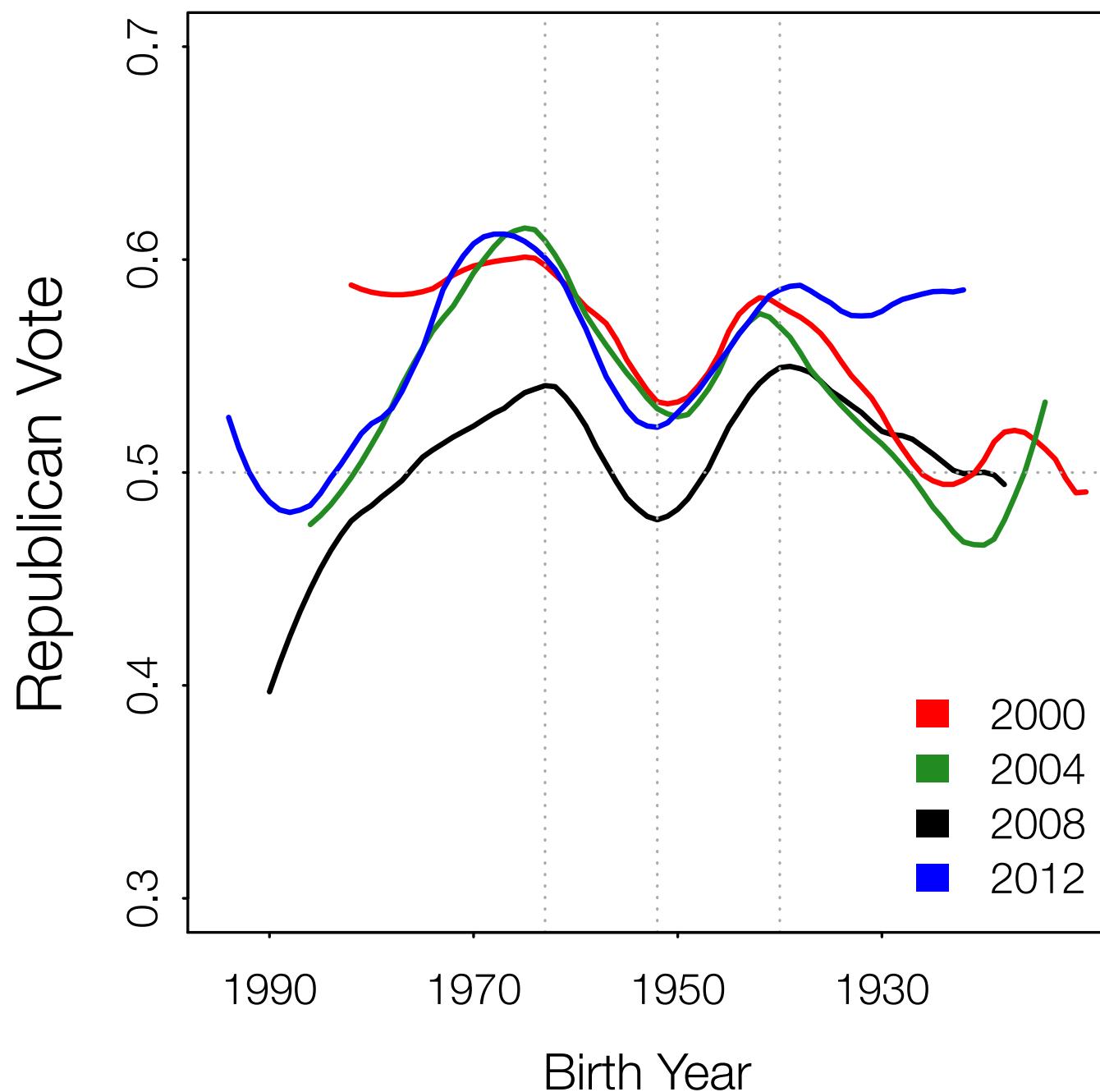


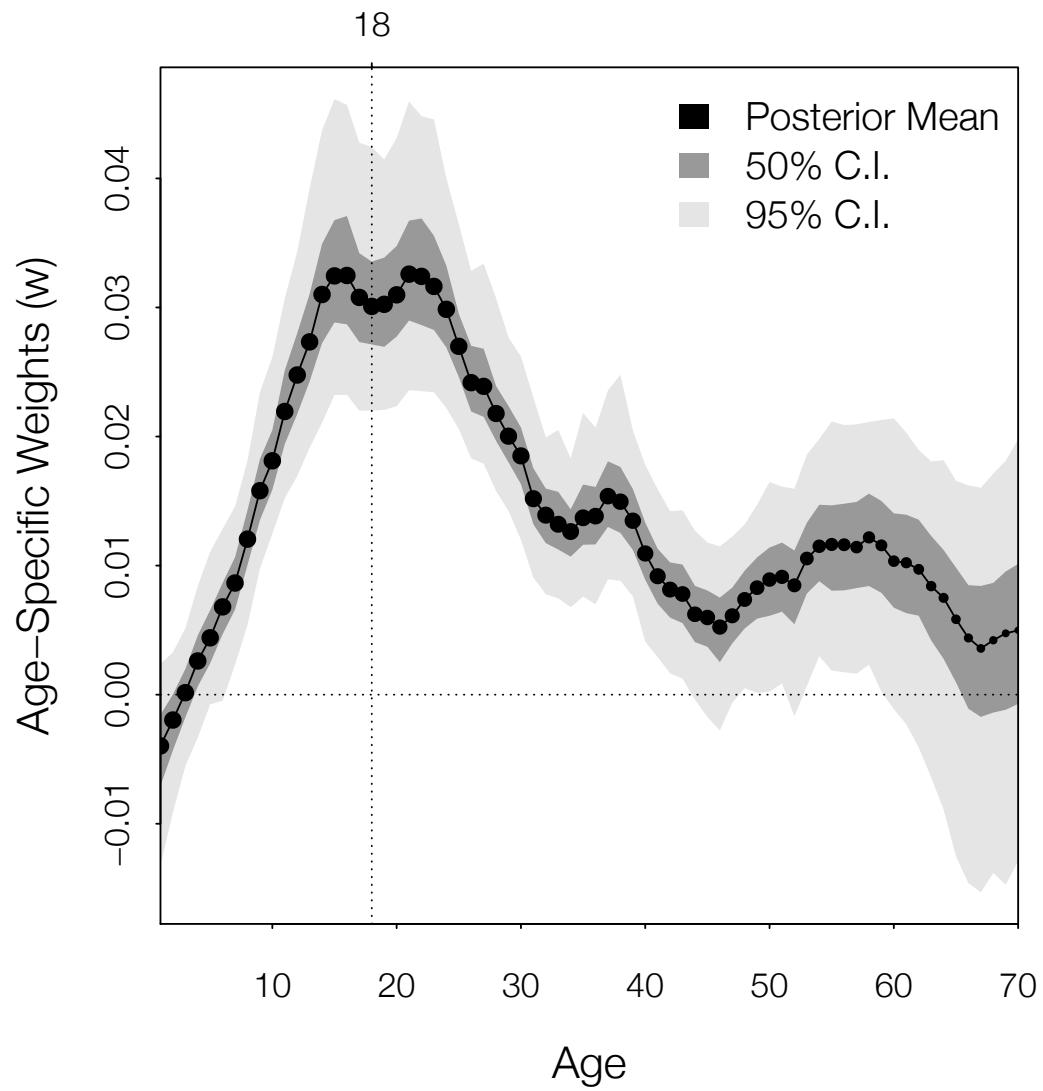
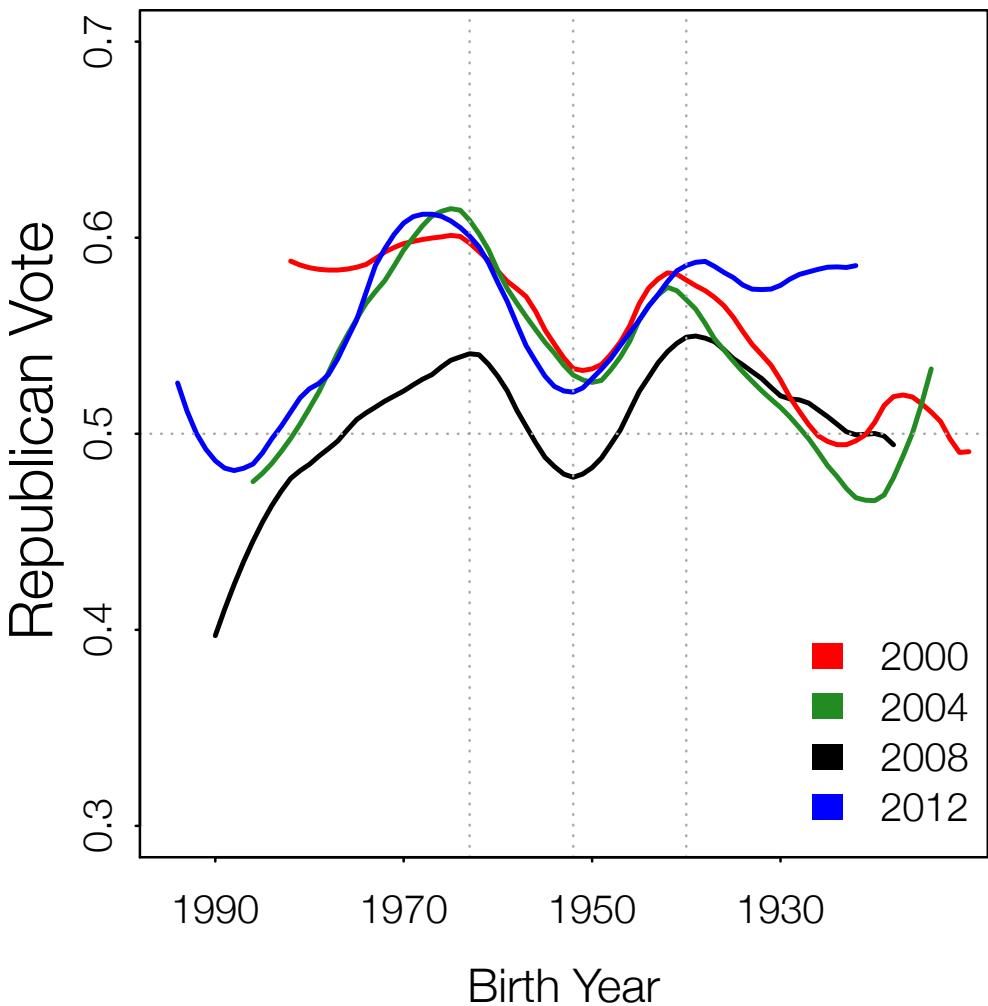
Statistical Rethinking

