

CSSS 569 · Visualizing Data

GALLERY 7:
TERNARY PLOTS FOR COMPOSITIONAL DATA ANALYSIS

Christopher Adolph

Department of Political Science

and

Center for Statistics and the Social Sciences

University of Washington, Seattle



Compositional outcomes: A central banks example

In *Bankers, Bureaucrats, and Central Bank Politics*,

I argue central bankers' career backgrounds explain their monetary policy choices:

Central bankers with financial sector backgrounds choose more conservative policies, leading to lower inflation but potentially higher unemployment

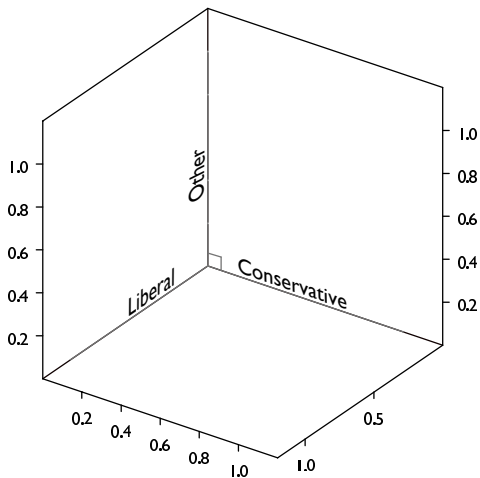
But how do central banks end up with governors whose careers are conservative?

My claim: more conservative governments should prefer to appoint more conservative central bankers, e.g., those with financial sector backgrounds

For this model, central bankers career backgrounds are composed of shares from liberal, conservative, and "other" career types

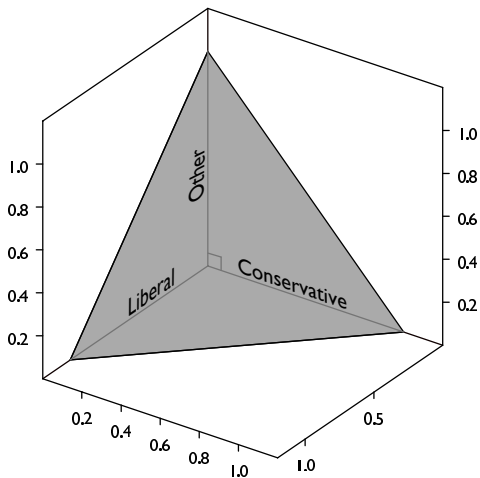
The model has a multivariate outcome:

a 3-part composition $\{ \text{Conservative}, \text{Liberal}, \text{Other} \}$ that sums to 1



I collect career compositions of central bankers at appointment
from 20 countries over 30 years

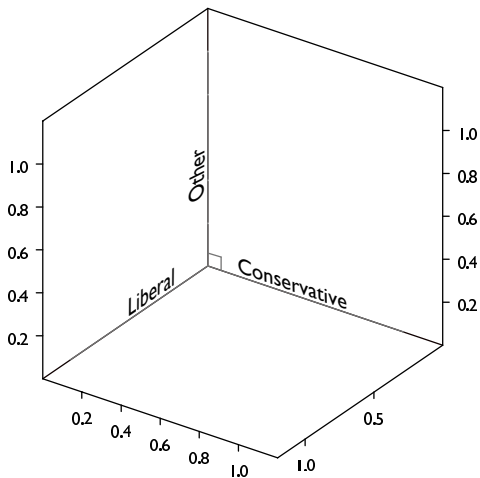
How do I visualize a three-part outcome?



