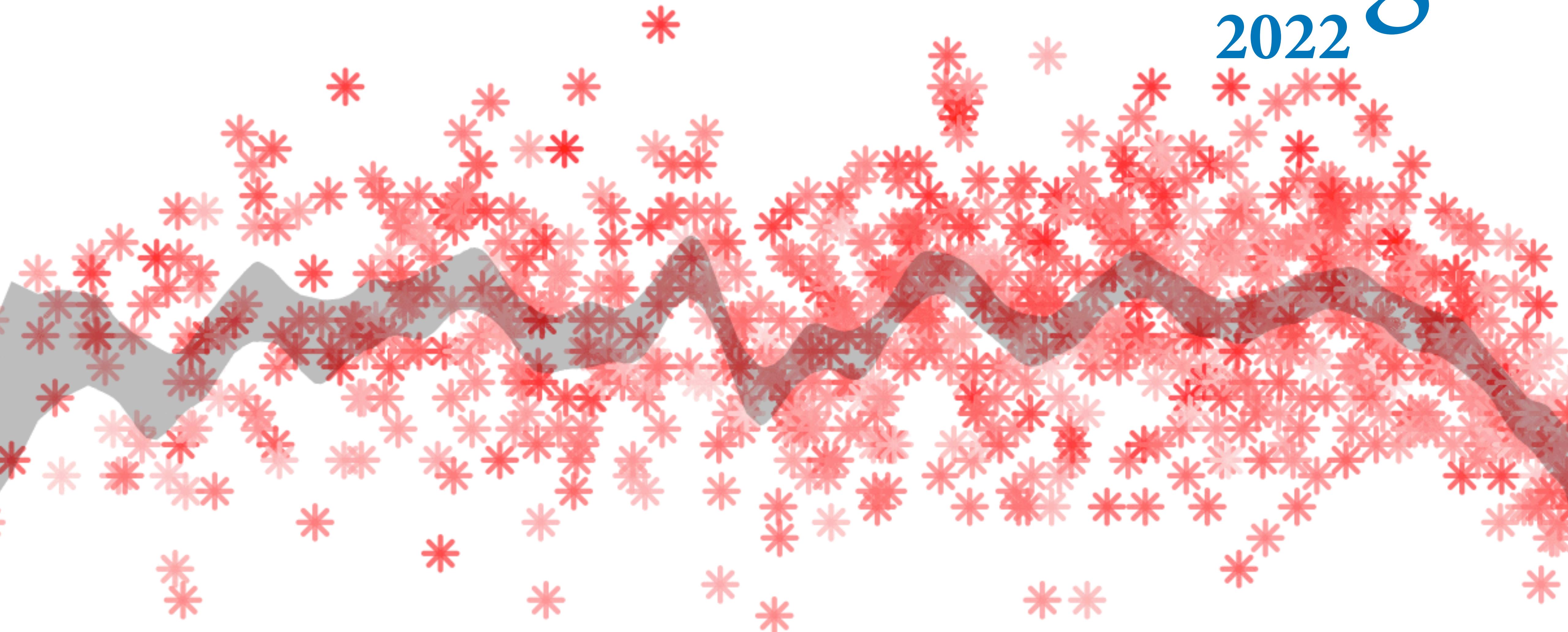


# Statistical Rethinking

2022



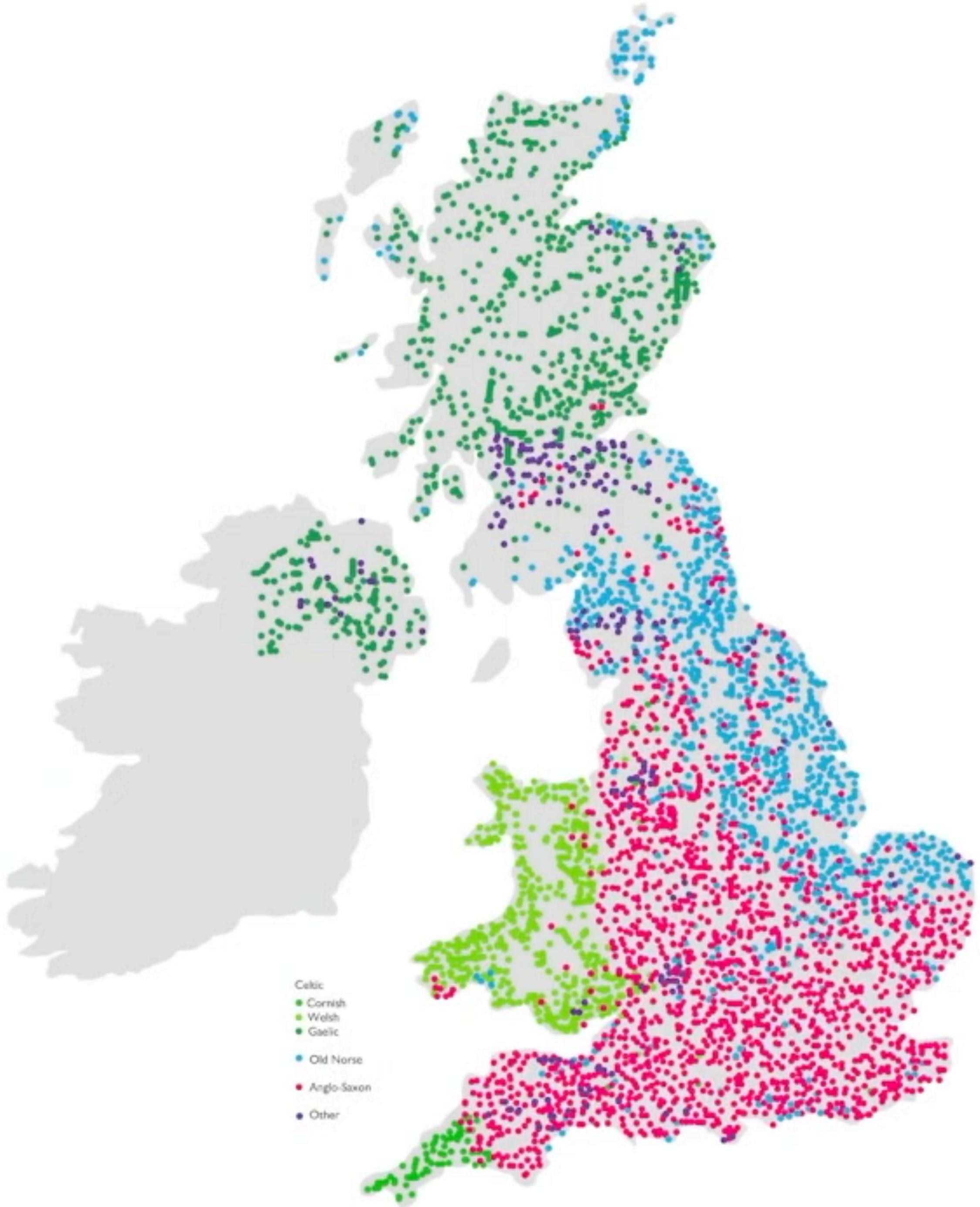
16: Space, Time and Phylogeny