Winter 2019

Lecture 19 / Week 10

Gaussian Processes





The Great Society, Reagan's Revolution, and Generations of Presidential Voting

Yair Ghitza*

Andrew Gelman†

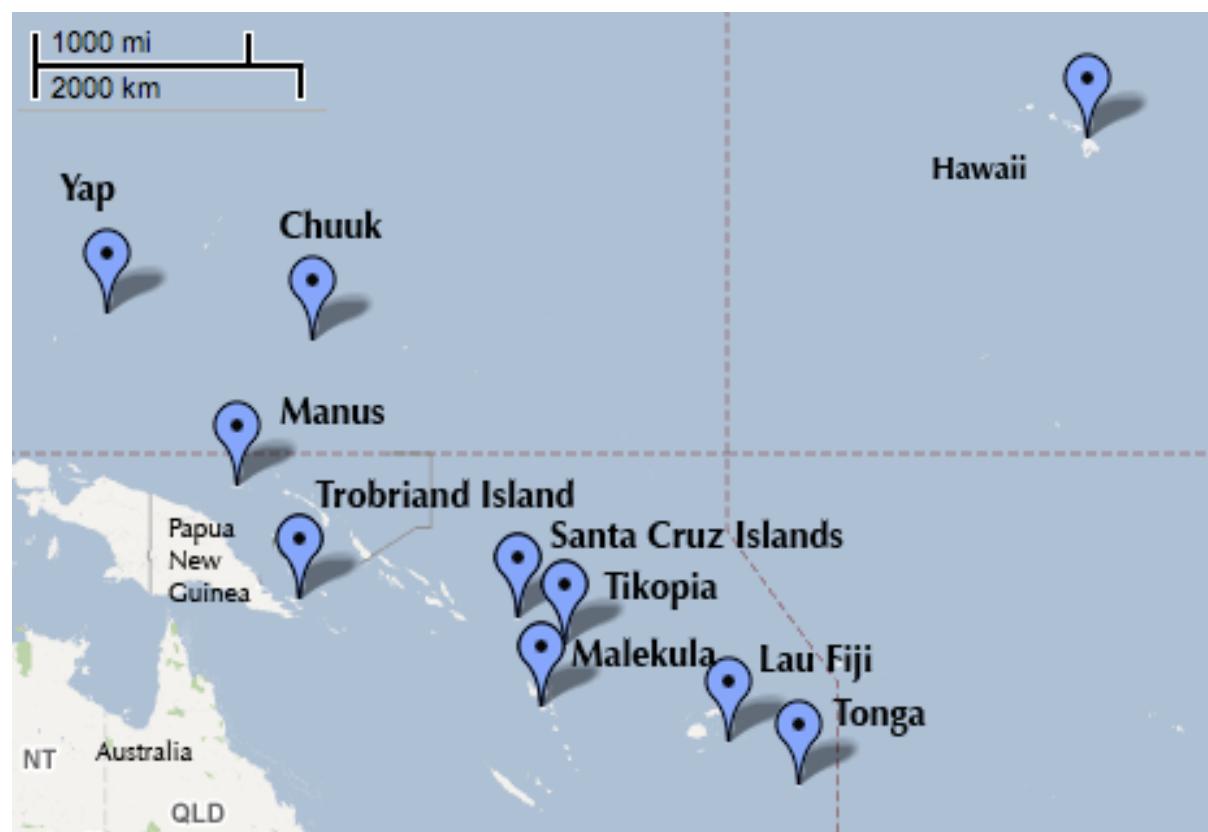
July 7, 2014

Continuous categories

- Traditional clusters discrete, unordered => every category equally different from all others (in prior)
- Continuous dimensions of difference:
 - Age, income, location, phylogenetic distance, social network distance, many others
 - No obvious cut points in continuum, but close values share common exposures/covariates/interactions
- Would like to exploit pooling in these cases as well
- Common approach: **Gaussian process regression**

Spatial autocorrelation

- Relationship between tool complexity and population
- Close societies may share tools because of contact or similar geology/ecology
- Use space as proxy
- Spatial autocorrelation



1000 mi
2000 km



	Ml	Ti	SC	Ya	Fi	Tr	Ch	Mn	To	Ha
Malekula	0.0	0.5	0.6	4.4	1.2	2.0	3.2	2.8	1.9	5.7
Tikopia	0.5	0.0	0.3	4.2	1.2	2.0	2.9	2.7	2.0	5.3
Santa Cruz	0.6	0.3	0.0	3.9	1.6	1.7	2.6	2.4	2.3	5.4
Yap	4.4	4.2	3.9	0.0	5.4	2.5	1.6	1.6	6.1	7.2
Lau Fiji	1.2	1.2	1.6	5.4	0.0	3.2	4.0	3.9	0.8	4.9
Trobriand	2.0	2.0	1.7	2.5	3.2	0.0	1.8	0.8	3.9	6.7
Chuuk	3.2	2.9	2.6	1.6	4.0	1.8	0.0	1.2	4.8	5.8
Manus	2.8	2.7	2.4	1.6	3.9	0.8	1.2	0.0	4.6	6.7
Tonga	1.9	2.0	2.3	6.1	0.8	3.9	4.8	4.6	0.0	5.0
Hawaii	5.7	5.3	5.4	7.2	4.9	6.7	5.8	6.7	5.0	0.0

distances in thousand km

$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = \alpha P_i^\beta / \gamma$$

expected tools *population size*



$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = \exp(k_{\text{SOCIETY}[i]}) \alpha P_i^\beta / \gamma$$

*expected
tools*

Varying factor



$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = \exp(k_{\text{SOCIETY}[i]}) \alpha P_i^\beta / \gamma$$

expected tools *Varying factor*

- Works like a proportion:

$$k = 0 \quad \exp(0) = 1 \quad \text{As expected}$$

$$k = -0.5 \quad \exp(-0.5) = 0.6 \quad 60\% \text{ of expected}$$

$$k = 0.25 \quad \exp(0.25) = 1.3 \quad 130\% \text{ of expected}$$

Unfamiliar prior

- Gaussian process prior:
 - Multivariate Gaussian
 - Means all zero (usually)
 - Model the covariance matrix using pairwise distances

vector of factors

$$\begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ \dots \\ k_{10} \end{pmatrix}$$

$$\sim \text{MVNormal} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \mathbf{K} \right)$$

covariance matrix

$$\mathbf{K}_{ij} = \eta^2 \exp(-\rho^2 D_{ij}^2) + \delta_{ij}\sigma^2$$

Modeling covariance

$$K_{ij} = \eta^2 \exp(-\rho^2 D_{ij}^2) + \delta_{ij}\sigma^2$$

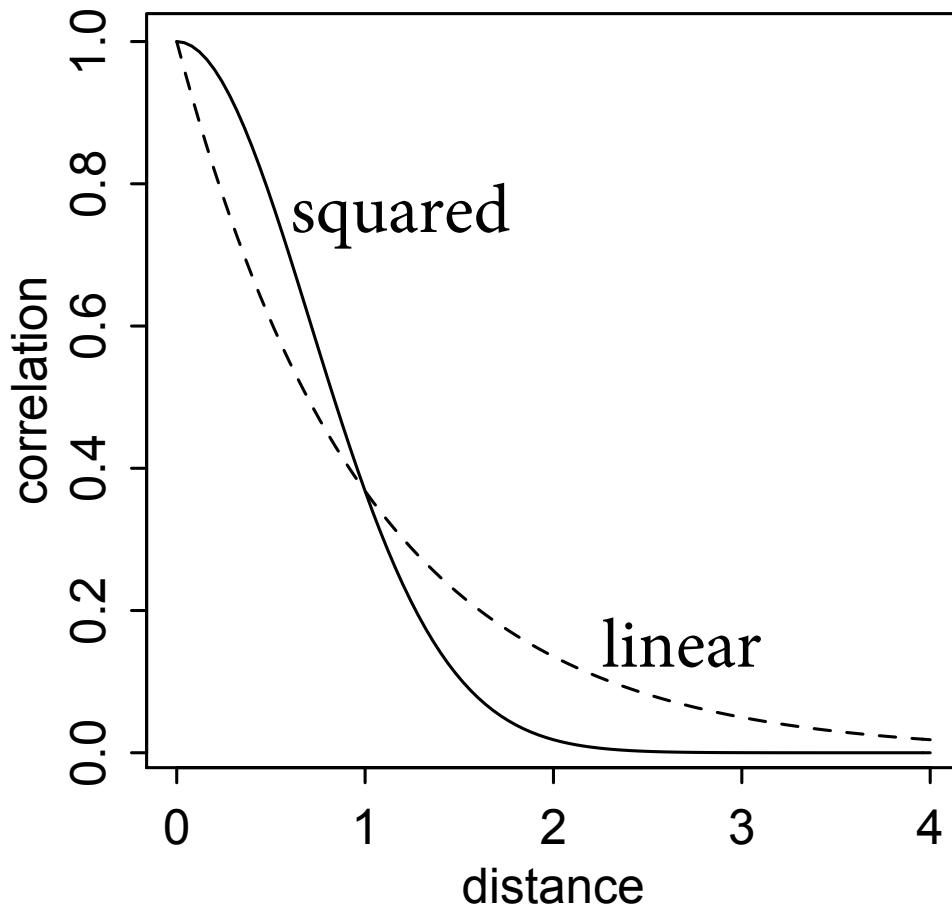
Diagram illustrating the components of the covariance model:

- covariance btw islands i & j* points to K_{ij}
- max cov* points to η^2
- squared distance* points to D_{ij}^2
- rate of decline with distance* points to ρ^2
- variance self* points to $\delta_{ij}\sigma^2$



Modeling covariance

$$K_{ij} = \eta^2 \exp(-\rho^2 D_{ij}^2) + \delta_{ij}\sigma^2$$



Linear: Cov declines fastest at near distances.

Squared: Cov declines fastest at intermediate distances.