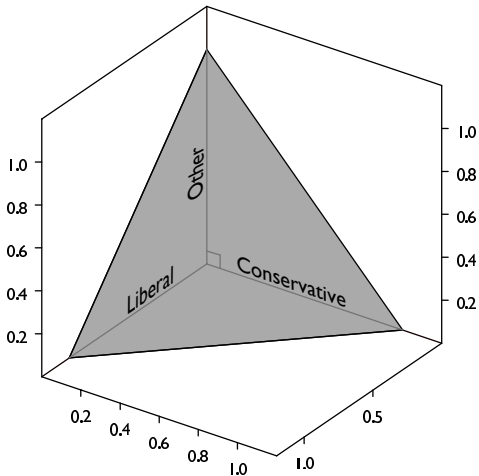
I collect career compositions of central bankers at appointment
from 20 countries over 30 years

How do I visualize a three-part outcome?

Exploit the compositional constraint!

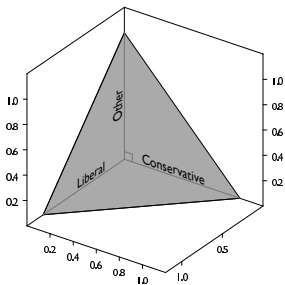


While each of the components {Liberal, Conservative, and Other} can range between 0 and 100%,



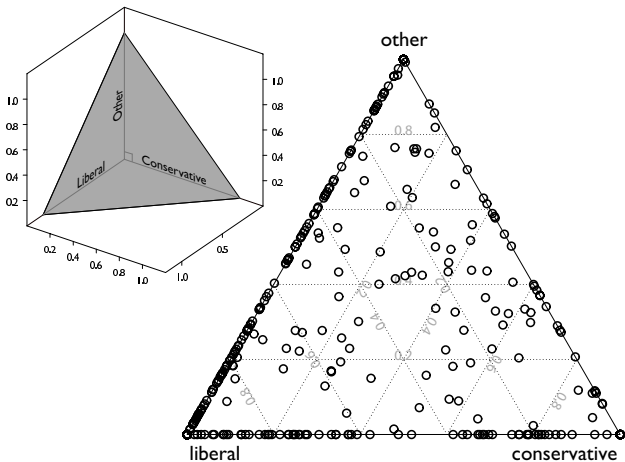
While each of the components {Liberal, Conservative, and Other} can range between 0 and 100%, their sum *must* be 100%

This constrains the possible compositions to the simplex “triangle,” which can be represented in 2D even for 3 components



How do I visualize a three-part outcome?

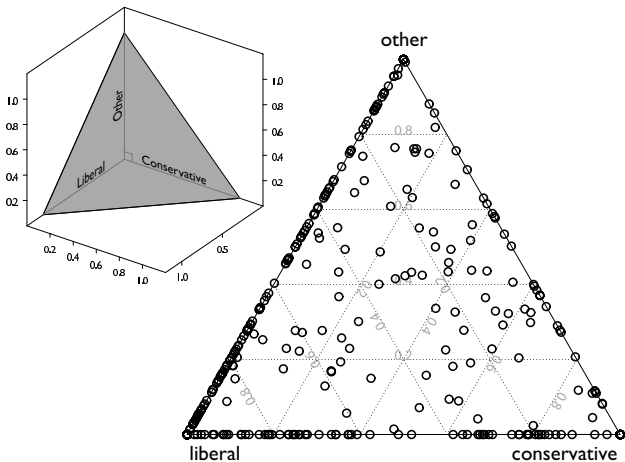
Exploit the compositional constraint!



How do I visualize a three-part outcome?

Exploit the compositional constraint!

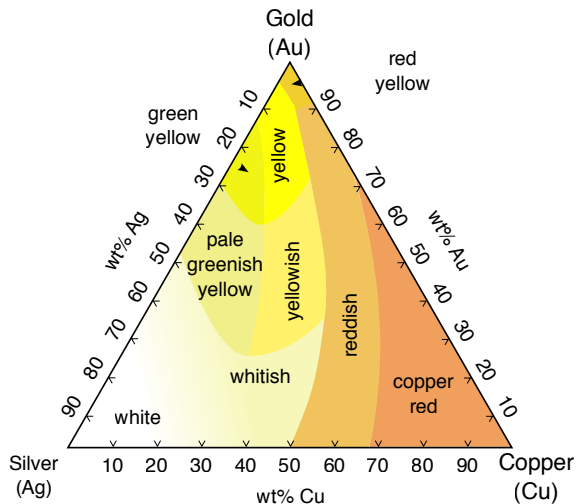
We can “pull out” the **simplex**, or the set of points meeting the compositional constraint