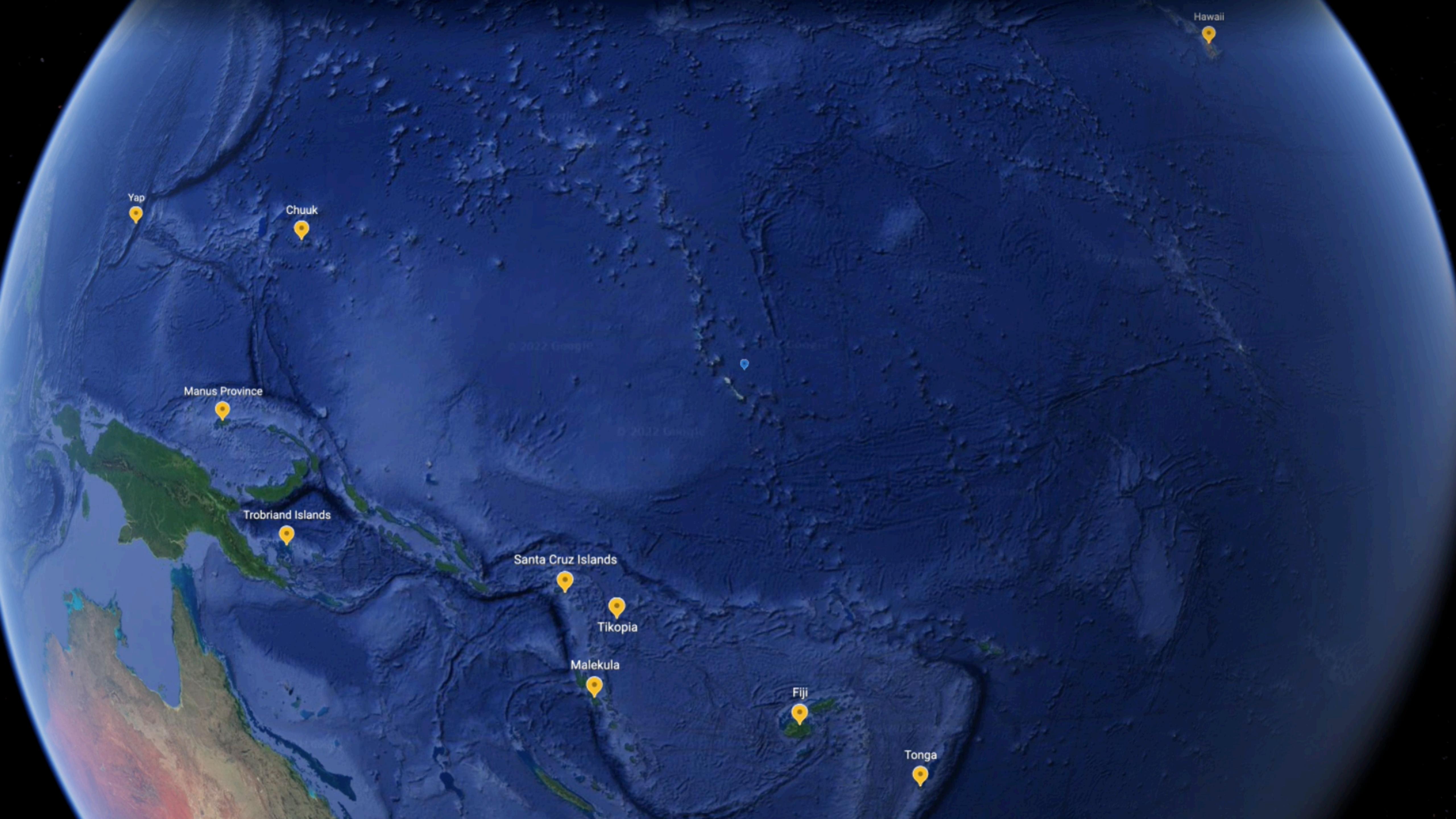
though tough trough

through thought

thorough bough hough

hiccough lough





Hawaii



Yap



Chuuk



Manus Province