Putting it all together

$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = \exp(k_{\text{SOCIETY}[i]}) \alpha P_i^\beta / \gamma$$

$$\mathbf{k} \sim \text{MVNormal}((0, \dots, 0), \mathbf{K})$$

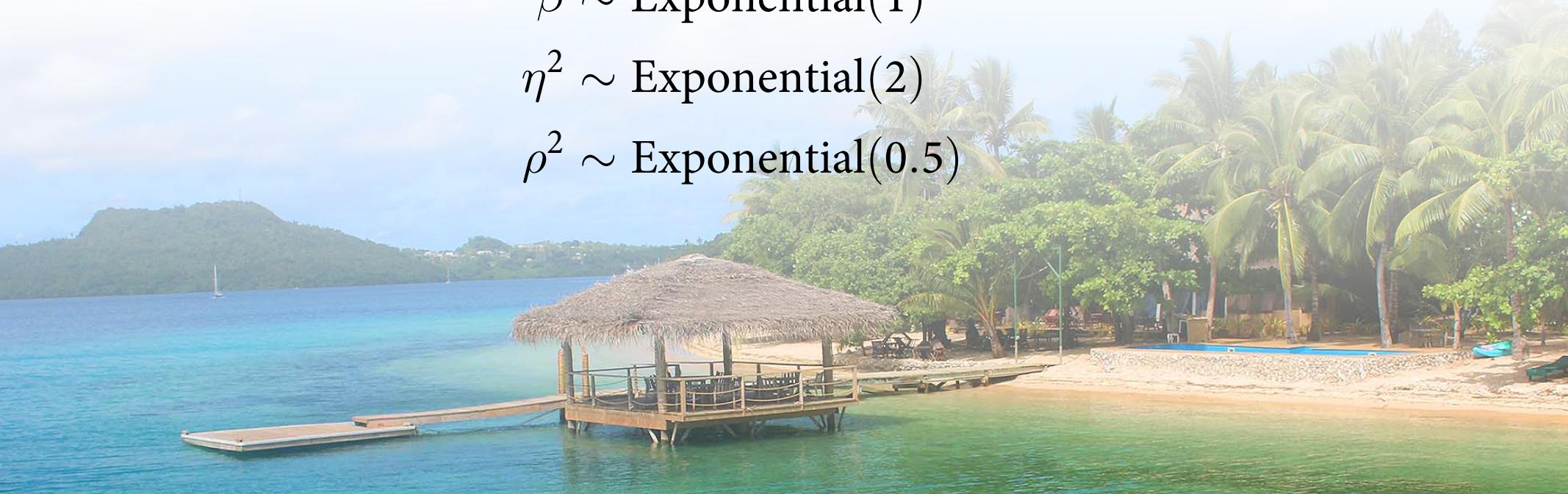
$$\mathbf{K}_{ij} = \eta^2 \exp(-\rho^2 D_{ij}^2) + \delta_{ij}(0.01)$$

$$\alpha \sim \text{Exponential}(1)$$

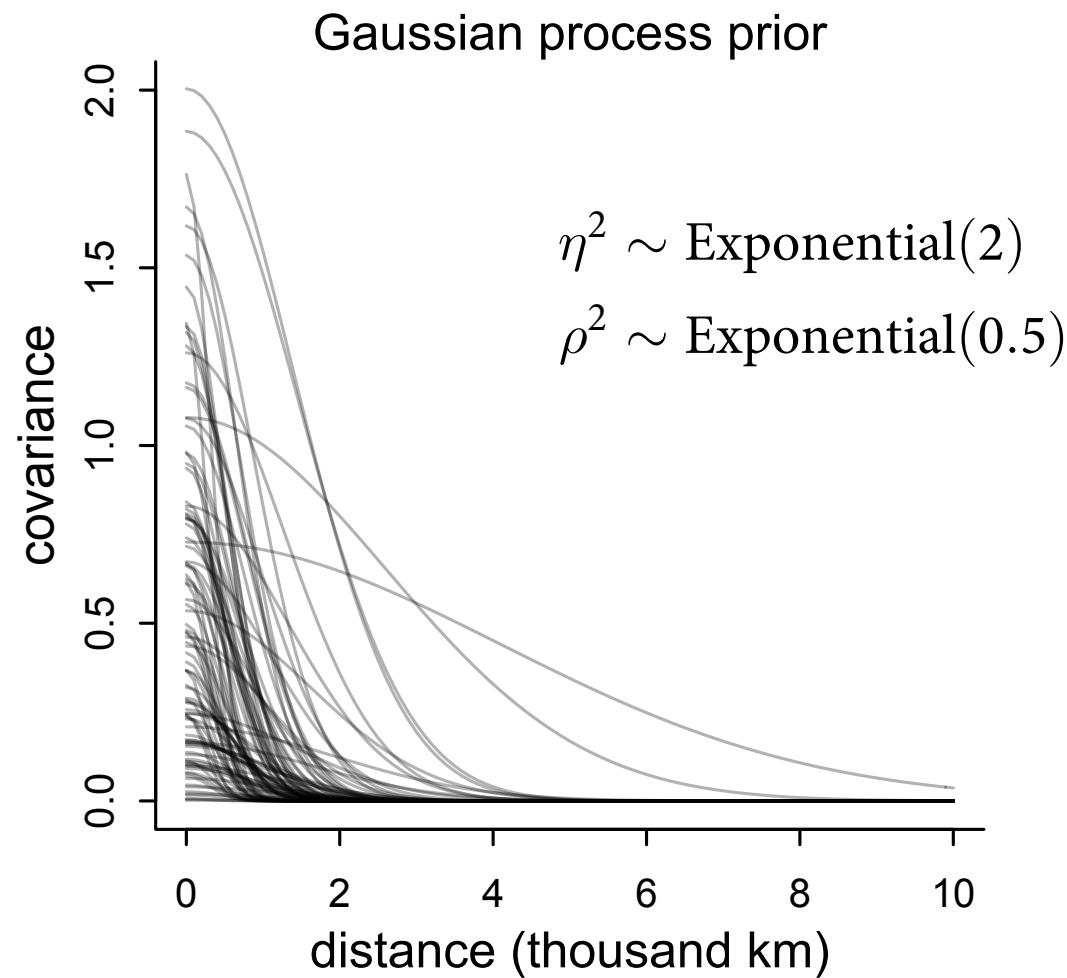
$$\beta \sim \text{Exponential}(1)$$

$$\eta^2 \sim \text{Exponential}(2)$$

$$\rho^2 \sim \text{Exponential}(0.5)$$



Prior covariance functions



Coding

```
m14.7 <- ulam(  
  alist(  
    T ~ dpois(lambda),  
    lambda <- (a*P^b/g)*exp(k[society]),  
  
    vector[10]:k ~ multi_normal( 0 , SIGMA ),  
    matrix[10,10]:SIGMA <- cov_GPL2( Dmat , etasq , rhosq , 0.01 ),  
  
    c(a,b,g) ~ dexp( 1 ),  
    etasq ~ dexp( 2 ),  
    rhosq ~ dexp( 0.5 )  
  ), data=dat_list , chains=4 , cores=4 , iter=2000 )
```

Coding

```
m14.7 <- ulam(  
  alist(  
    T ~ dpois(lambda),  
    lambda <- (a*P^b/g)*exp(k[society]) ,  
  
    vector[10]:k ~ multi_normal( 0 , SIGMA ),  
    matrix[10,10]:SIGMA <- cov_GPL2( Dmat , etasq , rhosq , 0.01 ) ,  
  
    c(a,b,g) ~ dexp( 1 ),  
    etasq ~ dexp( 2 ),  
    rhosq ~ dexp( 0.5 )  
  ), data=dat_list , chains=4 , cores=4 , iter=2000 )
```

Coding

```
m14.7 <- ulam(  
  alist(  
    T ~ dpois(lambda),  
    lambda <- (a*P^b/g)*exp(k[society]),  
  
    vector[10]:k ~ multi_normal( 0 , SIGMA ),  
    matrix[10,10]:SIGMA <- cov_GPL2( Dmat , etasq , rhosq , 0.01 ),  
  
    c(a,b,g) ~ dexp( 1 ),  
    etasq ~ dexp( 2 ),  
    rhosq ~ dexp( 0.5 )  
  ), data=dat_list , chains=4 , cores=4 , iter=2000 )
```

Coding

```
m14.7 <- ulam(  
  alist(  
    T ~ dpois(lambda),  
    lambda <- (a*P^b/g)*exp(k[society]) ,  
  
    vector[10]:k ~ multi_normal( 0 , SIGMA ),  
    matrix[10,10]:SIGMA <- cov_GPL2( Dmat , etasq , rhosq , 0.01 ) ,  
  
    c(a,b,g) ~ dexp( 1 ) ,  
    etasq ~ dexp( 2 ) ,  
    rhosq ~ dexp( 0.5 )  
  ), data=dat_list , chains=4 , cores=4 , iter=2000 )
```

Marginal posterior

- k values on log scale, so a bit opaque

R code
14.39

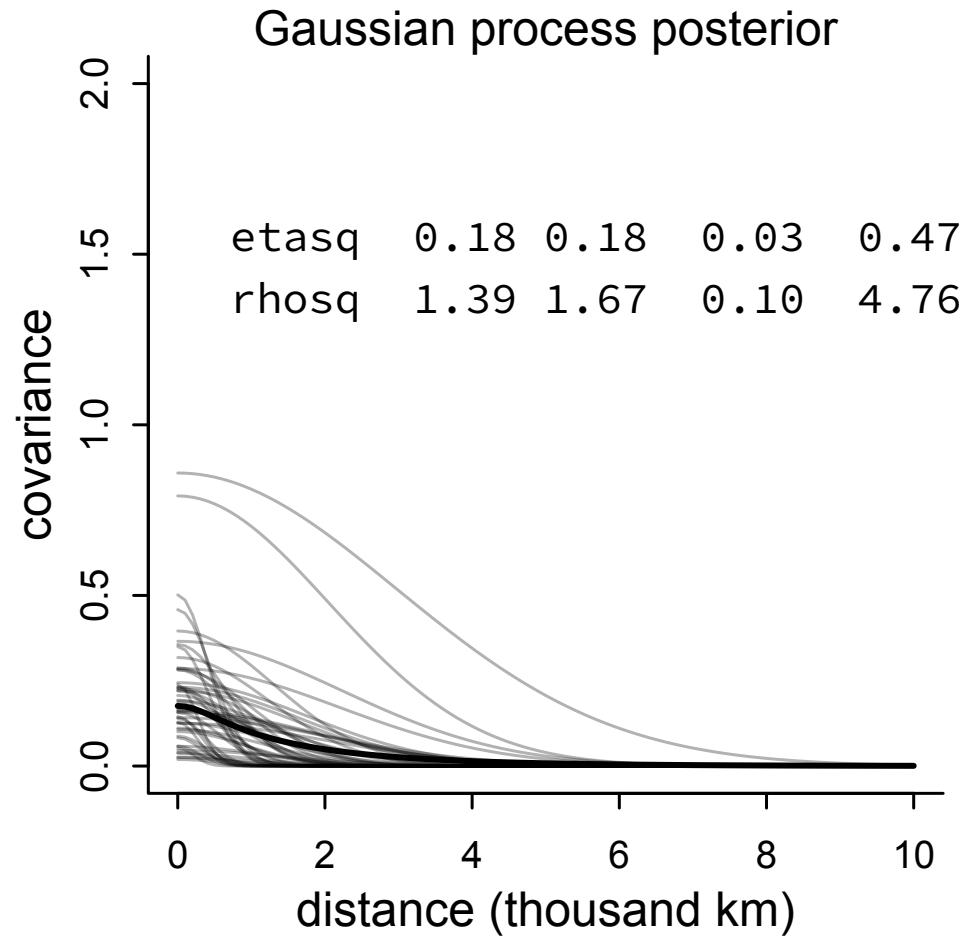
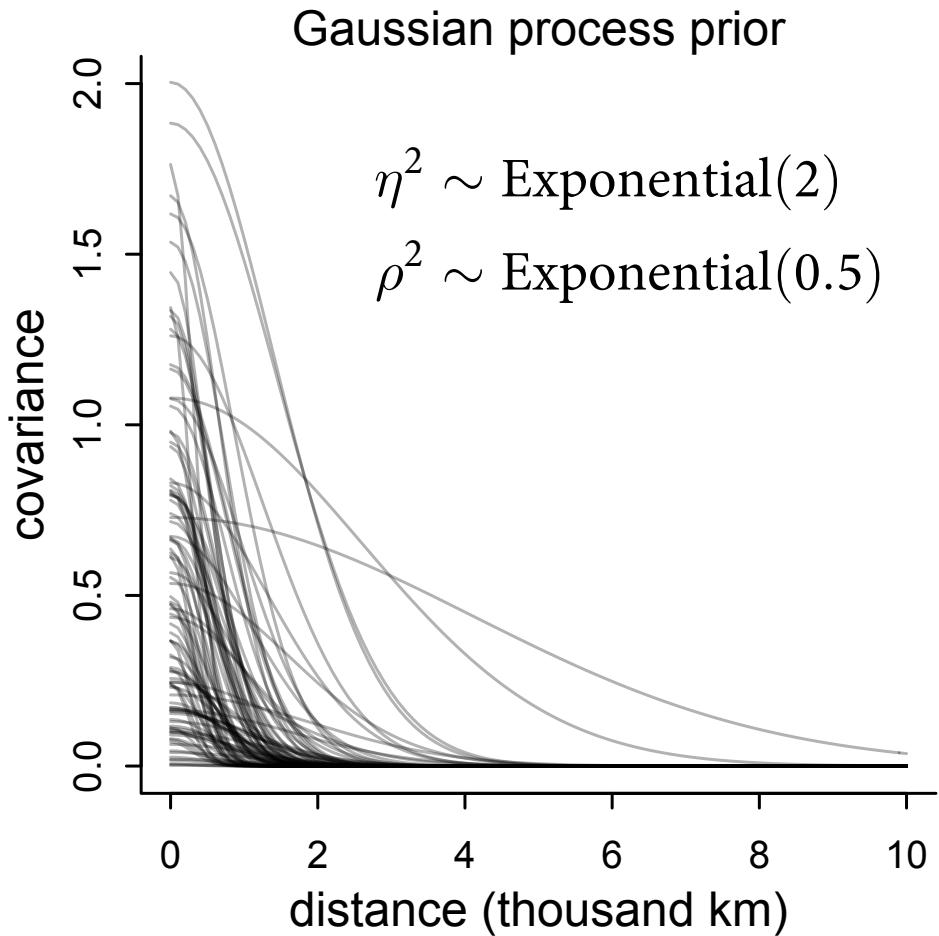
```
precis( m14.7 , depth=3 )
```