The simplex has 1 less dimension than the composition

The plot of the simplex for a 3-part composition is a *ternary plot*, also known as a *triangle* or *barycentric plot*

Aside on ternary plots



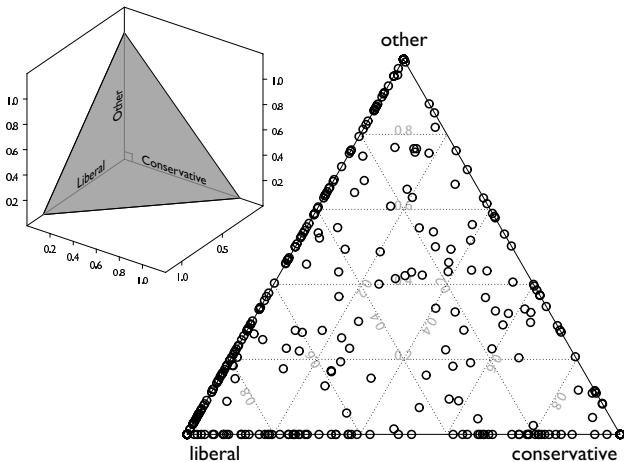
Source: Wikipedia

Ternary plots are most used in geology and metallurgy

This plot shows colors of alloys composed on Gold, Silver, and Copper

Key limitation: only works for 3-part compositions

...but you could make one category a "catch-all"



Note that many cases have one or more components at 0

This greatly complicates modeling:
 most compositional data models assume all components are non-zero

Key predictor:
Partisanship of
government (PCoG):
higher values = more
conservative

My claim:

$$\{\text{Cons, Lib, Oth}\} \\ = f(\text{PCoG, controls})$$

I estimate a *zero-inflated
compositional
regression...*

Table 8.3. Zeros-included compositional data analysis of central banker appointments.

