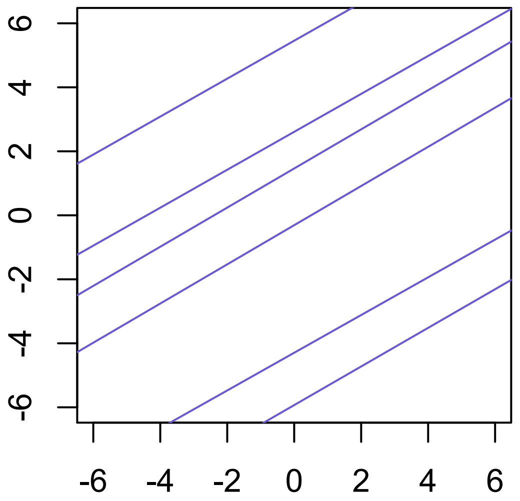
Statistical Rethinking

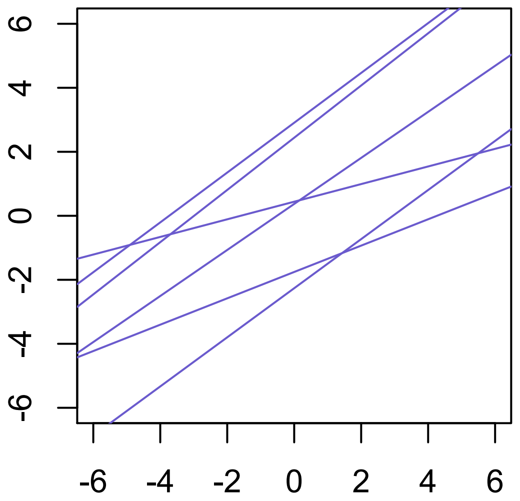
Week 9: Multilevel Models II
Adventures in Covariance

Richard McElreath

Kinds of varying effects

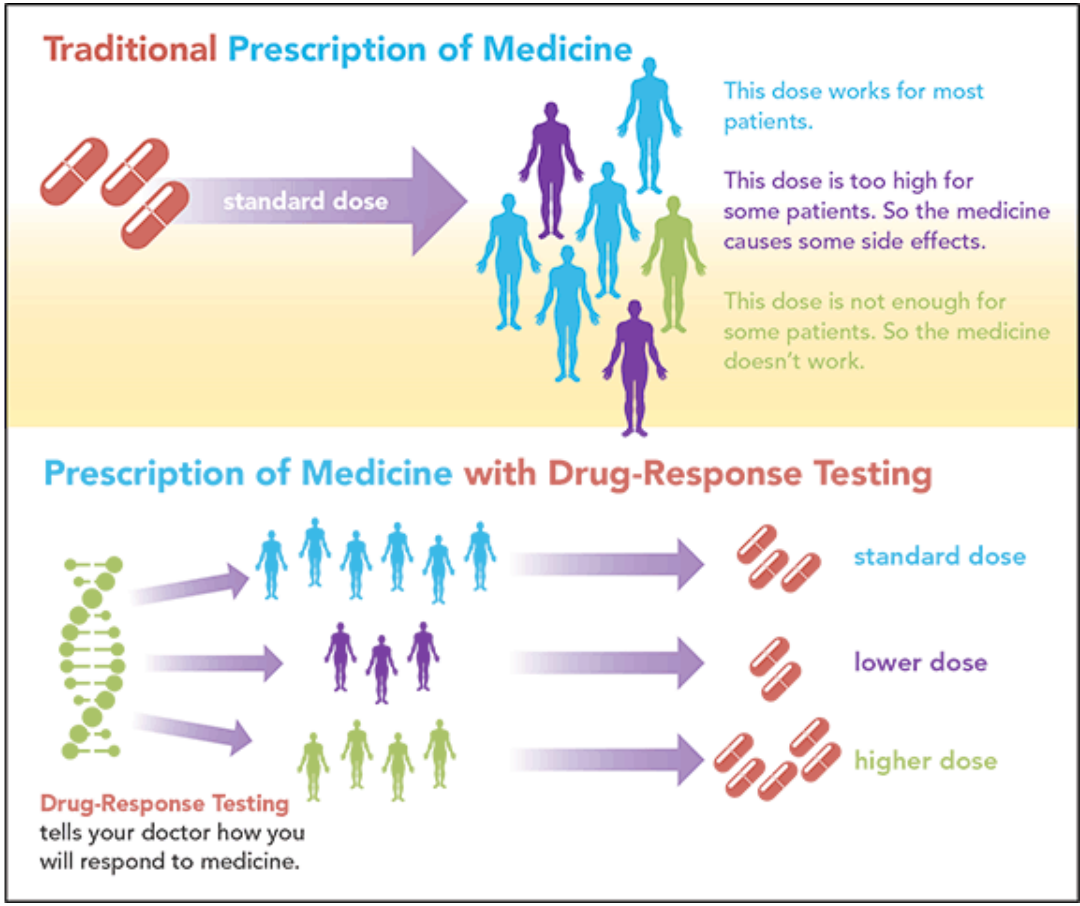
- *Varying intercepts*: means differ by cluster
- *Varying slopes*: effects of predictors vary by cluster
- Any parameter can be made into a varying effect
  - (1) split into vector of parameters by cluster
  - (2) define population distribution