	mean	sd	5.5%	94.5%	n_eff	Rhat
k[1]	-0.14	0.29	-0.60	0.30	793	1.01
k[2]	0.00	0.27	-0.40	0.43	676	1.01
k[3]	-0.04	0.27	-0.43	0.37	687	1.01
k[4]	0.37	0.26	0.00	0.77	715	1.01
k[5]	0.10	0.25	-0.26	0.48	729	1.01
k[6]	-0.35	0.26	-0.75	0.02	834	1.00
k[7]	0.16	0.25	-0.19	0.54	732	1.01
k[8]	-0.18	0.25	-0.57	0.18	796	1.01
k[9]	0.29	0.23	-0.04	0.66	742	1.01
k[10]	-0.14	0.32	-0.65	0.33	1135	1.01
a	1.44	1.07	0.27	3.40	1984	1.00
b	0.28	0.08	0.15	0.41	1212	1.01
g	0.63	0.57	0.08	1.69	1685	1.01
etasq	0.18	0.18	0.03	0.47	951	1.00
rhosq	1.39	1.67	0.10	4.76	1805	1.00



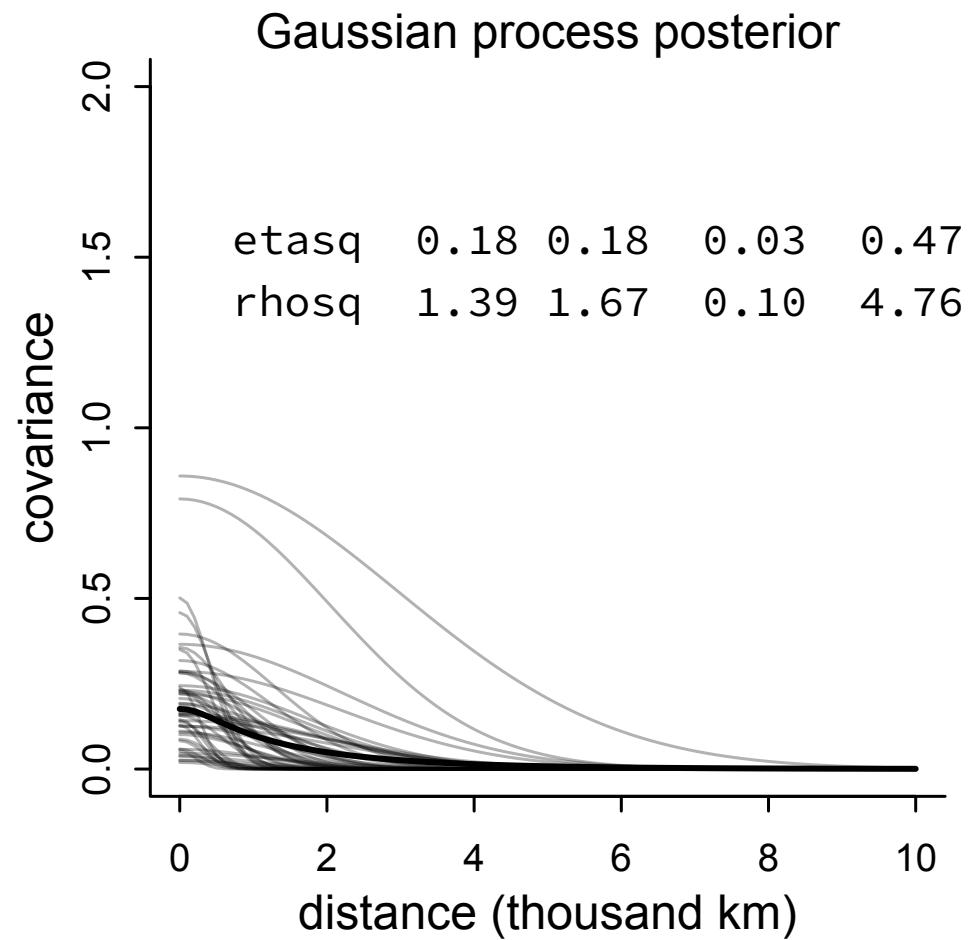
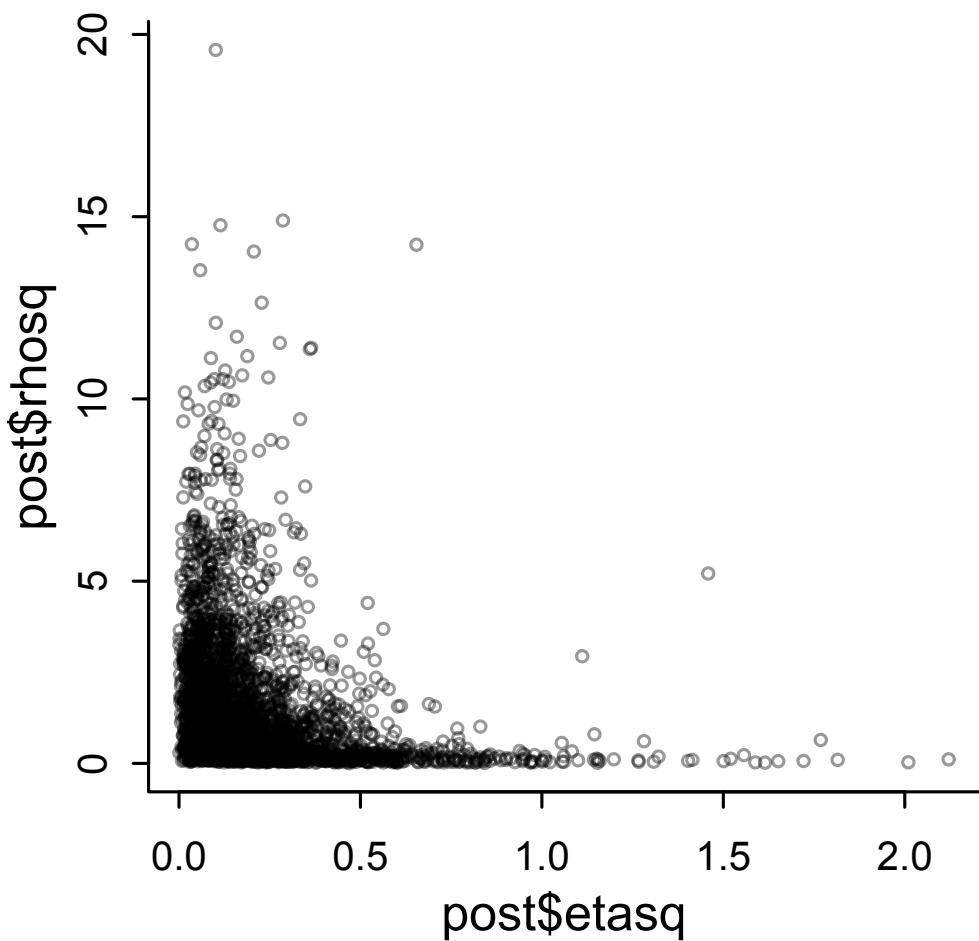
Covariance function