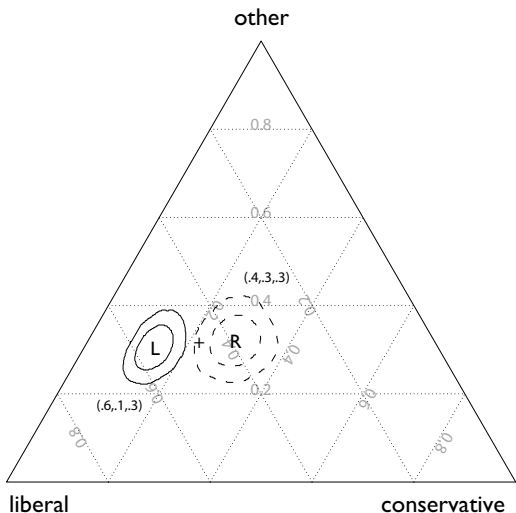
Response	Covariates	E(sign)	Model				
			1	2	3	4	
Model of non-zeros	LibExp > 0	Constant	1.142 ^{0.124}	1.142 ^{0.124}	1.308 ^d	-0.139 ^{0.389}	
		PCoG	-	-1.487 ^{0.343}	-1.487 ^{0.342}	-0.691 ^{0.486}	-1.770 ^{0.371}
		ConExp _{pre}				1.520 ^{0.678}	
		LibExp _{pre}				1.980 ^{0.596}	
	ConExp > 0	Constant	-0.242 ^{0.101}	-0.242 ^{0.101}	-0.440 ^d	-1.050 ^{0.377}	
		PCoG	+	0.171 ^{0.288}	0.170 ^{0.280}	0.648 ^{0.444}	-0.208 ^{0.322}
		ConExp _{pre}				2.350 ^{0.645}	
		LibExp _{pre}				0.573 ^{0.539}	
	OthExp > 0	Constant	0.482 ^{0.105}	0.482 ^{0.105}	0.452 ^d	1.710 ^{0.402}	
		PCoG	-	-0.662 ^{0.302}	-0.662 ^{0.302}	-0.163 ^{0.434}	-0.461 ^{0.327}
		ConExp _{pre}				-1.960 ^{0.653}	
		LibExp _{pre}				-1.500 ^{0.564}	
Model of composition	ln (LibExp / OthExp)	Constant	0.415 ^{0.128}	0.381 ^{0.124}	0.497 ^d	-0.443 ^{0.440}	
		PCoG	-	-0.390 ^{0.419}	-0.252 ^{0.419}	-0.561 ^{0.482}	-0.147 ^{0.414}
		ConExp _{pre}				0.314 ^{0.775}	
		LibExp _{pre}				1.470 ^{0.612}	
	ln (ConExp / OthExp)	Constant	-0.111 ^{0.160}	-0.112 ^{0.147}	0.085 ^d	-0.152 ^{0.498}	
		PCoG	+	0.557 ^{0.471}	0.491 ^{0.446}	0.057 ^{0.495}	0.546 ^{0.445}
		ConExp _{pre}				0.071 ^{0.818}	
		LibExp _{pre}				0.007 ^{0.722}	
Est. t dfs			7.779 ^{2.698}	4.730 ^{1.920}	6.900 ^{3.160}		
Composition Model			Normal	Student's t	Student's t	Student's t	
Notes					a,b	b	
N			411	411	411	391	
ln likelihood			-1414.82	-1066.29	-962.84	-985.80	
p-value of LR test			0.000	0.000	0.000	0.000	
against model lacking			PCoG	t-dist	f.e.	prev exp	

Key predictor:
Partisanship of
government (PCoG):
higher values = more
conservative

My claim:

$$\{Cons, Lib, Oth\} \\ = f(PCoG, controls)$$

I estimate a zero-inflated
compositional
regression...

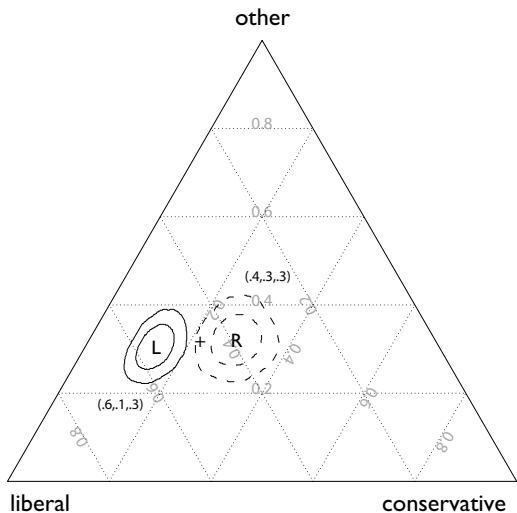


5 nonlinear coefficients aren't the quantity of interest – the expected career composition under partisan government is!

- L Left Gov't (-1.5 sd)
- R Right Gov't (+1.5 sd)
- + Average Gov't

Simulation of these components from the model & a ternary plot = Clear results

1- and 2-se confidence regions are computed by kde2d



We find the expected relationship:

Right-wing govts prefer conservative career types

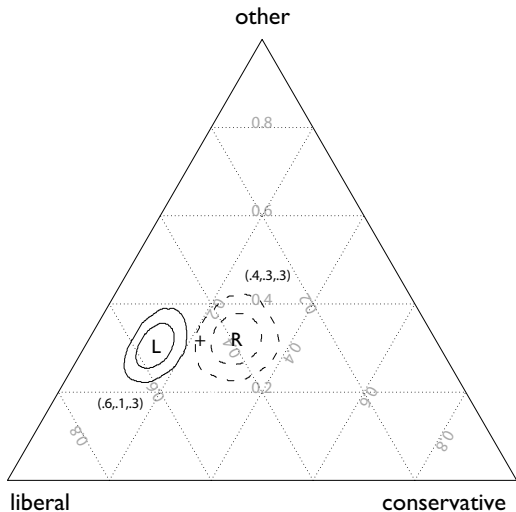
Left-wing govts prefer liberal career types

More on what this gap mean substantively later...