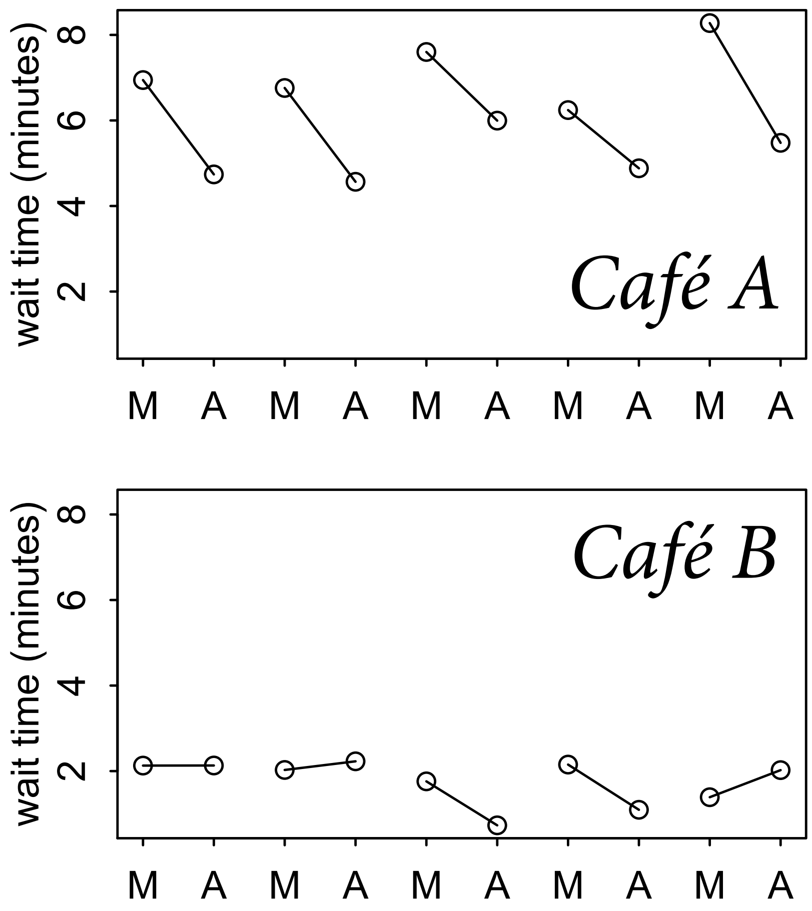
Varying slopes

- Why varying slopes?
  - drugs affect people differently
  - after school programs don't work for everyone
  - not every unit has same relationship to predictor
  - variation is important, whether for intervention or inference
- *Average* effect misleading?
- Pooling, shrinkage, mnesia

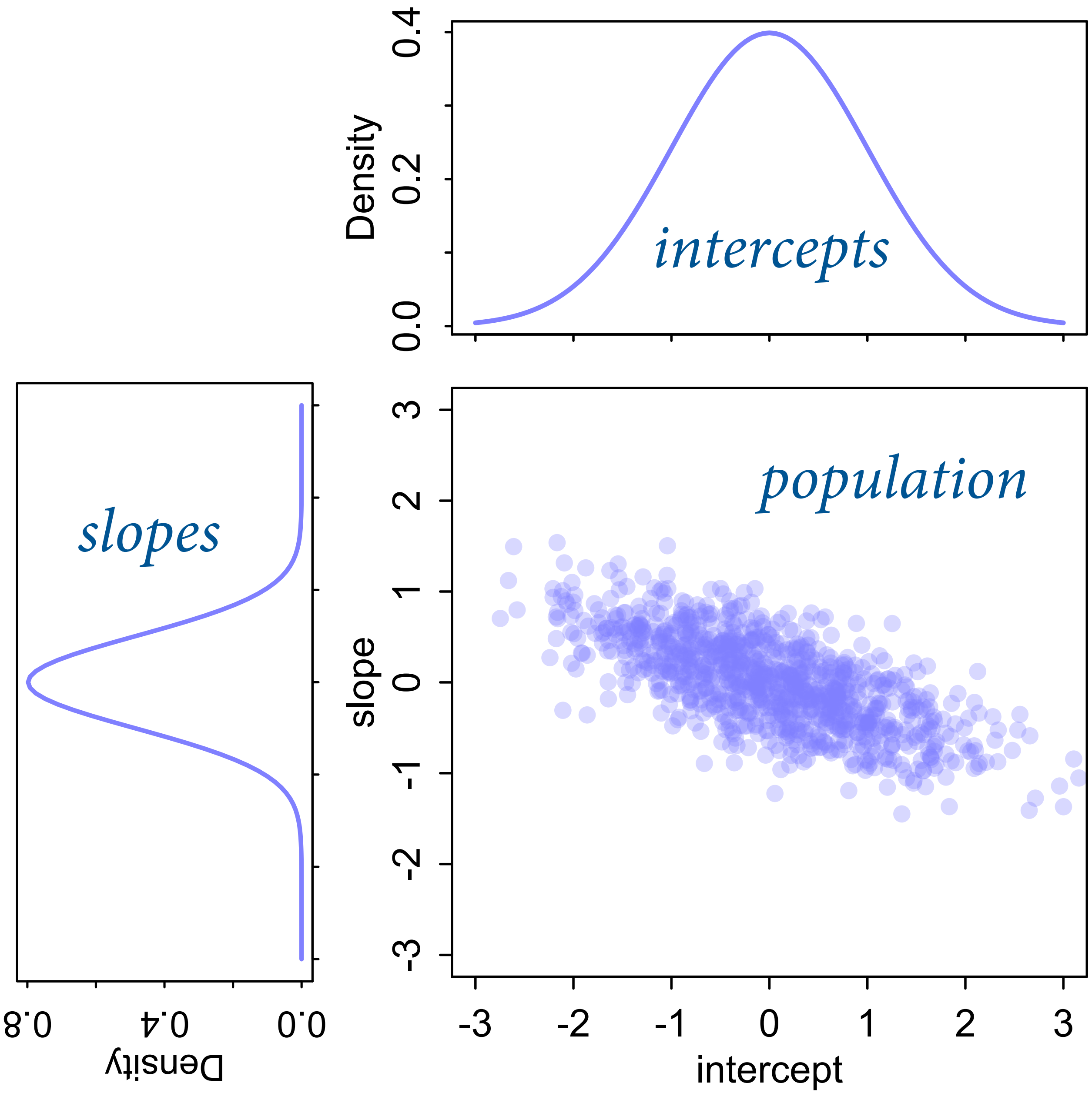


Café Robot

- Robot programmed to visit cafés, order coffee, record wait time
- Visits in *morning* and *afternoon*
- Intercepts: avg morning wait
- Slopes: avg difference btw afternoon and morning
- Are intercepts and slopes related?
    - Yes => pooling across parameter types!



Population of Cafés



Population of Cafés

- 2-dimensional Gaussian distribution
  - vector of means
  - variance-covariance matrix

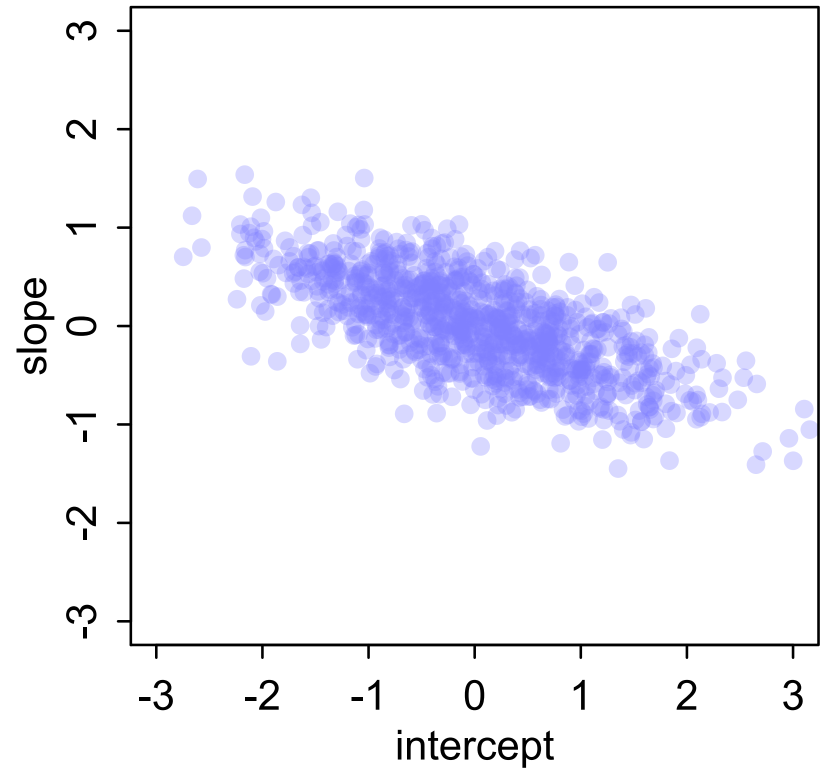
intercepts variance

$$\begin{pmatrix} \sigma_\alpha^2 & \sigma_\alpha \sigma_\beta \rho \\ \boxed{\sigma_\alpha \sigma_\beta \rho} & \sigma_\beta^2 \end{pmatrix}$$

