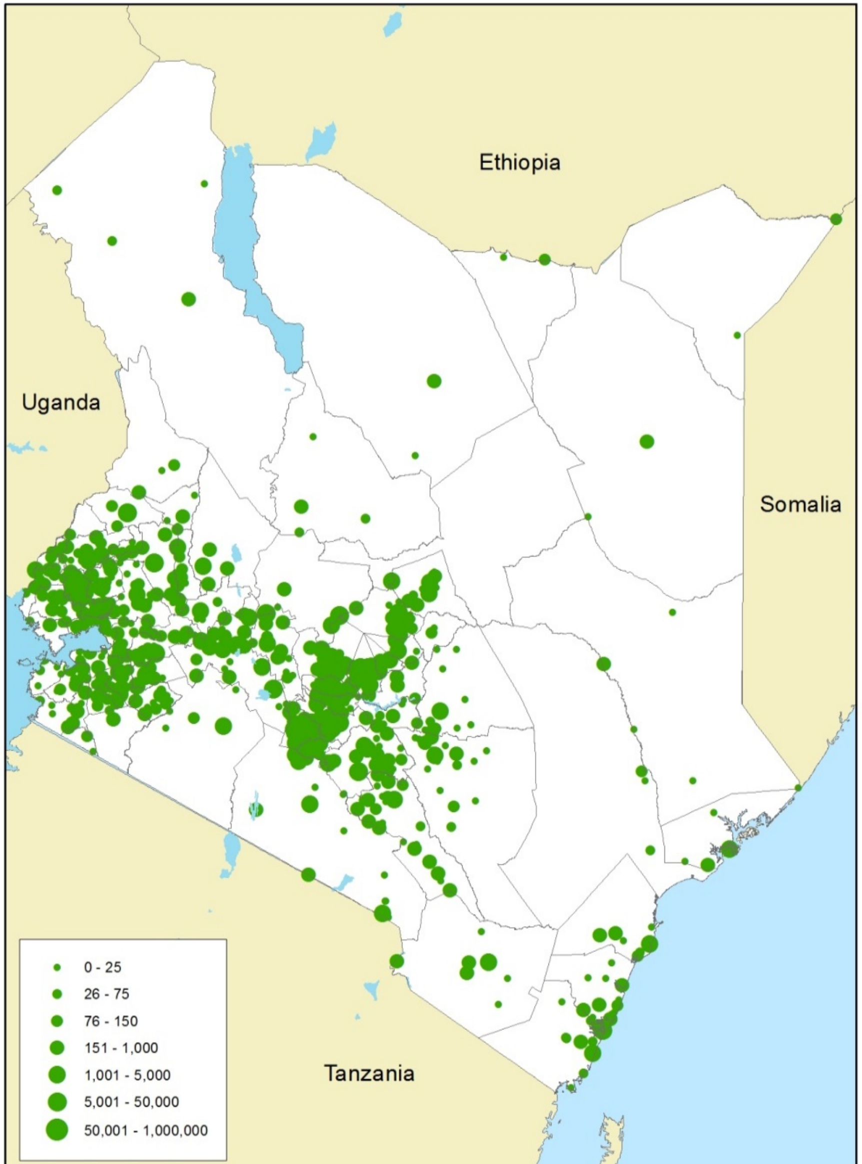


Population Density

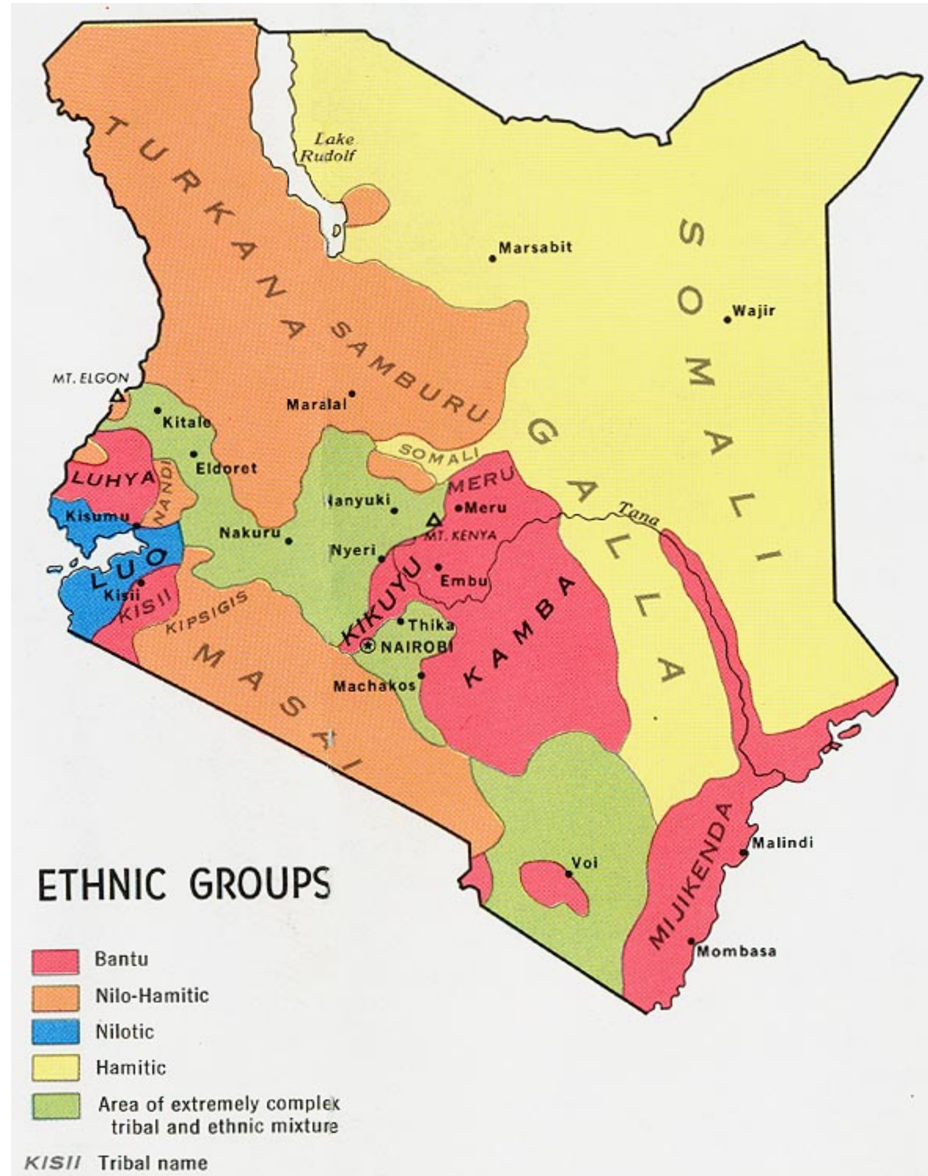


New investors learn about investing from