What else can we use ternary plots to display?

Another well-known composition:

Predicted probabilities from a three-category logit model



Example: Allocation of Authority for Health Policy

Adolph, Greer, and Fonseca consider the problem of explaining whether local, regional, or national European government have power over specific health policy areas and instruments

Areas: Pharmaceuticals, Secondary/Tertiary, Primary Care, Public Health

Instruments: Frameworks, Finance, Implementation, Provision

Each combination for each country is a case

Example: Allocation of Authority for Health Policy

Adolph, Greer, and Fonseca consider the problem of explaining whether local, regional, or national European government have power over specific health policy areas and instruments

Areas: Pharmaceuticals, Secondary/Tertiary, Primary Care, Public Health

Instruments: Frameworks, Finance, Implementation, Provision

Each combination for each country is a case

Fiscal federalism suggests lower levels for information-intensive policies and higher levels for policies with spillovers or public goods

Also control for country characteristics and country random effects

With 3 nominal outcomes for each case, need a multilevel multinomial logit

Example: Allocation of Authority for Health Policy

Covariates:

Policy area	Nominal
Policy instrument	Nominal
Regions old or new	Binary
Country size	Continuous
Number of regions	Continuous
Mountains	Continuous
Ethnic heterogeneity	Continuous

Tricky part to the model:

some cases have structural zeros for regions (when they don't exist!)

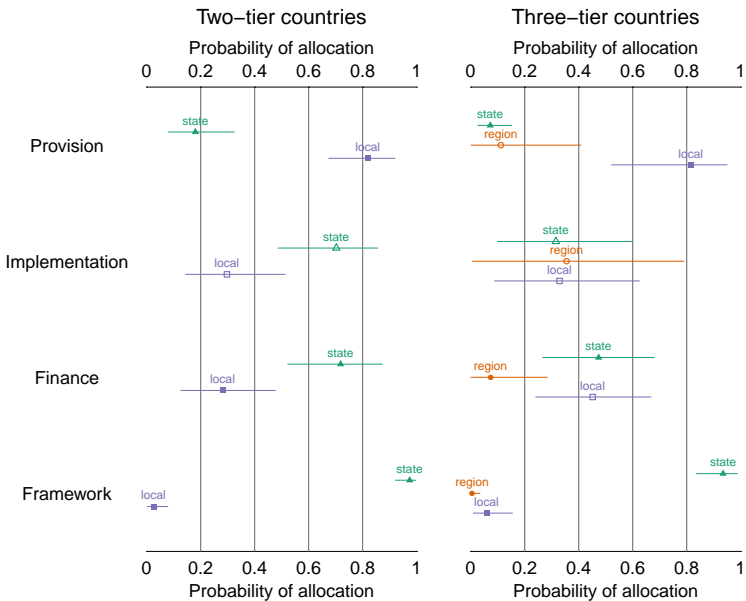
How to set up counterfactuals?

