Example: Brazil Model

	Means	Log Yields	Production / 1000
GDDs / 1000	2.946	0.018^{*}	0.716
	(0.931)	(0.008)	(0.450)
KDDs / 1000	0.149	-0.140^{***}	-2.828
	(0.146)	(0.042)	(1.821)
Frosts	0.944	0.005^{***}	0.036^{**}
	(3.499)	(0.001)	(0.014)
Precip. (m)	1.421	-0.045^{*}	0.882
	(0.719)	(0.021)	(0.646)
Precip.^{2}	2.538	0.015^{***}	-0.265
	(2.439)	(0.004)	(0.163)
State cubic trends		Yes	Yes
R^2		0.138	0.950
Adj. R^2		0.136	0.950
Num. obs.		42582	42600

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Motivation

Preparing the global analysis.

- Want to give each variety in each country its own parameters: γ_{iv}
- Want to partially pool across countries for a given variety: $\gamma_{iv} \sim \mathcal{N}(\gamma_v, \tau_{\gamma_v})$
- Want to partially pool variety hyperparameters $\gamma_v \sim \mathcal{N}(\gamma, \tau_\gamma)$
- Want to do this for every parameter $\gamma, \kappa, \phi, \pi, \psi$
- Full-Bayes is too slow to simultaneously perform the regression

Terminology Note: Traditionally, "Hierarchical linear modeling" and "multilevel modeling" describe model substitutions where data only exists at lowest level.

The Model

Pooled:

$$\log y_{it} = \alpha_i + beta_v + \gamma g_{it} + \kappa k_{it} + \phi f_{it} + \pi p_{it} + \psi p_{it}^2 + \epsilon_{it}$$

Unpooled:

$$\log y_{ivt} = \alpha_i + \beta_v + \gamma_{iv} g_{it} + \kappa_{iv} k_{it} + \phi_{iv} f_{it} + \pi_{iv} p_{it} + \psi_{iv} p_{it}^2 + \epsilon_{ivt}$$

Partial-pooling relationships:

$$\begin{aligned}
\gamma_{iv} &= \gamma_v + \epsilon_{iv} \\
\gamma_a &= \gamma_c + \epsilon_a \\
\gamma_r &= \gamma_c + \epsilon_r \\
&\vdots
\end{aligned}$$

The Model

$$
\begin{array}{rlllr}
\log y_{ivt} & = \alpha_i & + \gamma_{iv} g_{it} & & + \cdots \\
\hline
\log y_{ivt} & = \sum_j \alpha_j \mathbf{1}_{j=i} & + \sum_{ju} \gamma_{ju} g_{it} \mathbf{1}_{ju=iv} & + \gamma_a 0 + \gamma_r 0 + \gamma_c 0 & + \cdots \\
\hline
0 & = \sum_j \alpha_j 0 & + \sum_{ju} \gamma_{ju} \mathbf{1}_{ju=1a} & - \gamma_a 1 - \gamma_r 0 - \gamma_c 0 & + \cdots \\
0 & = \sum_j \alpha_j 0 & + \sum_{ju} \gamma_{ju} \mathbf{1}_{ju=2a} & - \gamma_a 1 - \gamma_r 0 - \gamma_c 0 & + \cdots \\
 & & \vdots & & \\
0 & = \sum_j \alpha_j 0 & + \sum_{ju} \gamma_{ju} \mathbf{1}_{ju=1r} & - \gamma_a 0 - \gamma_r 1 - \gamma_c 0 & + \cdots \\
 & & \vdots & & \\
0 & = \sum_j \alpha_j 0 & + \sum_j u \gamma_{ju} \mathbf{1}_{ju=1c} & - \gamma_a 0 - \gamma_r 0 - \gamma_c 1 & + \cdots \\
 & & \vdots & & \\
0 & = \sum_j \alpha_j 0 & + \sum_j u \gamma_{ju} 0 & + \gamma_a 1 + \gamma_r 0 - \gamma_c 1 & + \cdots \\
0 & = \sum_j \alpha_j 0 & + \sum_j u \gamma_{ju} 0 & + \gamma_a 0 + \gamma_r 1 - \gamma_c 1 & + \cdots
\end{array}
$$

Time for an R Package!

Many use cases:

Partial pooling	`y ~ f	f - .`
Partial pooling to multiple	`y ~ f	f > g`
Pairwise coefficient relation	`y ~ a + b	a - b`
	`y ~ f	fa == fb`
Multilevel data	`y ~ f	y ~ g	f - g`
Smoothness criteria	???

```
model <- hierlm(logyield ~ 0 + region + variety + regionvariety:gdd1000 |
                regionvariety:gdd1000 > variety:gdd1000 |
                varietyarabica:gdd1000 == varietycombined:gdd1000 |
                varietyrobusta:gdd1000 == varietycombined:gdd1000,
                data, ratios=c(.1, .1, .1))
summary(model)
```

Pooling Effects