$$K_{ij} = \eta^2 \exp(-\rho^2 D_{ij}^2) + \delta_{ij}(0.01)$$



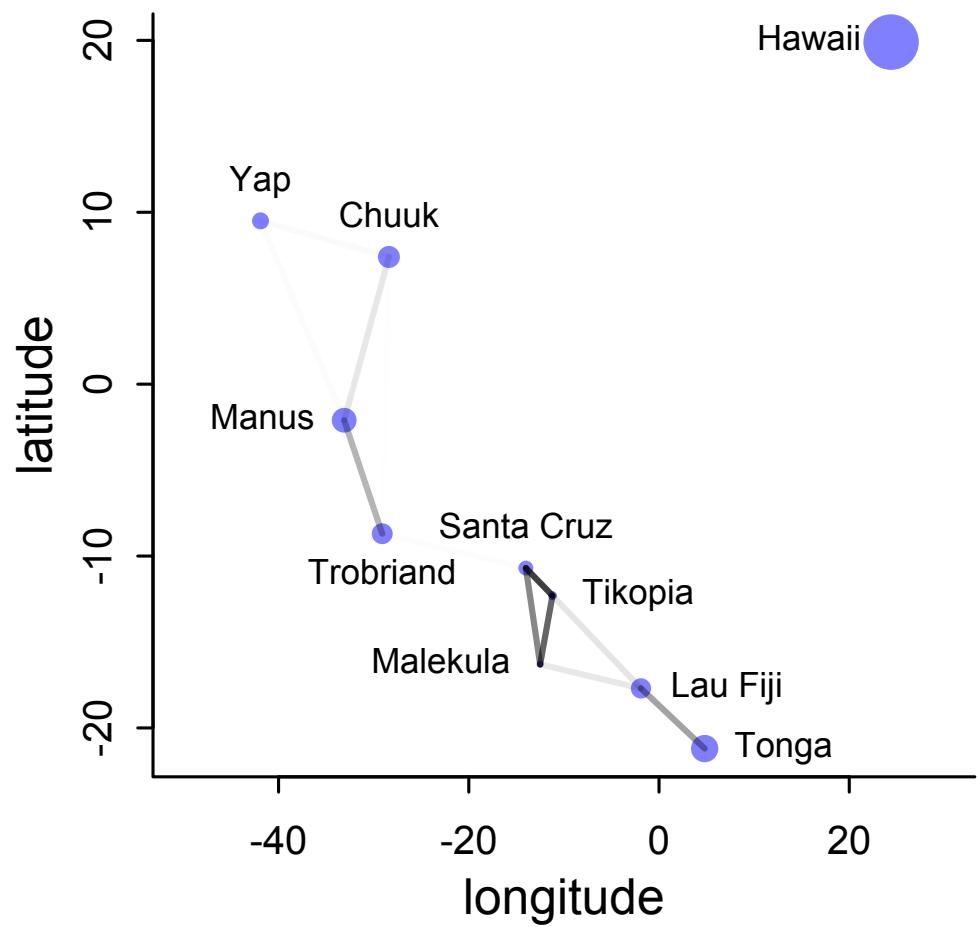
Covariance function

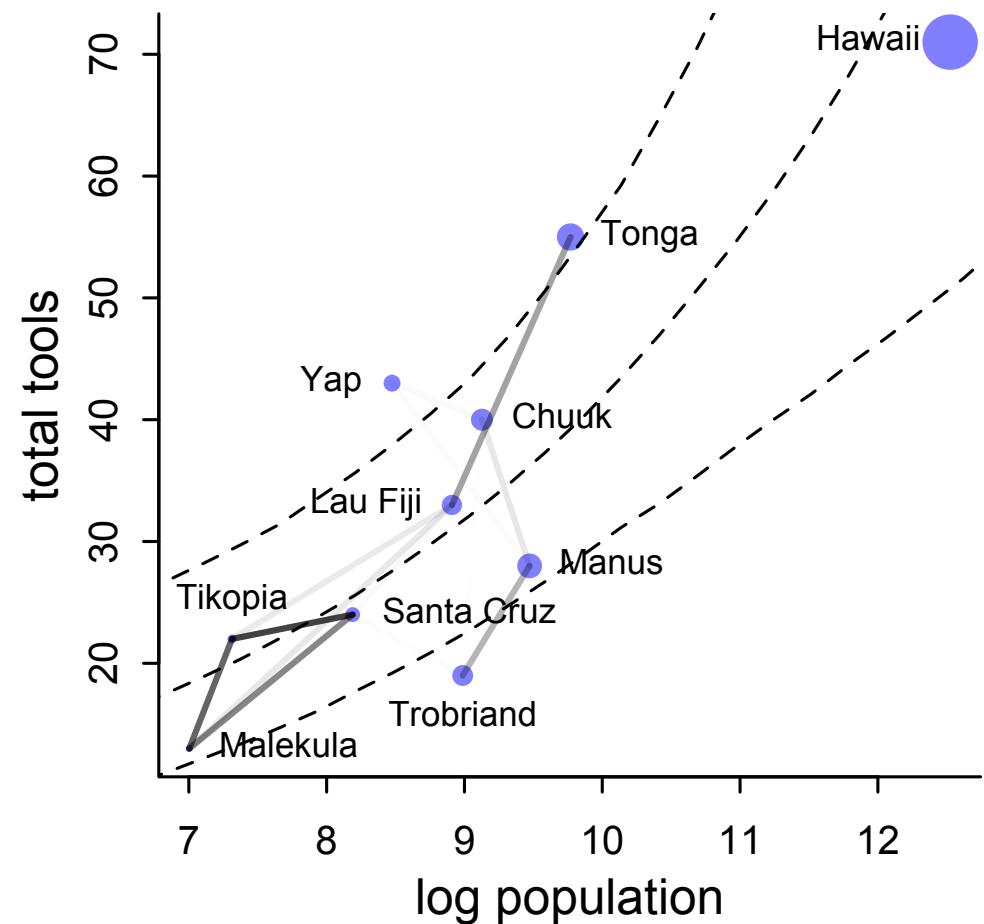
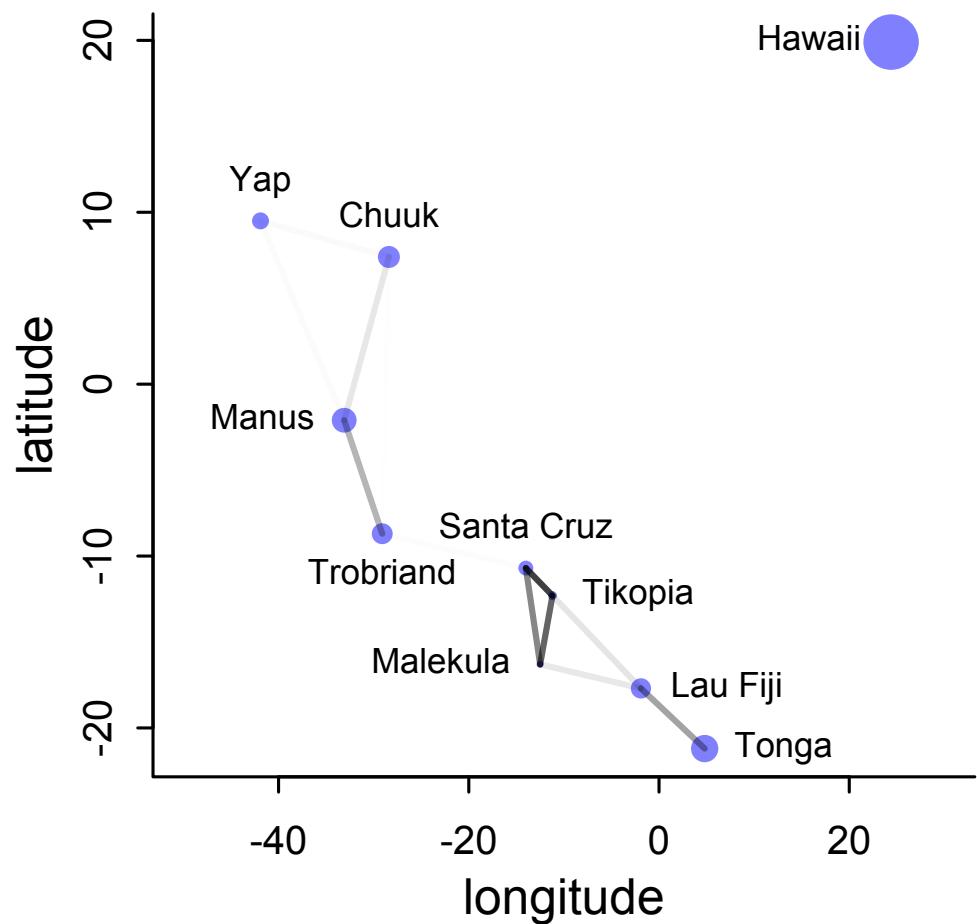
$$K_{ij} = \eta^2 \exp(-\rho^2 D_{ij}^2) + \delta_{ij}(0.01)$$



Implied correlations

- Covariance (and variance) on log scale, so hard to understand
 - Compute correlations at posterior median:





Gaussian process regression

- Many applications, many covariance functions
 - Periodic functions of time (seasonality)
 - Phylogenetic (*patristic*) distance => phylogenetic regression
 - Social networks
 - Non-parametric splines on any predictor
- Can use multiple dimensions in covariance, “automatic relevance determination”

$$K_{ij} = \eta^2 \exp\left(-(\rho_D^2 D_{ij}^2 + \rho_P^2 (\log P_i - \log P_j)^2)\right) + \delta_{ij}\sigma^2$$

Phylogenetic regression

- Phylogenetic relationships like unobserved confounds
- Phylogenetic distance a clue to covariation
- Ways to get phylogenetic information into a GLM
 - Brownian motion model (PGLS)
 - Ornstein–Uhlenbeck (OU) processes
 - Many others
- All use covariance matrix to represent phylogeny
- Each in principle Gaussian process regression

```
library(rethinking)  
data(Primates301)  
data(Primates301_nex)
```

```
library(ape)
```

```
plot( Primates301_nex
```