We could set all but one covariate to the mean, then predict the probability of each level of authority given varied levels of the remaining covariate

We should do this separately for countries with and without regional governments

Let's fix everything but policy instrument to the mean values, then simulate the probability of authority at each level for each instrument

We show the results using a "nested" dot plot, made using `ropeladder()` in the `tile` package



Compositional data simplexes *redux*

Probabilities have a special property: they sum to one

Variables that sum to a constraint are compositional

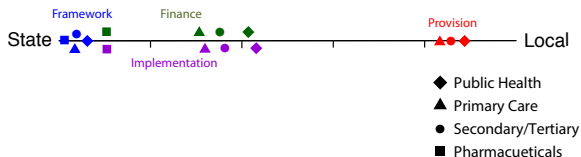
We can plot a two-part composition on a line,
and a three-part composition on a triangular plot

This makes it easier to show more complex counterfactuals, such as every combination of policy area and instrument

But we also need to work harder to explain these plots

Probability of allocation of authority by policy type

Let's start with countries that have no regions, just local and state levels:

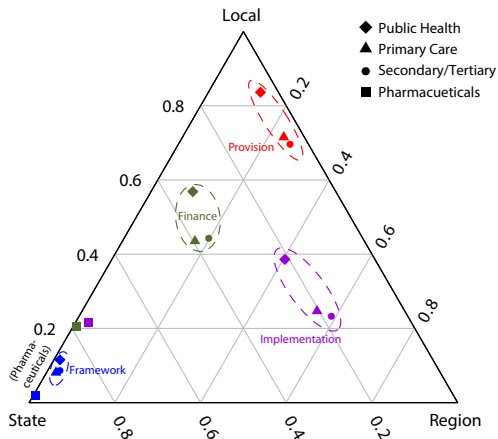


Above holds country characteristics at their means

And predicts the probability of state or local control

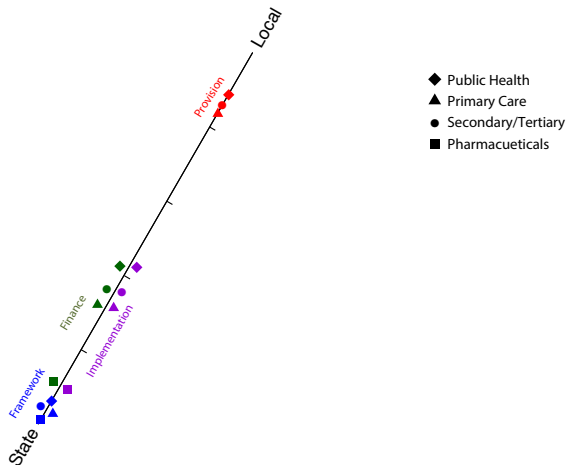
Because of the compositional constraint, these always sum to 1:
if the probability of local is p , the probability of state is $1 - p$

Probability of allocation of authority by policy type



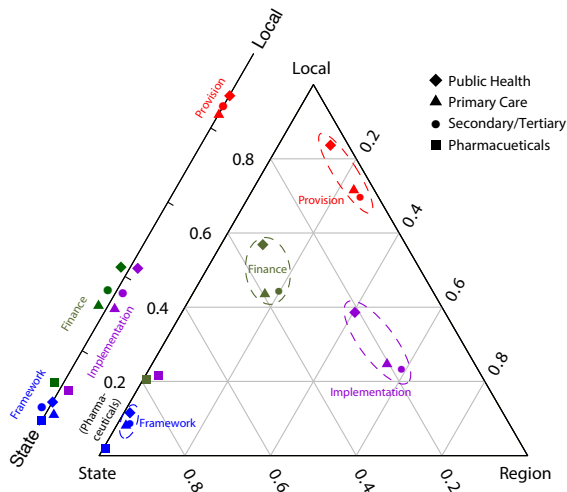
Next consider countries with all three levels, holding covariates at means

Probability of allocation of authority by policy type



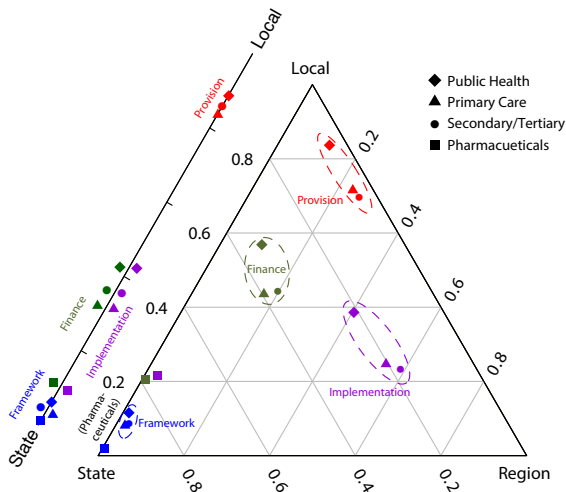
If we are clever...