Trobriand Islands



Santa Cruz Islands



Tikopia



Malekula



Fiji



Tonga

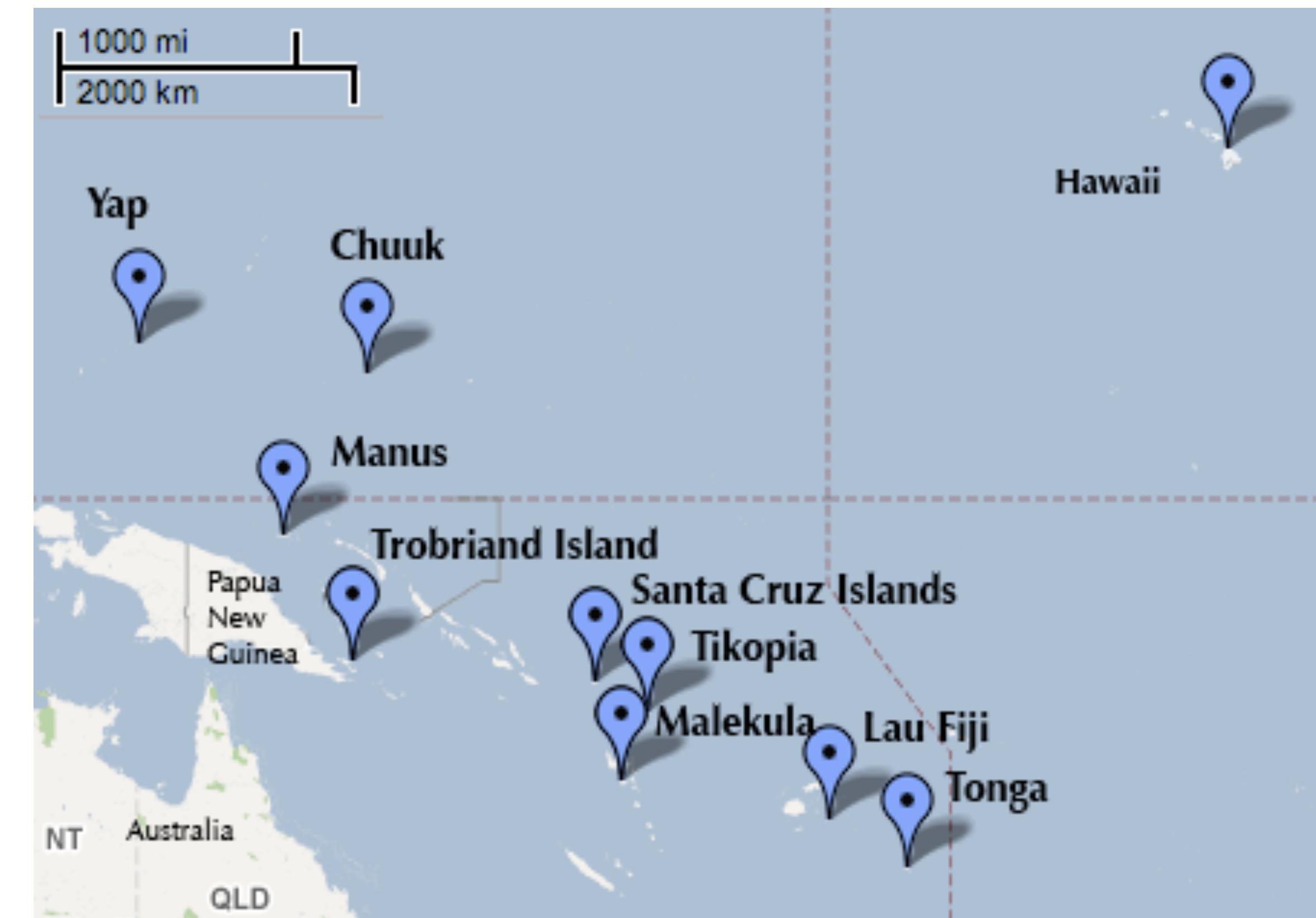


# Oceanic Technology

data(Kline2)

Number of **tool types** associated with **population size**

Spatial covariation: Islands close together share **unobserved confounds** and innovations