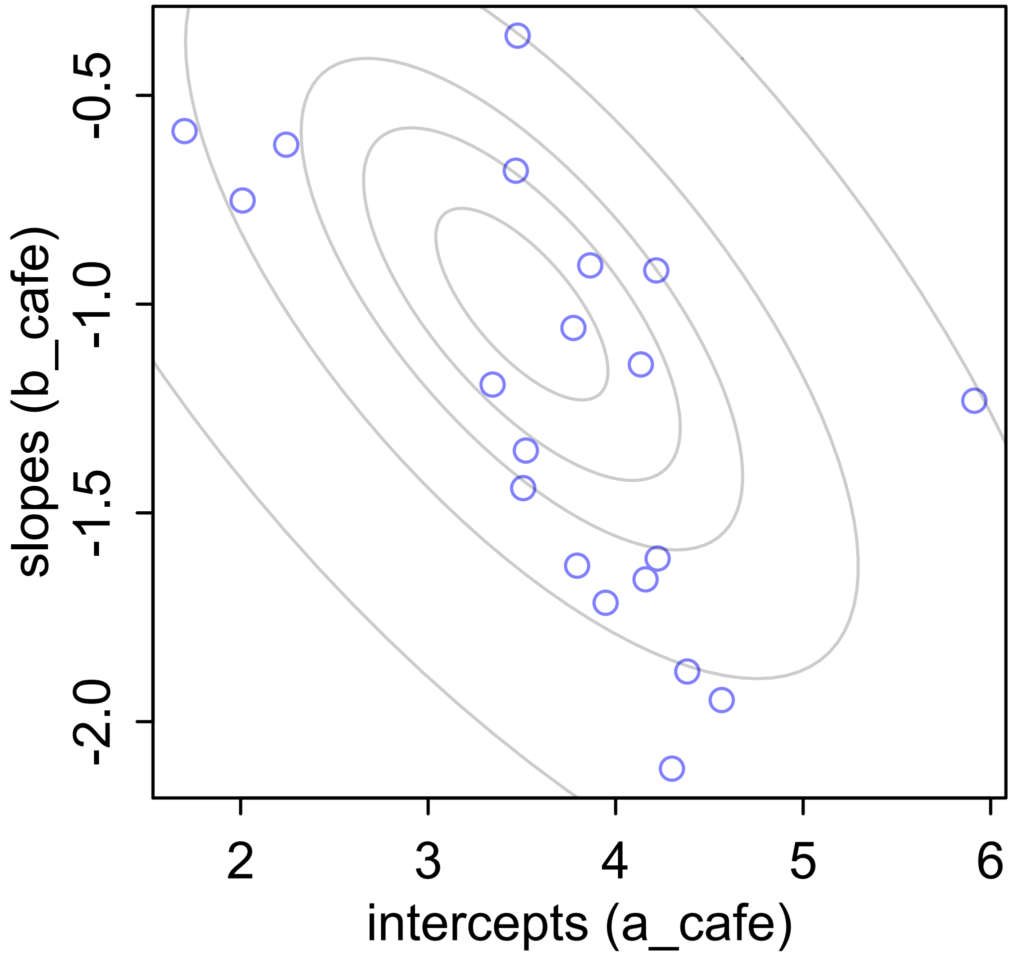
covariance

correlation

slopes variance



Simulated Cafés



20 cafés

5 days
morning & afternoon

200 observations

Varying slopes model

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} M_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \text{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

varying intercepts

varying slopes

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} M_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \text{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} M_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \text{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

multivariate prior

covariance matrix

pop avg intercept

$$
\begin{aligned}
W_i &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} M_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \mathrm{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \mathrm{Normal}(0, 10) \\
\beta &\sim \mathrm{Normal}(0, 10) \\
\sigma &\sim \mathrm{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \mathrm{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \mathrm{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \mathrm{LKJcorr}(2)
\end{aligned}
$$

pop avg slope

Covariance matrix shuffle

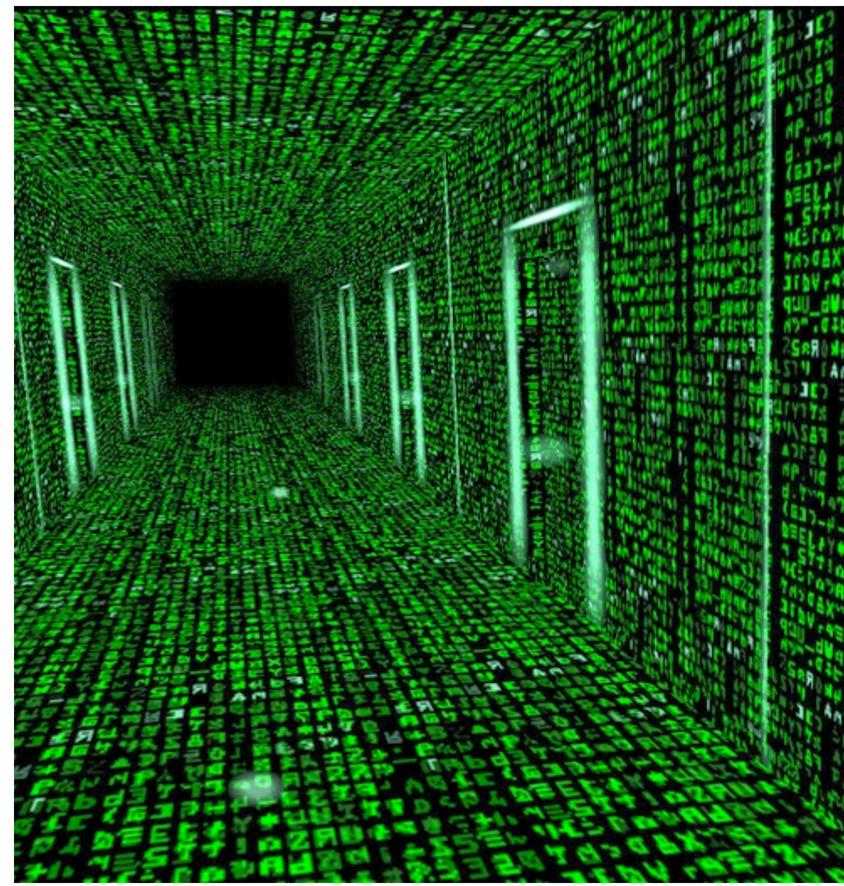
- m-by-m covariance matrix requires estimating
  - m standard deviations (or variances)
  - $(m^2 - m)/2$ correlations (or covariances)
  - total of $m(m + 1)/2$ parameters
- Several ways specify priors
  - Conjugate: inverse-Wishart (`inv_wishart`)
  - inverse-Wishart cannot pull apart stddev and correlations
  - Better to decompose:

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \underbrace{\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}}_{\mathbf{R}} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

Matrixes are nice

- Matrix algebra just shortcuts for working with lists of numbers
- A few simple rules
- Can you make an omelet?
  You can multiply matrixes.

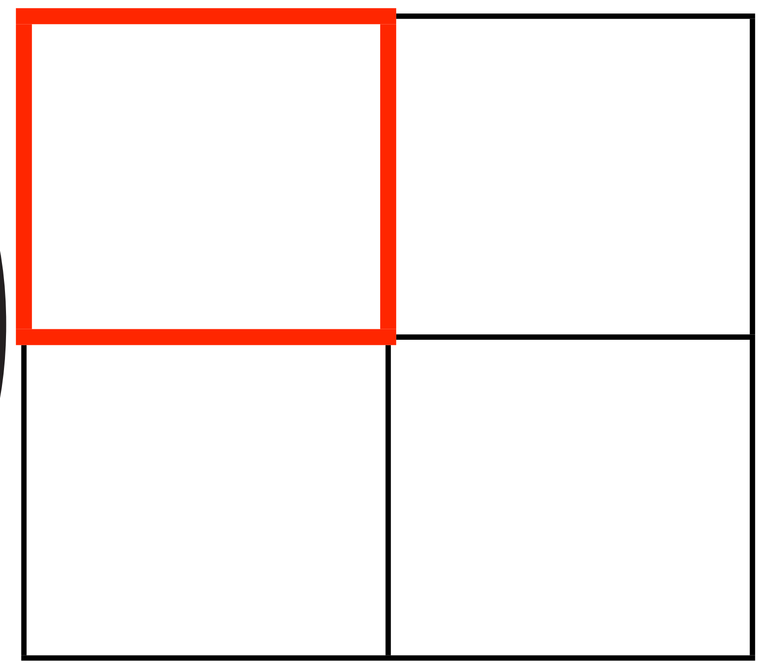
Case	Singular	Plural
nominative	mātrīx	mātrīcēs
genitive	mātrīcis	mātrīcum
dative	mātrīcī	mātrīcibus
accusative	mātrīcem	mātrīcēs
ablative	mātrīce	mātrīcibus
vocative	mātrīx	mātrīcēs



$$\begin{array}{cc} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ \begin{pmatrix} A & B \\ C & D \end{pmatrix} & \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \end{array}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$



$$\begin{array}{cc} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ \begin{pmatrix} A & B \\ C & D \end{pmatrix} & \begin{array}{|c|c|} \hline \boxed{Aa + Bc} & \\ \hline & \\ \hline \end{array} \end{array}$$

$$\begin{array}{cc} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ \begin{pmatrix} A & B \\ C & D \end{pmatrix} & \begin{array}{|c|c|} \hline Aa + Bc & Ab + Bd \\ \hline Ca + Dc & Cb + Dd \\ \hline \end{array} \end{array}$$

Matrixes are nice

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} = \mathbf{SRS}$$

Matrixes are nice

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} = \mathbf{SRS}$$

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \qquad ?$$

Matrixes are nice

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} = \mathbf{SRS}$$

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \quad \begin{pmatrix} \sigma_\alpha & \rho\sigma_\alpha \\ \rho\sigma_\beta & \sigma_\beta \end{pmatrix}$$

Matrixes are nice

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} = \mathbf{SRS}$$

$$\begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\alpha & \rho\sigma_\alpha \\ \rho\sigma_\beta & \sigma_\beta \end{pmatrix}$$

?

Matrixes are nice

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} = \mathbf{SRS}$$

$$\begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\alpha & \rho\sigma_\alpha \\ \rho\sigma_\beta & \sigma_\beta \end{pmatrix} \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}$$

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} M_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \text{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

build cov matrix

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} M_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \text{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \text{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