neighbors, who are coethnics or ethnic rivals.

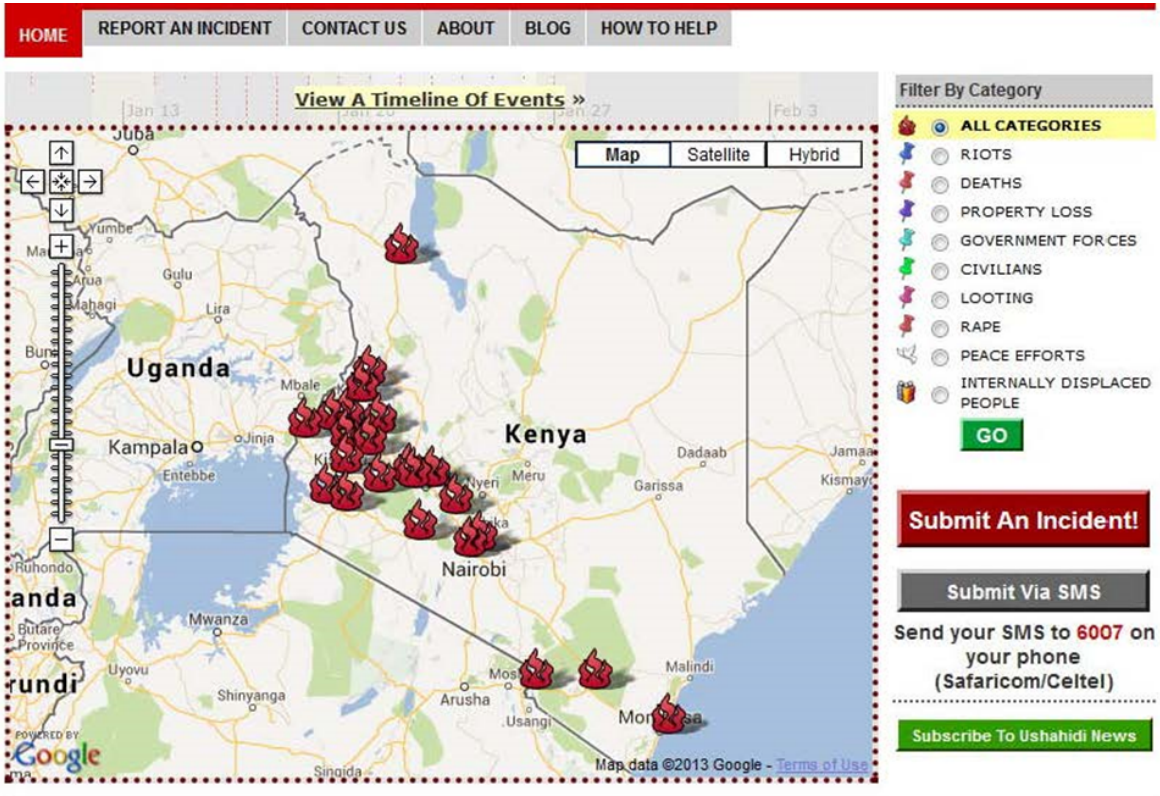
Geographic Distribution of Kenyan Shareholders: February 2009