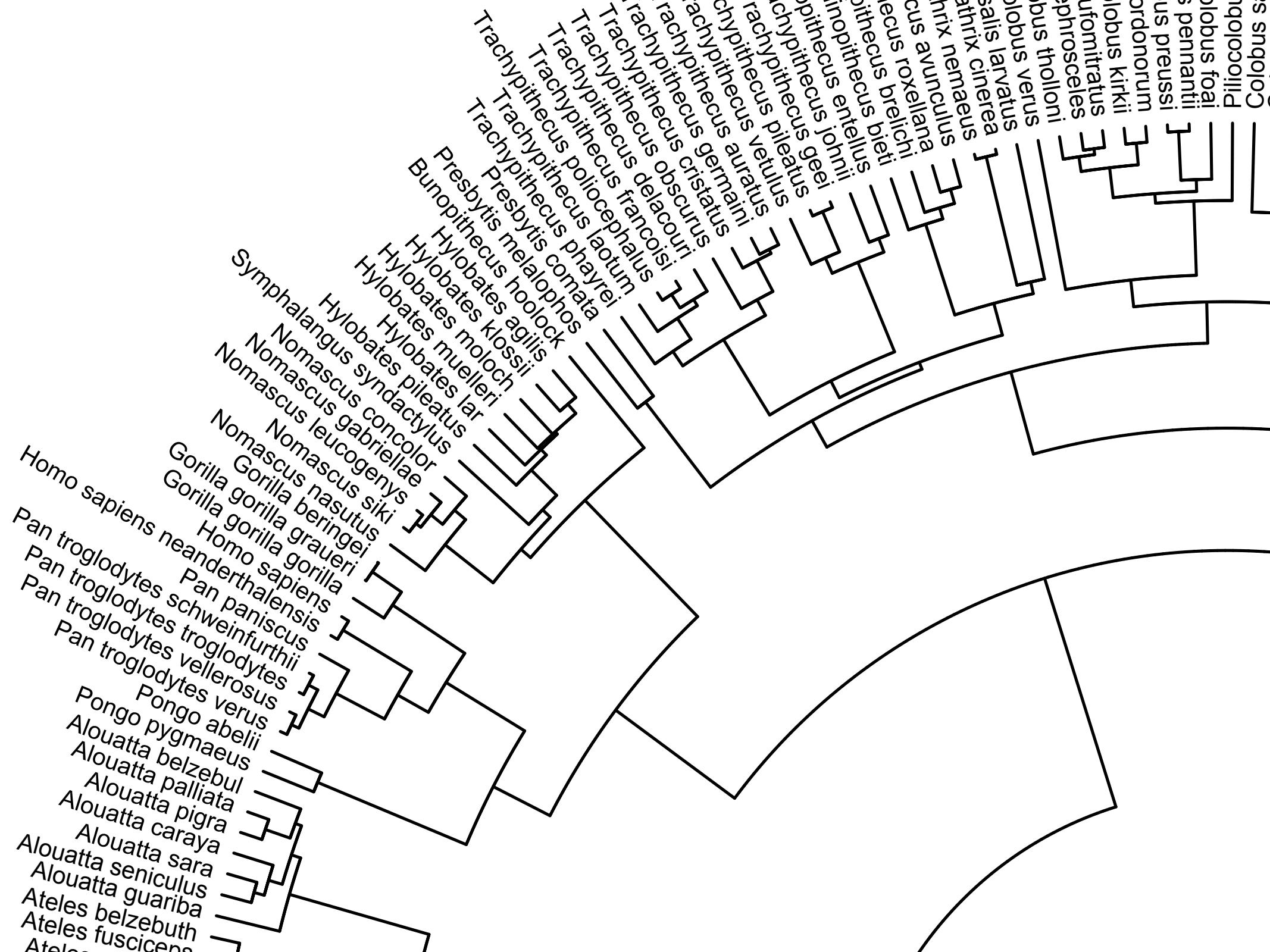
```
no.margin=TRUE , label.offset=1 , cex=0.5 )
```

```

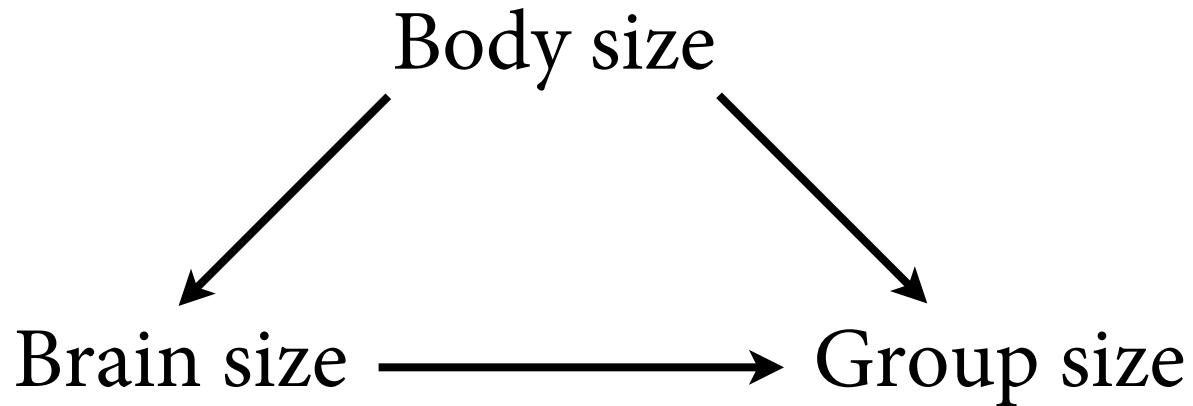
library(rethinking)
data(Primates301)
data(Primates301_nex)
library(ape)
plot( Primates301_nex , type="fan" , font=14,
      no.margin=TRUE , label.offset=1 , cex=0.8 )

```

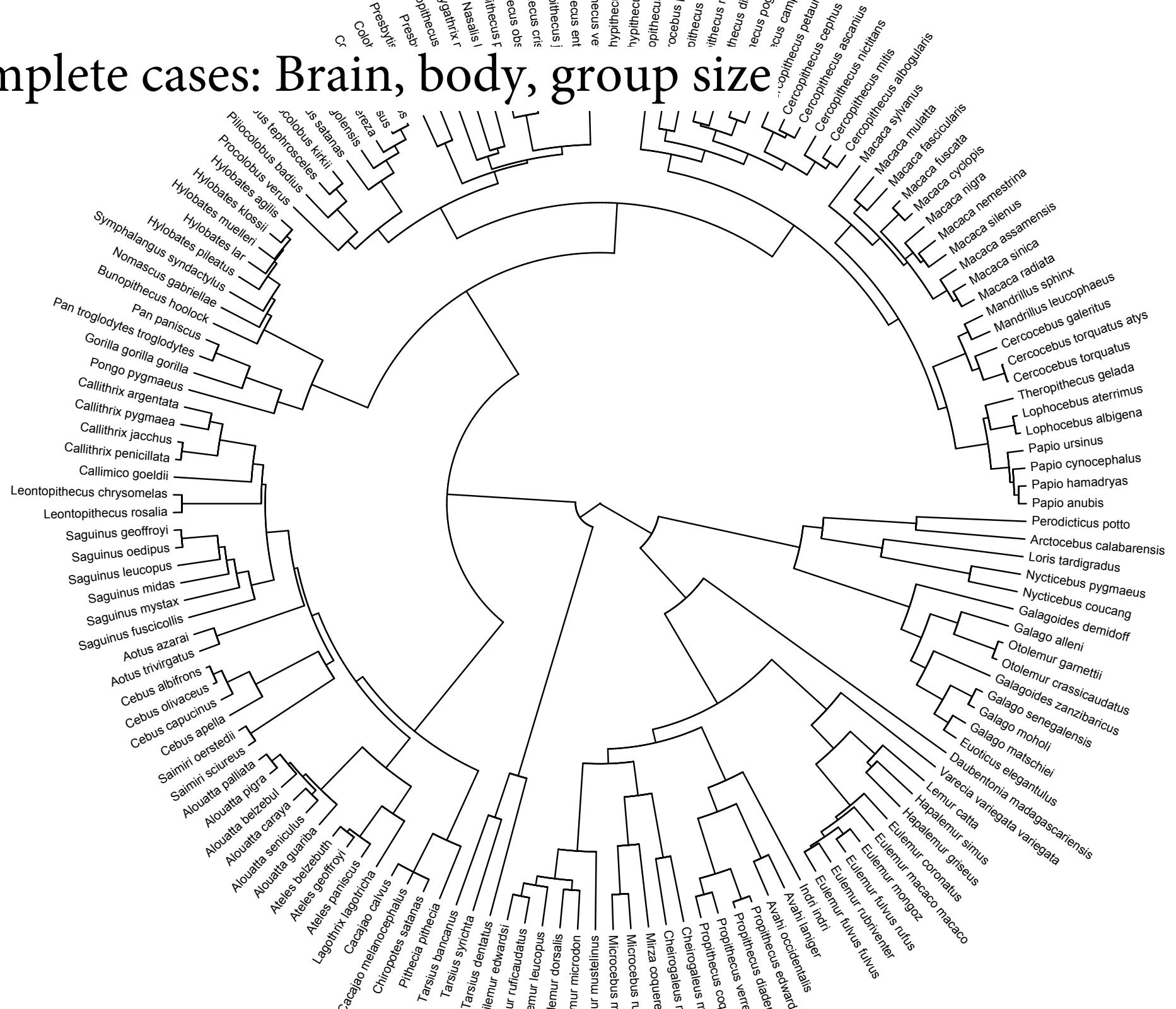
The figure displays a phylogenetic tree of primates, generated using the R package ape. The tree is rooted at the bottom and branches upwards, illustrating the evolutionary relationships among various primate species. The labels for the nodes represent different primate genera and species, such as Homo sapiens, Gorilla gorilla, Pan troglodytes, and numerous species from the families Cercopithecidae, Alouattidae, Callitrichidae, and so on. The tree is presented in a fan-like format, with the central node being the root. The labels are rotated for readability.



Group size



Complete cases: Brain, body, group size



Ordinary regression, weirded

- Any linear regression can be expressed as a multi-variate regression
- Consider outcome as a single vector:

All group sizes

Covariance matrix

$$\mathbf{G} \sim \text{MVNormal}(\boldsymbol{\mu}, \mathbf{S})$$

$$\mu_i = \alpha + \beta_B B_i + \beta_M M_i$$

$$\mathbf{S} = \sigma^2 \mathbf{I}$$

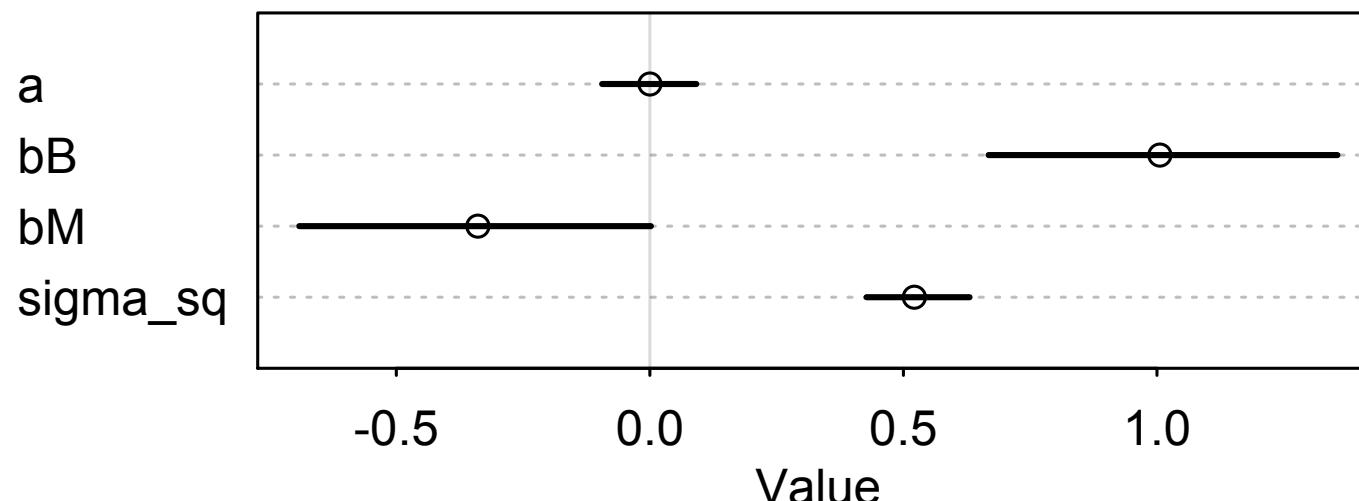
“Identity” matrix



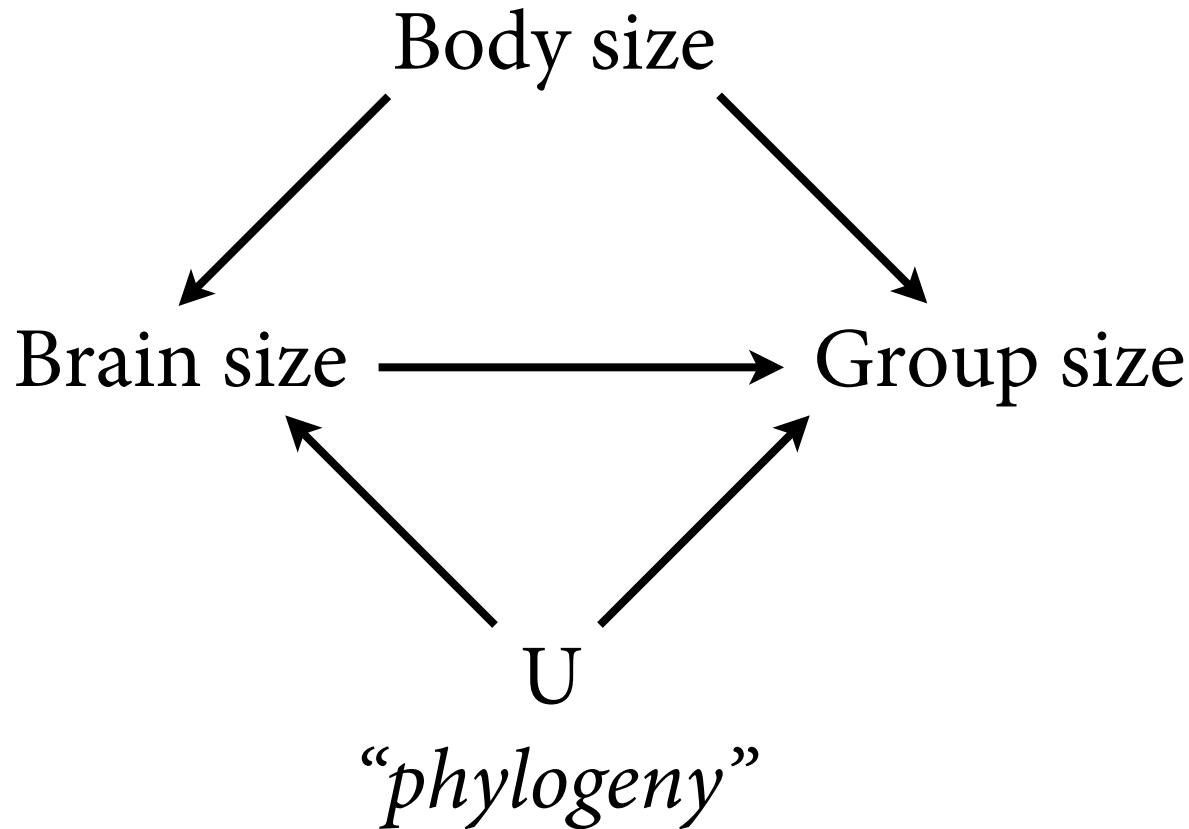
1	0	0	0	0	0	0
0	1	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	1	0	0
0	0	0	0	0	1	0
0	0	0	0	0	0	1

Ordinary regression, weirded

```
m14.8 <- ulam(  
  alist(  
    G ~ multi_normal( mu , SIGMA ) ,  
    mu <- a + bM*M + bB*B ,  
    matrix[N_spp,N_spp]: SIGMA <- Iimat * sigma_sq ,  
    a ~ normal( 0 , 1 ) ,  
    c(bM,bB) ~ normal( 0 , 0.5 ) ,  
    sigma_sq ~ exponential( 1 )  
  ) , data=dat_list , chains=4 , cores=4 )
```

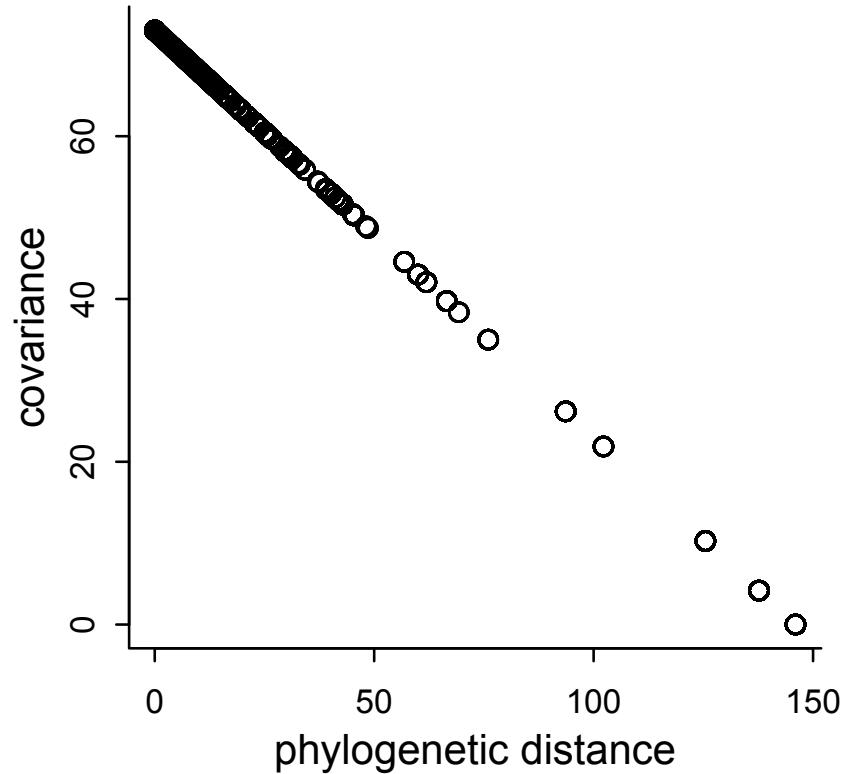


Group size



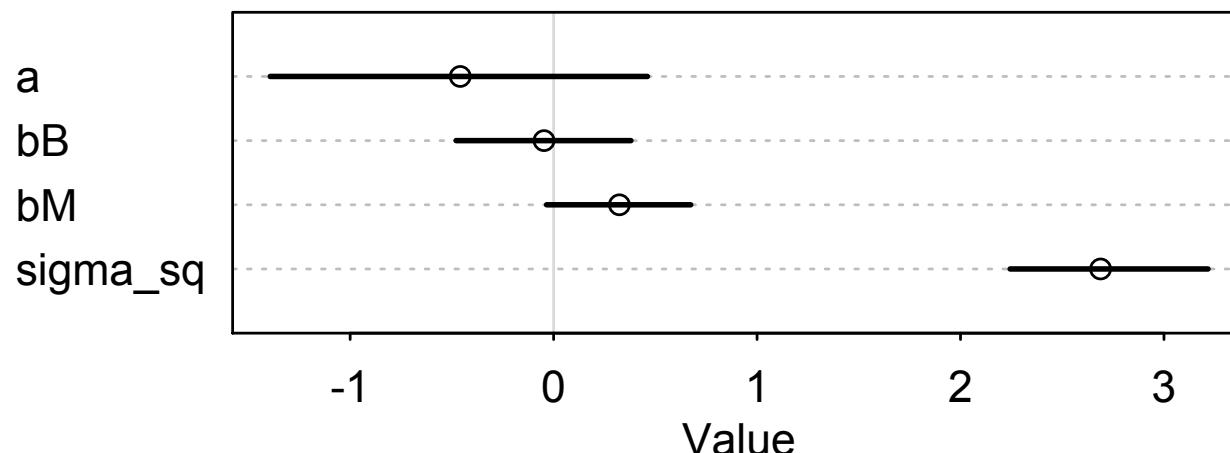
Brownian motion

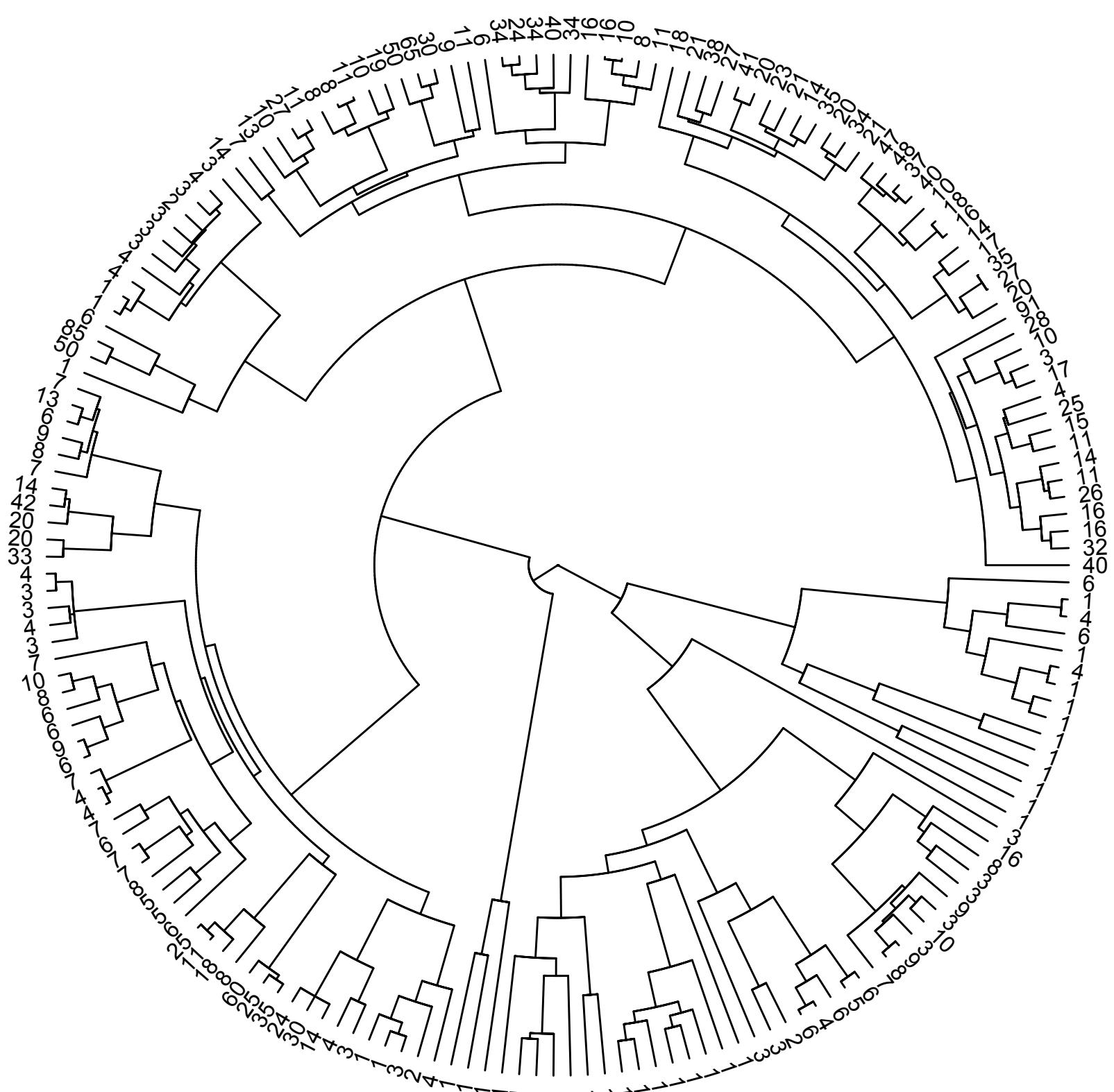
- Oldest & most conservative way to model covariance as function of phylogeny: Brownian motion
- Implies covariance declines linearly with phylogenetic distance
- Really no one is satisfied with this, but common