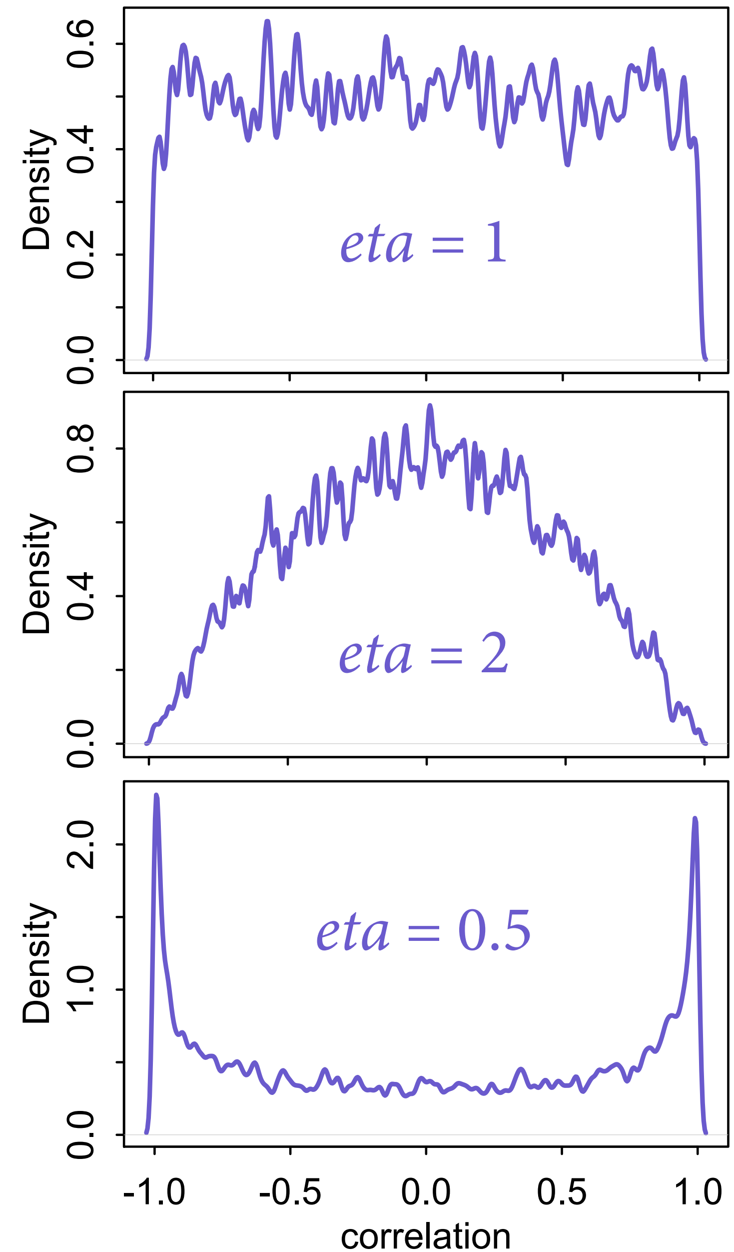
fixed (non-adaptive) priors

$$
\begin{aligned}
W_i &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\textsc{café}[i]} + \beta_{\textsc{café}[i]} M_i \\
\begin{bmatrix} \alpha_{\textsc{café}} \\ \beta_{\textsc{café}} \end{bmatrix} &\sim \mathrm{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \mathrm{Normal}(0, 10) \\
\beta &\sim \mathrm{Normal}(0, 10) \\
\sigma &\sim \mathrm{HalfCauchy}(0, 1) \\
\sigma_\alpha &\sim \mathrm{HalfCauchy}(0, 1) \\
\sigma_\beta &\sim \mathrm{HalfCauchy}(0, 1) \\
\mathbf{R} &\sim \mathrm{LKJcorr}(2)
\end{aligned}
$$

correlation matrix prior

LKJ Correlation prior

- After Lewandowski, Kurowicka, and Joe (LKJ) 2009
- One parameter, *eta*, specifies concentration or dispersion from *identity matrix* (zero correlations)
  - *eta* = 1, uniform correlation matrices
  - *eta* > 1, stomps on extreme correlations
  - *eta* < 1, elevates extreme correlations



Varying slopes estimation

```
m13.1 <- map2stan(
    alist(
        wait ~ dnorm( mu , sigma ),
        mu <- a_cafe[cafe] + b_cafe[cafe]*afternoon,
        c(a_cafe,b_cafe)[cafe] ~ dmvnorm2(c(a,b),sigma_cafe,Rho),
        a ~ dnorm(0,10),
        b ~ dnorm(0,10),
        sigma_cafe ~ dcauchy(0,2),
        sigma ~ dcauchy(0,2),
        Rho ~ dlkjcorr(2)
    ) ,
    data=d ,
    iter=5000 , warmup=2000 , chains=2 )
```

Varying slopes estimation

```
m13.1 <- map2stan(
    alist(
        wait ~ dnorm( mu , sigma ),
        mu <- a_cafe[cafe] + b_cafe[cafe]*afternoon,
        c(a_cafe,b_cafe)[cafe] ~ dmvnorm2(c(a,b),sigma_cafe,Rho),
        a ~ dnorm(0,10),
        b ~ dnorm(0,10),
        sigma_cafe ~ dcauchy(0,2),
        sigma ~ dcauchy(0,2),
        Rho ~ dlkjcorr(2)
    ) ,
    data=d ,
    iter=5000 , warmup=2000 , chains=2 )
```

Varying slopes estimation

```
m13.1 <- map2stan(
    alist(
        wait ~ dnorm( mu , sigma ),
        mu <- a_cafe[cafe] + b_cafe[cafe]*afternoon,
        c(a_cafe,b_cafe)[cafe] ~ dmvnorm2(c(a,b),sigma_cafe,Rho),
        a ~ dnorm(0,10),
        b ~ dnorm(0,10),
        sigma_cafe ~ dcauchy(0,2),
        sigma ~ dcauchy(0,2),
        Rho ~ dlkjcorr(2)
    ) ,
    data=d ,
    iter=5000 , warmup=2000 , chains=2 )
```

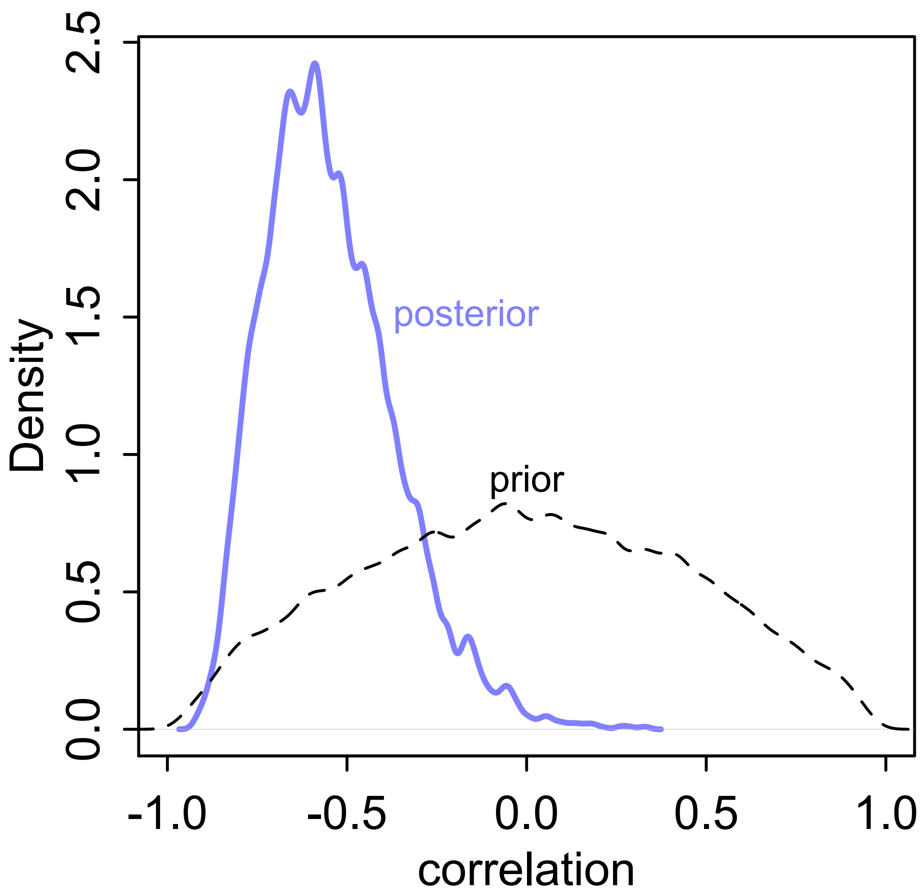
Varying slopes estimation

```
m13.1 <- map2stan(
    alist(
        wait ~ dnorm( mu , sigma ),
        mu <- a_cafe[cafe] + b_cafe[cafe]*afternoon,
        c(a_cafe,b_cafe)[cafe] ~ dmvnorm2(c(a,b),sigma_cafe,Rho),
        a ~ dnorm(0,10),
        b ~ dnorm(0,10),
        sigma_cafe ~ dcauchy(0,2),
        sigma ~ dcauchy(0,2),
        Rho ~ dlkjcorr(2)
    ) ,
    data=d ,
    iter=5000 , warmup=2000 , chains=2 )
```