Probability of allocation of authority by policy type



We can display everything in a single 2D figure

Probability of allocation of authority by policy type

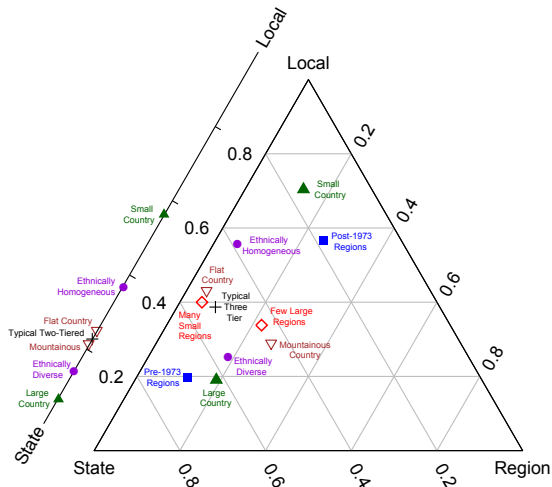


Note the 2-level "line" looks like a projection of the 3-level triangle down to 1D

Probability of allocation of authority by country type

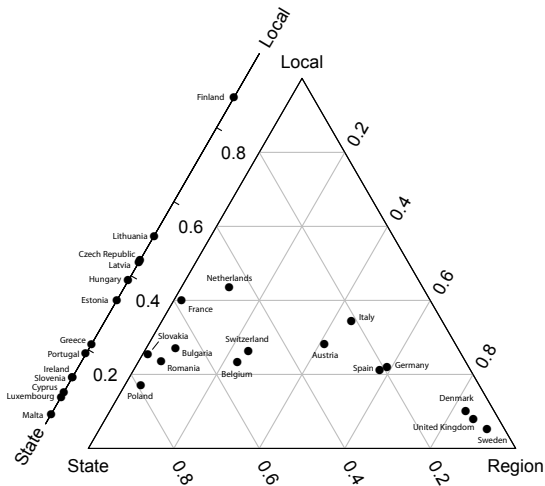
We can use the same framework to illustrate other implications of the model

Probability of allocation of authority by country type

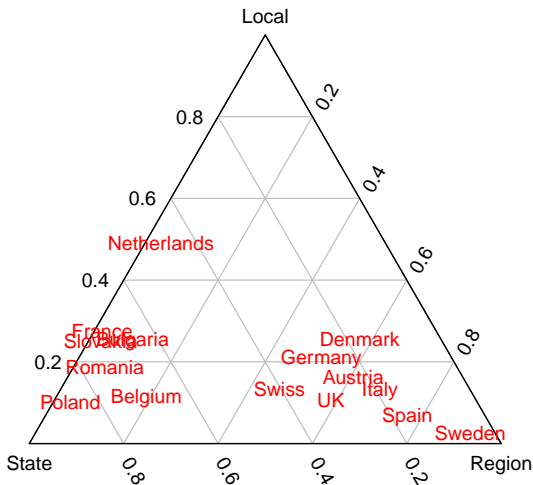


Above holds policy area and instrument "at their means"