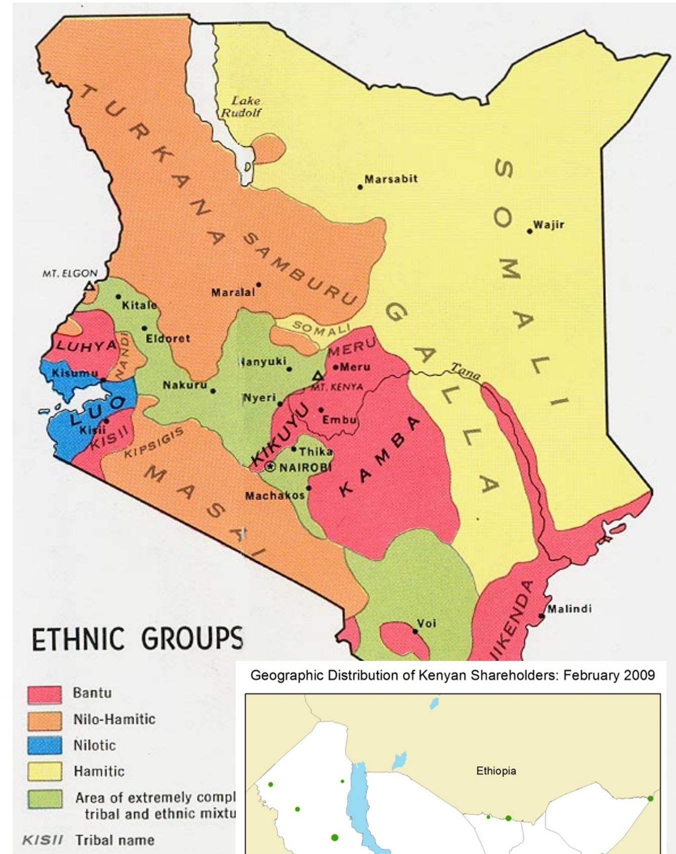


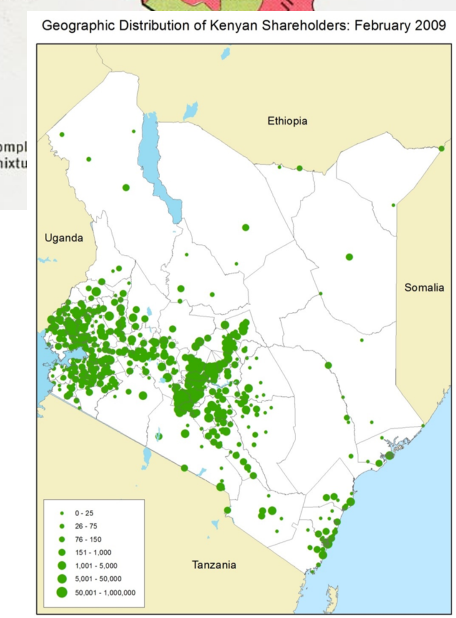
Source: U.S. CIA, "Ethnic map of Kenya" (1974)