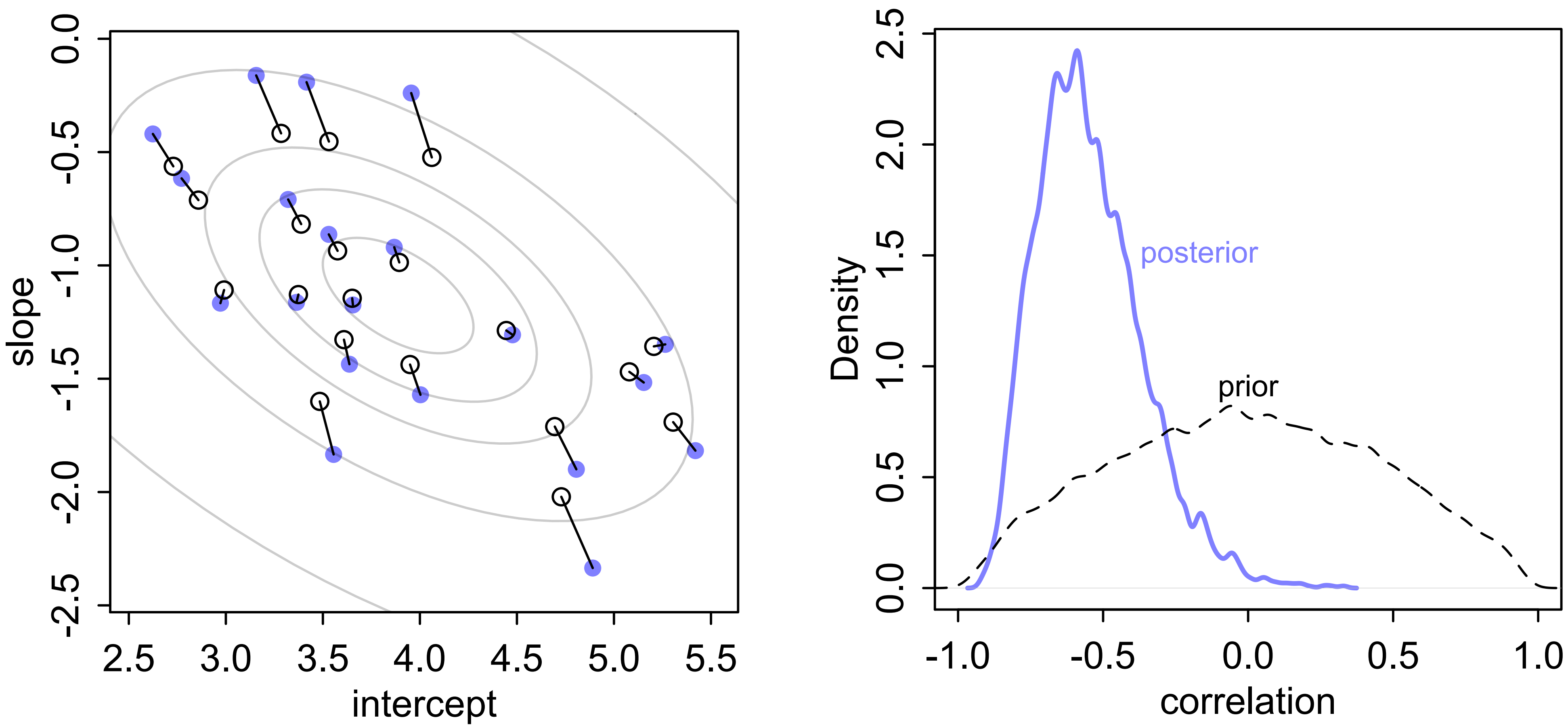
Posterior correlation

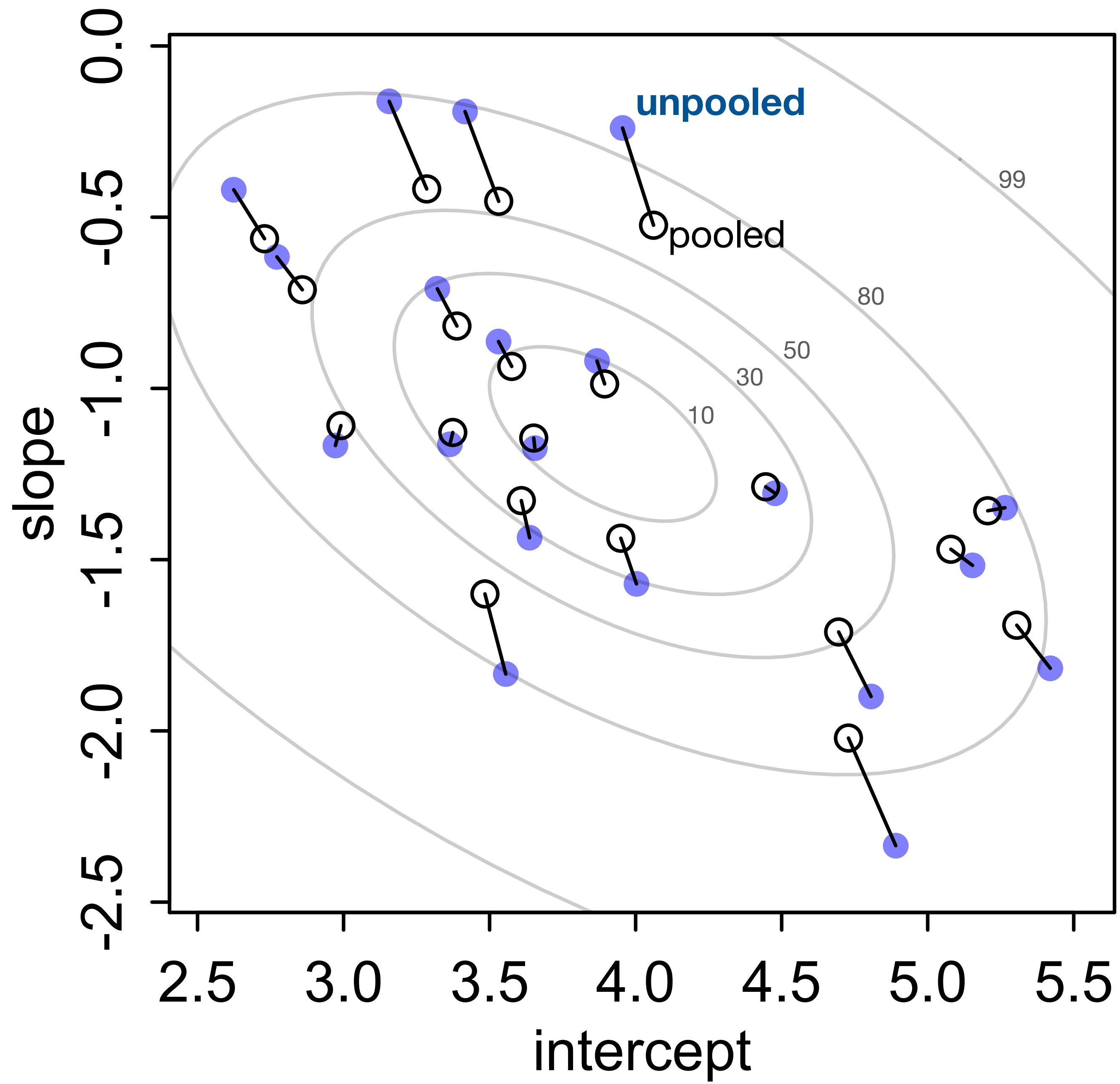
R code 13.12

```
post <- extract.samples(m13.1)
dens( post$Rho[,1,2] )
```