- Just replace identity matrix with matrix of correlations implied by linear function

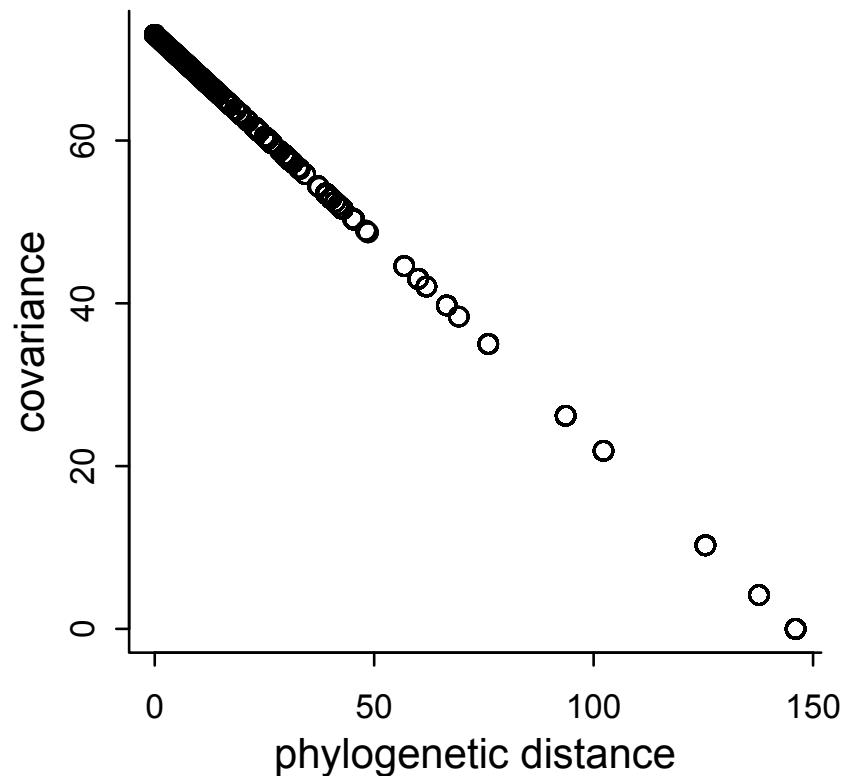
```
m14.9 <- ulam(
  alist(
    G ~ multi_normal( mu , SIGMA ) ,
    mu <- a + bM*M + bB*B ,
    matrix[N_spp,N_spp]: SIGMA <- R * sigma_sq ,
    a ~ normal( 0 , 1 ) ,
    c(bM,bB) ~ normal( 0 , 0.5 ) ,
    sigma_sq ~ exponential( 1 )
  ), data=dat_list , chains=4 , cores=4 )
```





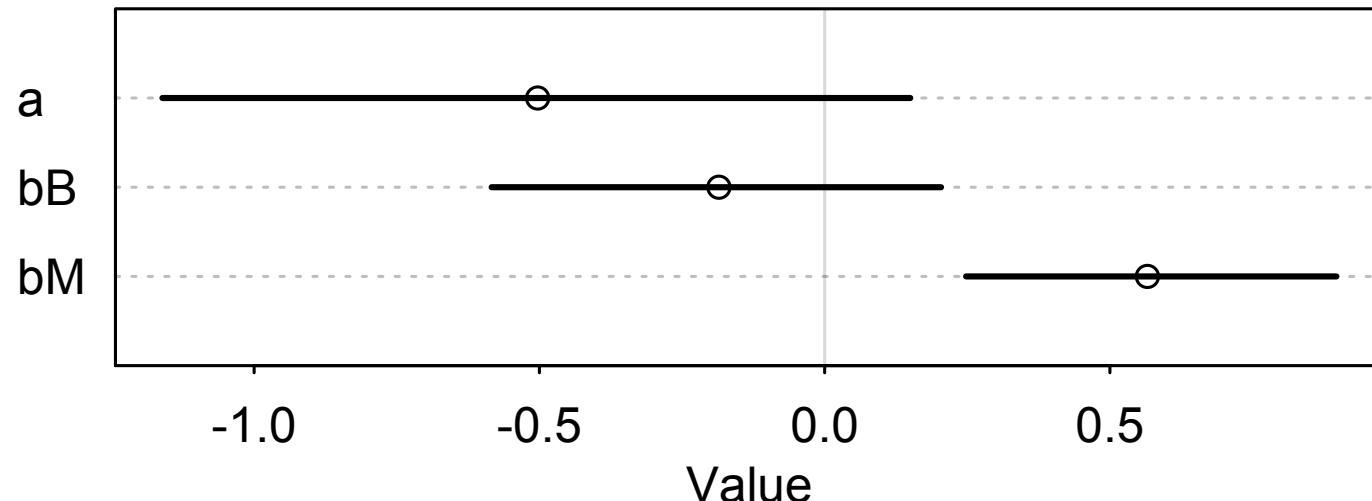
Gaussian process

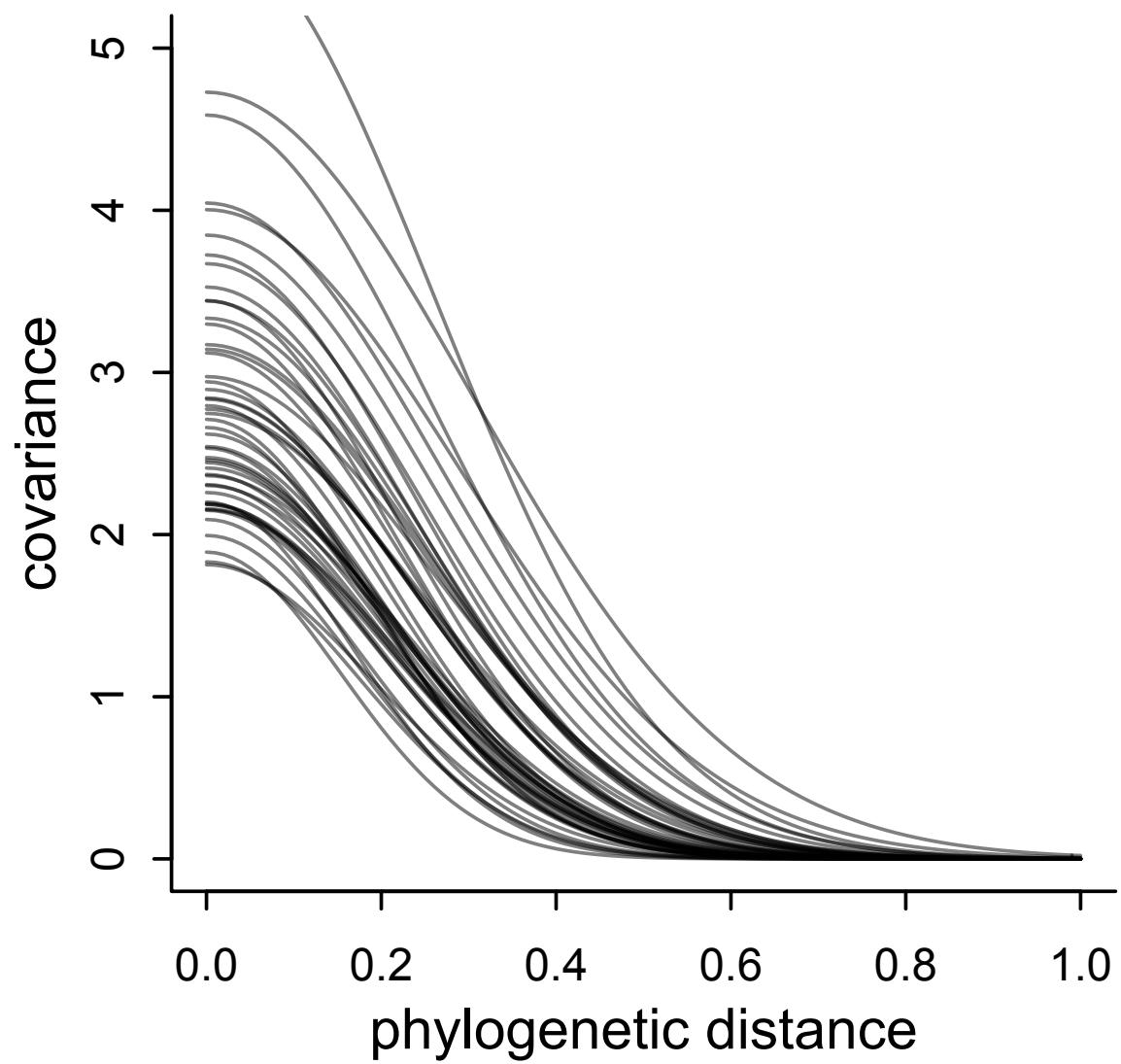
- No reason to assume linear decline with distance
- Many possible covariance functions
- Gaussian process considers an infinite number of specific form



```
# add scaled and reordered distance matrix
dat_list$Dmat <- Dmat[ spp_obs , spp_obs ] / max(Dmat)

m14.10 <- ulam(
  alist(
    G ~ multi_normal( mu , SIGMA ),
    mu <- a + bM*M + bB*B,
    matrix[N_spp,N_spp]: SIGMA <- cov_GPL2( Dmat , etasq , rhosq , 0.01 ),
    a ~ normal(0,1),
    c(bM,bB) ~ normal(0,0.5),
    etasq ~ exponential(1),
    rhosq ~ exponential(1)
  ), data=dat_list , chains=4 , cores=4 )
precis( m14.10 )
```





Phylogenetic regression

- Many possible covariance functions
 - Variable rates on branches
 - Different trees for different traits (*hemiplasy*)
 - Many equilibria
- No unique null model — p -values weird
- Causation: Organisms are silly machines (joint causation over time)