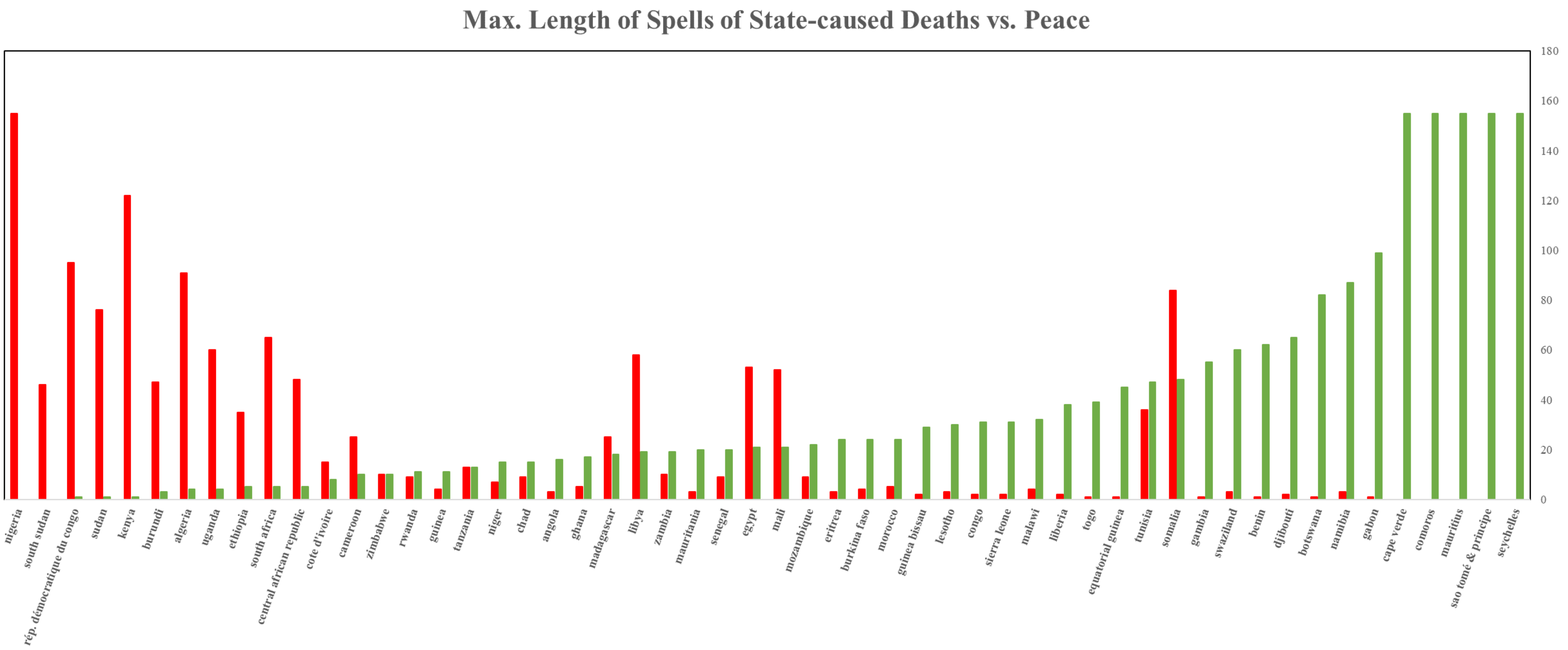
These rivalries spilled over into violent conflict at the time







Across Africa, want to know which countries experience political violence and which don't?



Takeaway #1: visualization for framing and theorizing

When working in new contexts, visualization can help get to know the place. That helps you find what's important, frame the research question, and then THEORIZE (see Swedberg 2014, *The Art of Social Theory*) something new and useful instead of rehashing old ideas in new places.

Goal #2: simplify results

Yenkey (2018, *AJS*): *The Outsider's Advantage*

Yenkey (2018, *ASQ*): *Fraud and Market Participation*

Two more papers from Kenyan stock market look at the causes and effects of misconduct (fraud) on market participation. A rogue stock broker cheats 25,000 investors. I want to know ***which*** clients are victimized, and *if* victims keep investing.