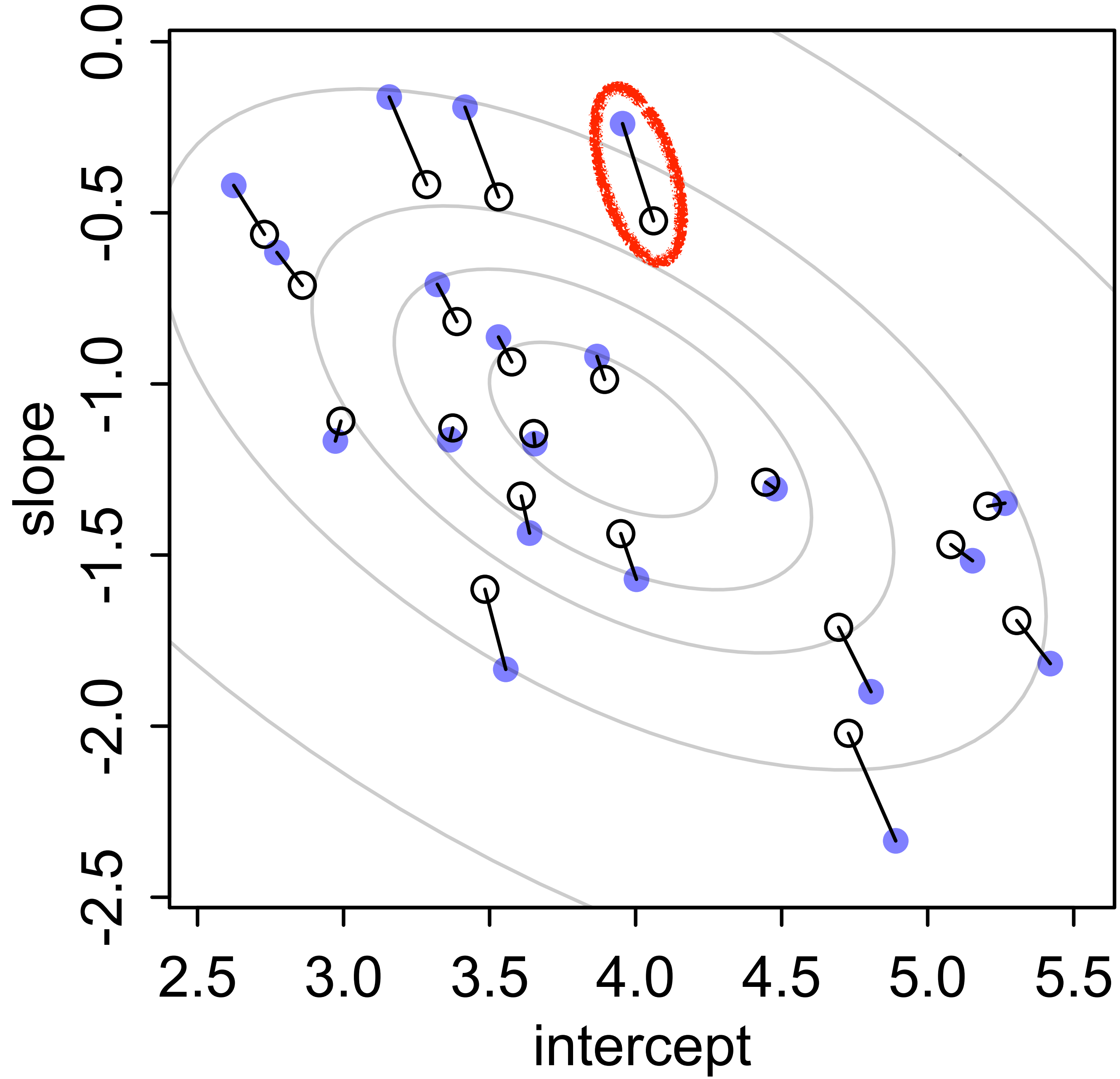


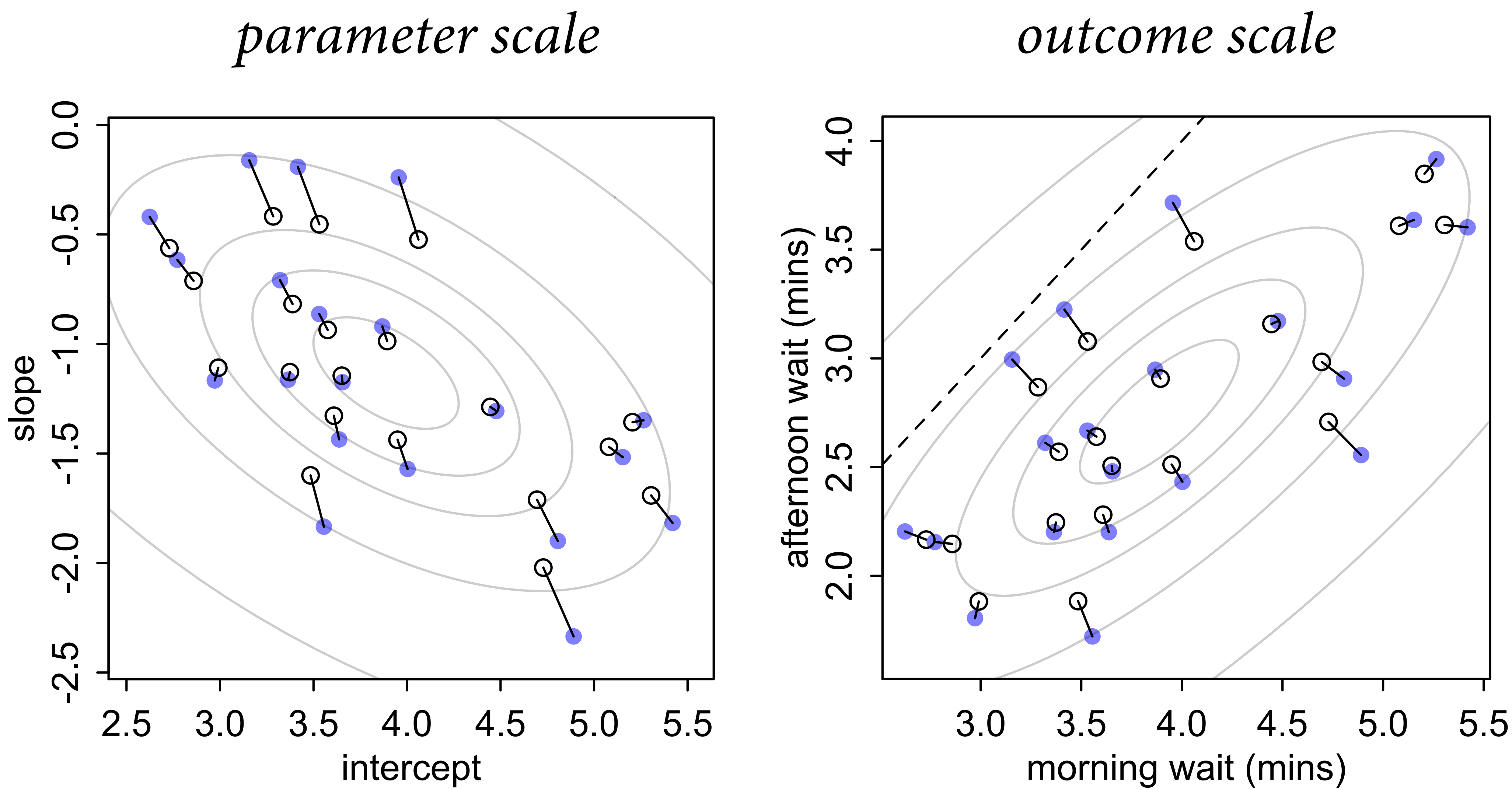
Posterior shrinkage





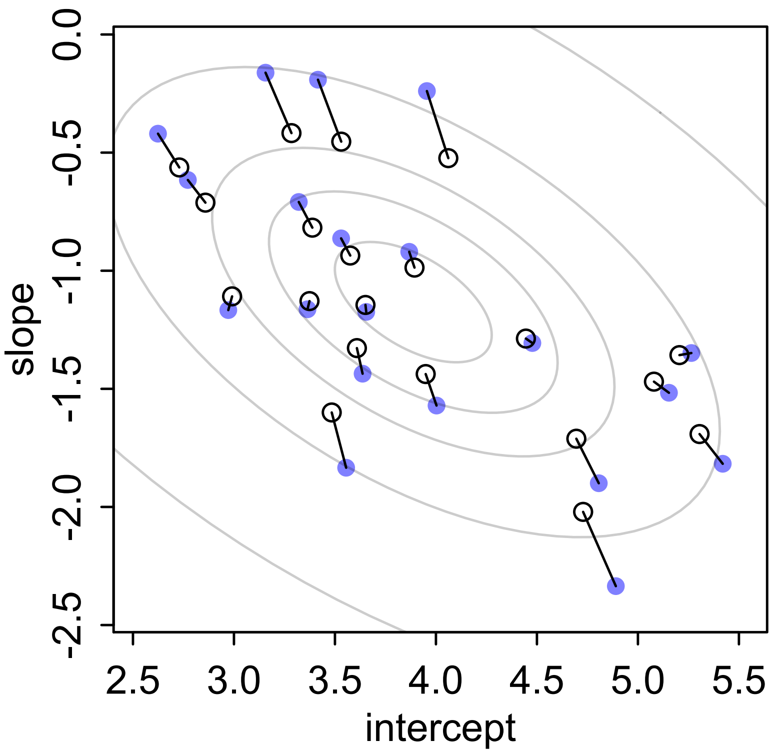
unpooled
pooled
99
80
50
30
10
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
slope
2.5
3.0
3.5
4.0
4.5
5.0
5.5
intercept


slope
intercept
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
2.5
3.0
3.5
4.0
4.5
5.0
5.5


parameter scale
outcome scale
slope
intercept
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
2.5
3.0
3.5
4.0
4.5
5.0
5.5
afternoon wait (mins)
morning wait (mins)
4.0
3.5
3.0
2.5
2.0
3.0
3.5
4.0
4.5
5.0
5.5

Multi-dimensional shrinkage

- Joint distribution of varying effects pools information across intercepts & slopes
- Correlation btw effects => shrinkage in one dimension induces shrinkage in others
- Improved accuracy, just like varying intercepts



Example: UCB admit data again

	dept	applicant.gender	admit	reject	applications	male	i	j
1	A	male	512	313	825	1	1	1
2	A	female	89	19	108	0	2	1
3	B	male	353	207	560	1	3	2
4	B	female	17	8	25	0	4	2