Residual country effects



Looking at the country random effects might suggest omitted variables

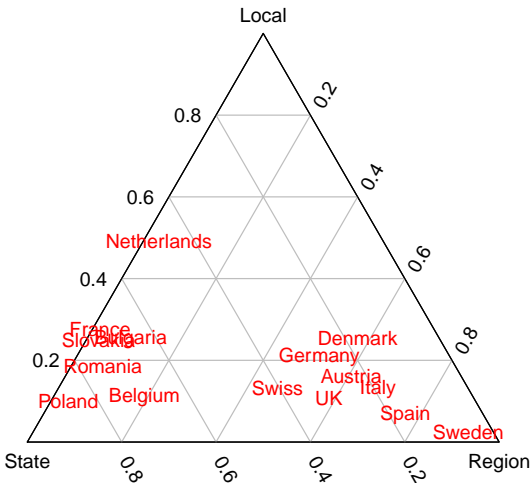


This is the random effects plot from an earlier iteration of the model

(Notice I often don't beautify plots until they are "final")

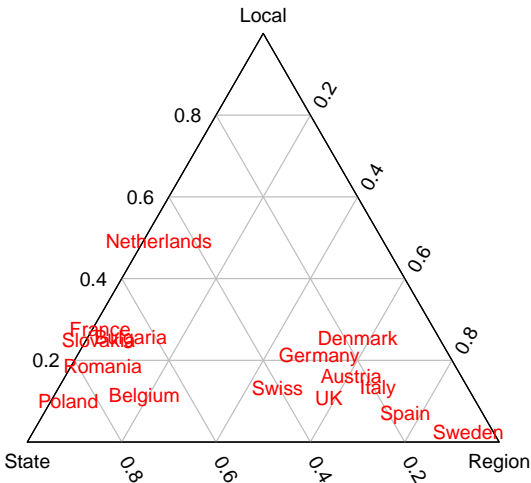
Some residuals appear "clustered" towards regional devolution

What missing variable does this plot suggest?



What missing variable does this plot suggest?

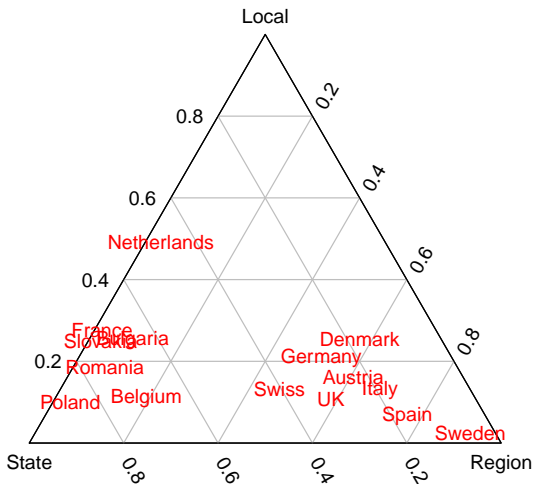
Do the clustered countries have something in common?



What missing variable does this plot suggest?

Do the clustered countries have something in common?

I couldn't see it



What missing variable does this plot suggest?

Do the clustered countries have something in common?

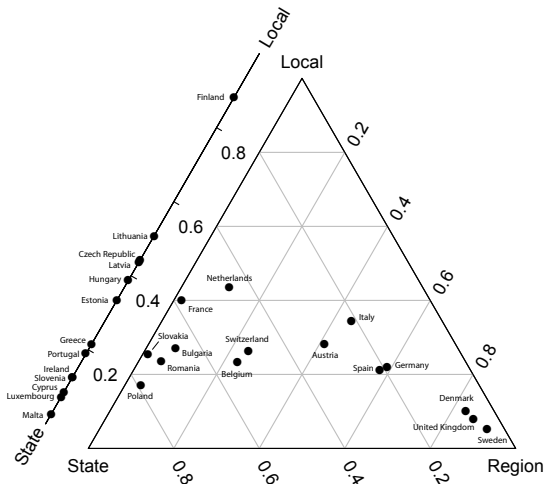
I couldn't see it

My wife did: good skiing

And historically mountains strengthened regional autonomy!

Lesson: Share your diagnostic plots!

Residual country effects



After controlling for mountainousness, no clusters of similar countries remain