I have ***BODACIOUS*** data: excellent, admirable, attractive.

I see the same information that the corrupt brokerage sees! ***Frankly, the challenge is to present complex results to a lazy yet hard to satisfy audience.***

Compare 9 coefficients w/out the reference group, or one visual w/ confidence bars?

60 Administrative Science Quarterly 63 (2018)

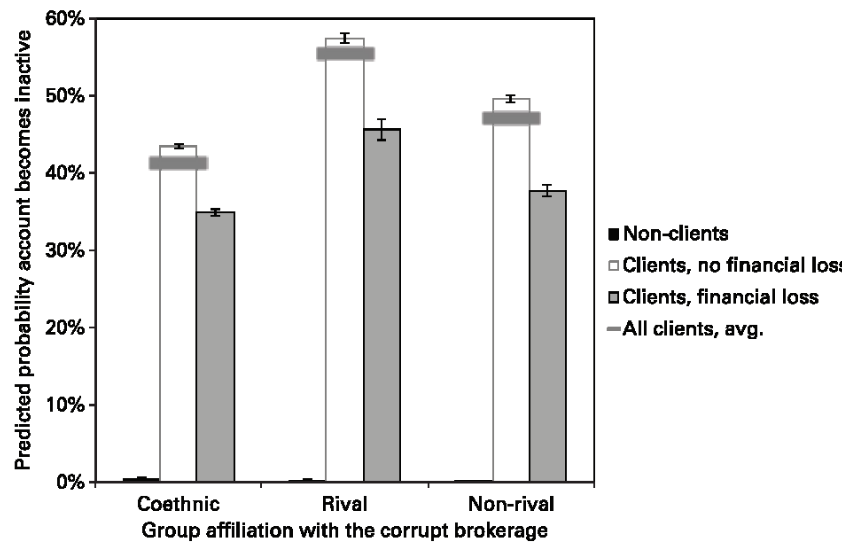
Table 2. Linear Probability Estimates of Investor's Account Becoming Inactive Post-fraud*

Variable	Coethnics Model 1	Rivals Model 2	Non-rivals Model 3
Client of the corrupt brokerage, no financial loss	.430***	.572***	.494***
	(.002)	(.003)	(.002)
Client of the corrupt brokerage, financial loss	.344***	.454***	.374***
	(.002)	(.007)	(.004)
Controls			
Client of the previous corrupt brokerage	.301***	.501***	.348***
	(.005)	(.012)	(.007)
Coethnic clients of corrupt brokerage (district, ln)	.005	.001	–.004***
	(.004)	(.002)	(.001)
All clients of corrupt brokerage (district, ln)	–.002	.001	.005***
	(.006)	(.003)	(.001)
Tenure in the market (months)	–.001***	–.000	–.001***
	(.000)	(.000)	(.000)
Value of prior investments: 2nd quintile	–.012***	–.003	–.008**
	(.002)	(.004)	(.003)
Value of prior investments: 3rd quintile	–.027***	–.012**	–.021***
	(.002)	(.004)	(.003)
Value of prior investments: 4th quintile	–.033***	–.011*	–.025***
	(.003)	(.004)	(.003)
Value of prior investments: 5th quintile	–.021***	–.010*	–.018***
	(.003)	(.005)	(.003)
Profit on prior investments (ln)	–.001***	–.000	–.001***
	(.000)	(.000)	(.000)
No. past transactions (ln)	–.006**	–.001	–.001
	(.002)	(.004)	(.003)
Female	–.000	.000	.007*
	(.003)	(.005)	(.003)
Male	.014***	.007	.017***
	(.003)	(.004)	(.003)
Constant	.087	.014	.044**
	(.044)	(.022)	(.014)
Degrees of freedom	130	127	133
No. of observations	173,537	30,977	92,606
R-squared	.357	.531	.434

* *p* < .05; ** *p* < .01; *** *p* < .001.

* Standard errors are in parentheses. All models include fixed effect dummies for investor's district of residence and each stock held.

Figure 3. Predicted probability of investor's account becoming inactive.



Source: Models 1–3, control variables estimated at mean values.