5	C	male	120	205	325	1	5	3
6	C	female	202	391	593	0	6	3
7	D	male	138	279	417	1	7	4
8	D	female	131	244	375	0	8	4
9	E	male	53	138	191	1	9	5
10	E	female	94	299	393	0	10	5
11	F	male	22	351	373	1	11	6
12	F	female	24	317	341	0	12	6

Varying intercepts by dept

$$
\begin{aligned}
A_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{DEPT}[i]} + \beta m_i \\
\alpha_{\text{DEPT}} &\sim \text{Normal}(\alpha, \sigma) \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Normal}(0, 1) \\
\sigma &\sim \text{HalfCauchy}(0, 2)
\end{aligned}
$$

Varying slopes by dept

$$
\begin{aligned}
A_i &\sim \mathrm{Binomial}(n_i, p_i) \\
\mathrm{logit}(p_i) &= \alpha_{\mathrm{DEPT}[i]} + \beta_{\mathrm{DEPT}[i]} m_i \\
\begin{bmatrix} \alpha_{\mathrm{DEPT}} \\ \beta_{\mathrm{DEPT}} \end{bmatrix} &\sim \mathrm{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \mathrm{Normal}(0, 10) \\
\beta &\sim \mathrm{Normal}(0, 1) \\
(\sigma_\alpha, \sigma_\beta) &\sim \mathrm{HalfCauchy}(0, 2) \\
\mathbf{R} &\sim \mathrm{LKJcorr}(2)
\end{aligned}
$$