Same strategy, different DV

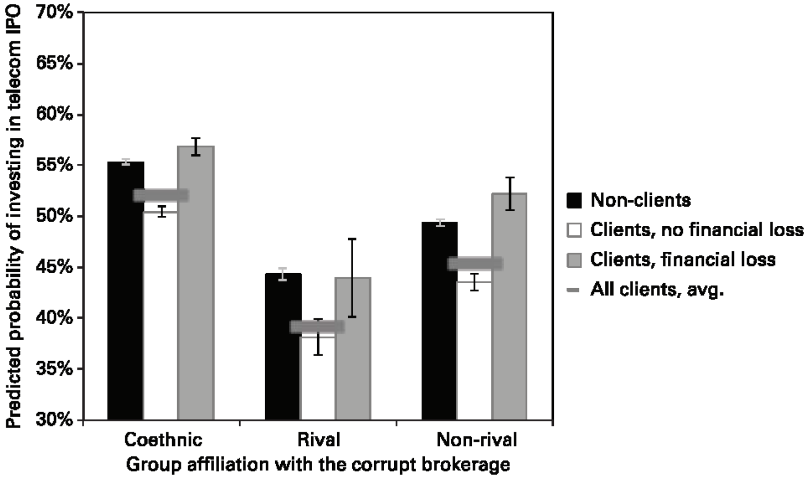
Table 3. Linear Probability Estimates of Investor's Participation in Post-fraud Telecom IPO*

Variable	Coethnics Model 4	Rivals Model 5	Non-rivals Model 6
Client of the corrupt brokerage, no financial loss	–.048***	–.062***	–.059***
	(.003)	(.009)	(.005)
Client of the corrupt brokerage, financial loss	.015***	–.004	.028
	(.005)	(.019)	(.008)
Controls			
Client of the previous corrupt brokerage	.021*	–.068*	–.021
	(.009)	(.034)	(.016)
Coethnic clients of corrupt brokerage (district, ln)	–.027***	–.003	.026***
	(.007)	(.006)	(.002)
All clients of corrupt brokerage (district, ln)	.054***	.033***	.014***
	(.012)	(.007)	(.003)
Tenure in the market (months)	.000	–.000	–.000
	(.000)	(.001)	(.000)
Value of prior investments: 2nd quintile	.020***	.024*	–.016**
	(.004)	(.010)	(.006)
Value of prior investments: 3rd quintile	.062***	.066***	.025***
	(.005)	(.011)	(.006)
Value of prior investments: 4th quintile	.095***	.100***	.056***
	(.005)	(.012)	(.007)
Value of prior investments: 5th quintile	.058***	.086***	.029***
	(.006)	(.014)	(.008)
Profit on prior investments (ln)	.002***	.002**	.003***
	(.000)	(.001)	(.000)
No. past transactions (ln)	.012*	.009	.018**
	(.004)	(.011)	(.006)
Female	.299***	.228***	.151***
	(.006)	(.013)	(.008)
Male	.252***	.188***	.109***
	(.006)	(.012)	(.007)
Constant	–.264**	–.292***	–.197***
	(.083)	(.060)	(.032)
Degrees of freedom	130	127	133
No. of observations	173,537	30,977	92,606
R-squared	.102	.097	.096

* $p < .05$; ** $p < .01$; *** $p < .001$.
* Standard errors are in parentheses. All models include fixed effect dummies for investor's district of residence and each stock held.

Figure 4. Investors' participation in post-fraud telecom IPO.

Panel A. Predicted probabilities, by ethnic group and victimization.



Source: Models 4-6, control variables estimated at mean values.

When things get really busy, give model numbers

American Journal of Sociology

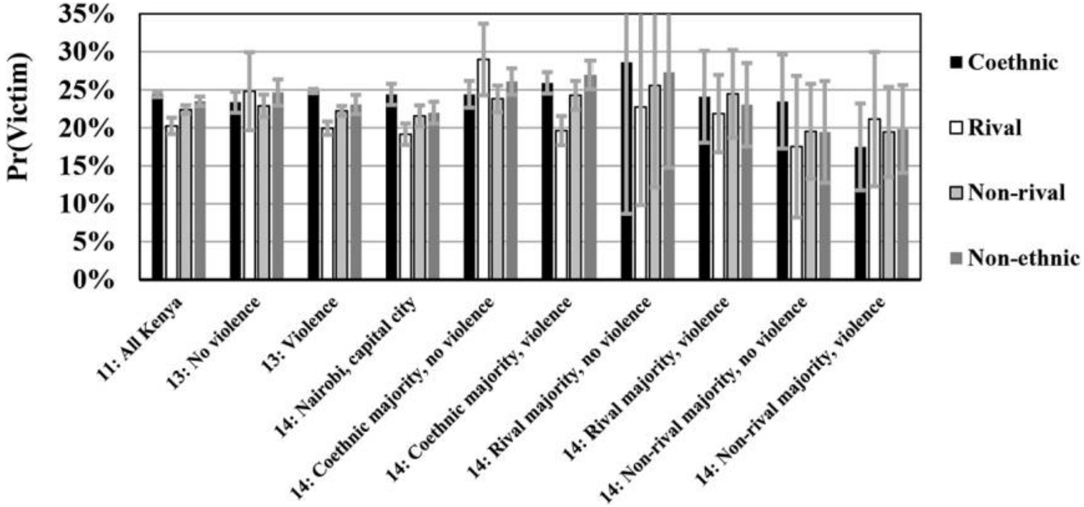


FIG. 3.—Predicted probability of fraud victimization.

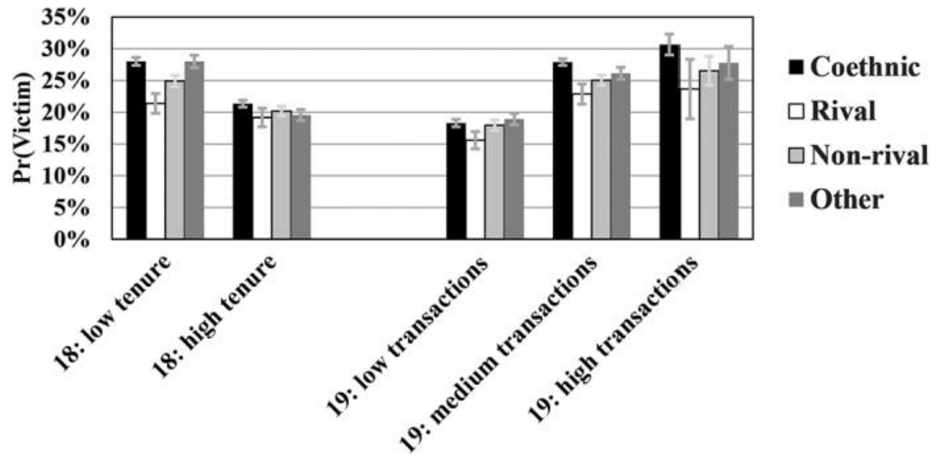


FIG. 5.—Predicted probability of fraud victimization, by ethnic group and relational embedding.

Takeaway #2: visualizing complex answers

A picture is worth 1,000 words

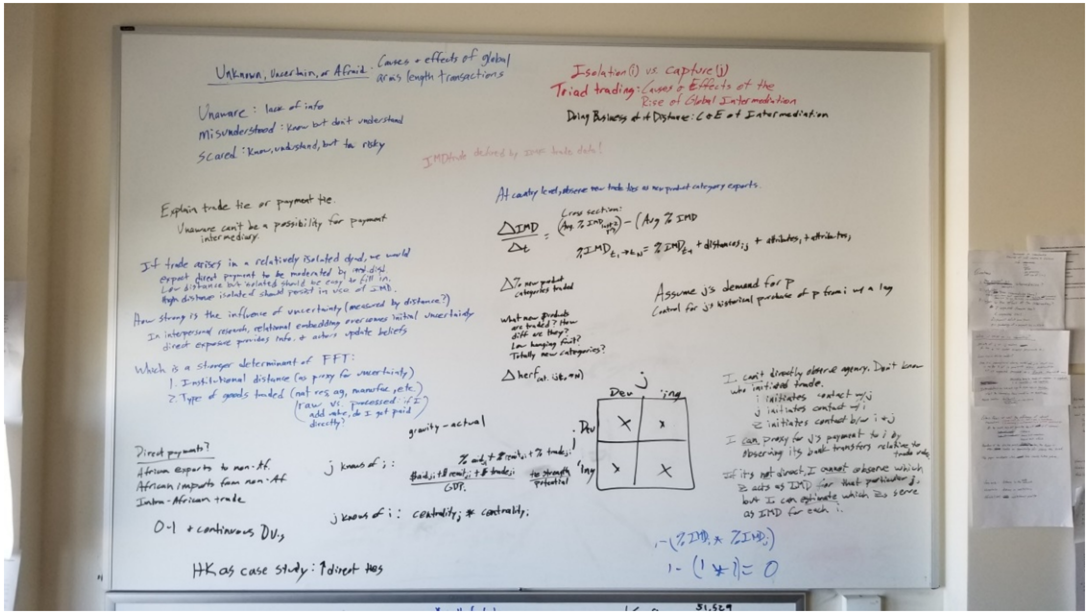
Use visuals to tell an accurate, complete, but FOCUSED story. Then give the reader the coefficients, tables, etc. necessary to go back and double check you and sift through the details. Don't start w/ a discussion of 50 things and then expect the reader to still be with you when you get to the main point. ***Provide the main point, then show them how to check on the 50 things if they want.***

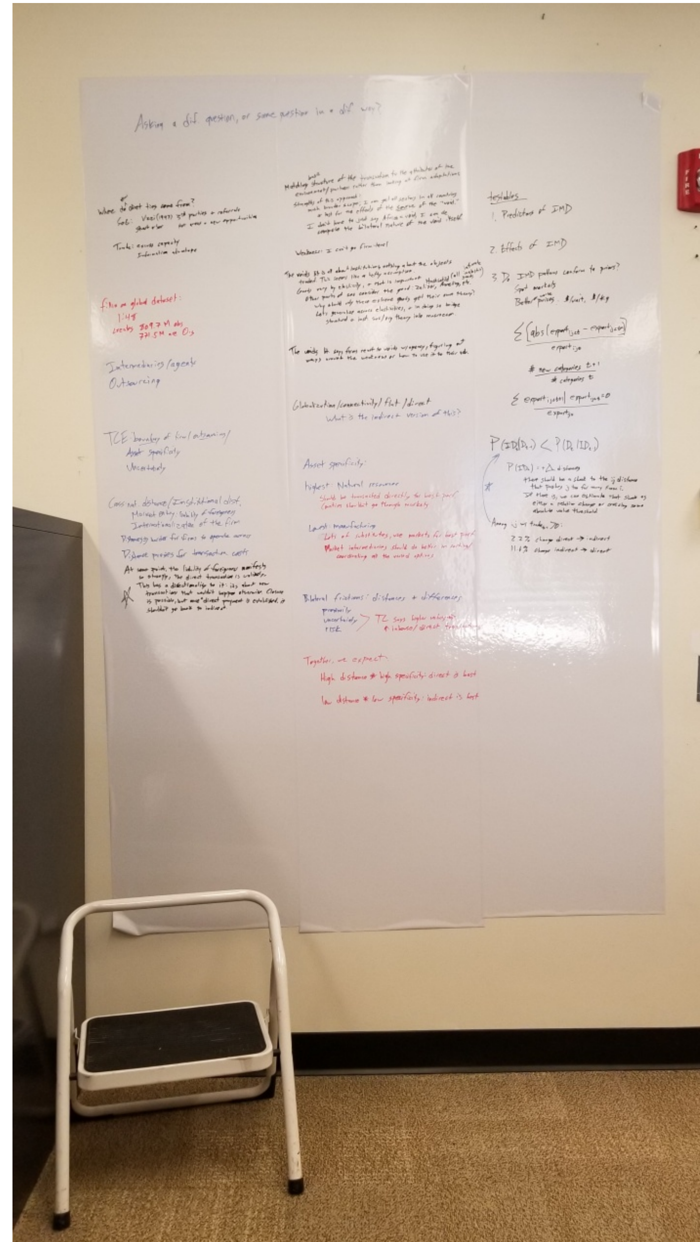
The simple command is "margins" in STATA, and you can specify specific values for specific variables; e.g.:

- predicted value of x when y = -1SD, mean, +1SD
- probability of male or female doing x when y = 1, 2, or 3
- predicted outcomes hold all other vars to mean values, so you get the value of the full model, which captures influence of all predictors

Show the magnitude of the answers to the questions you asked

Make visuals a part of your everyday





cyenkey@moore.sc.edu

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business