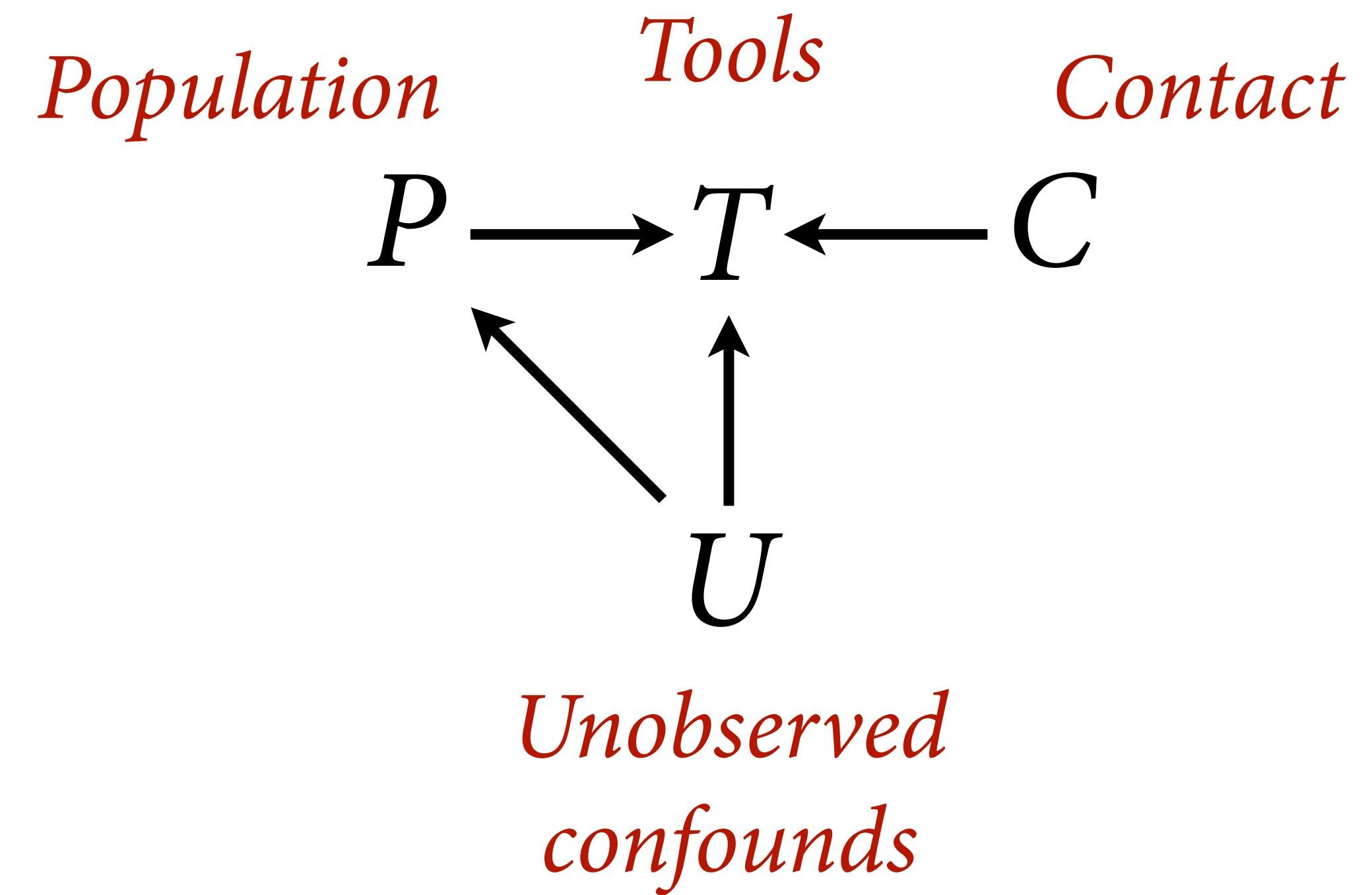


# Oceanic Technology

`data(Kline2)`

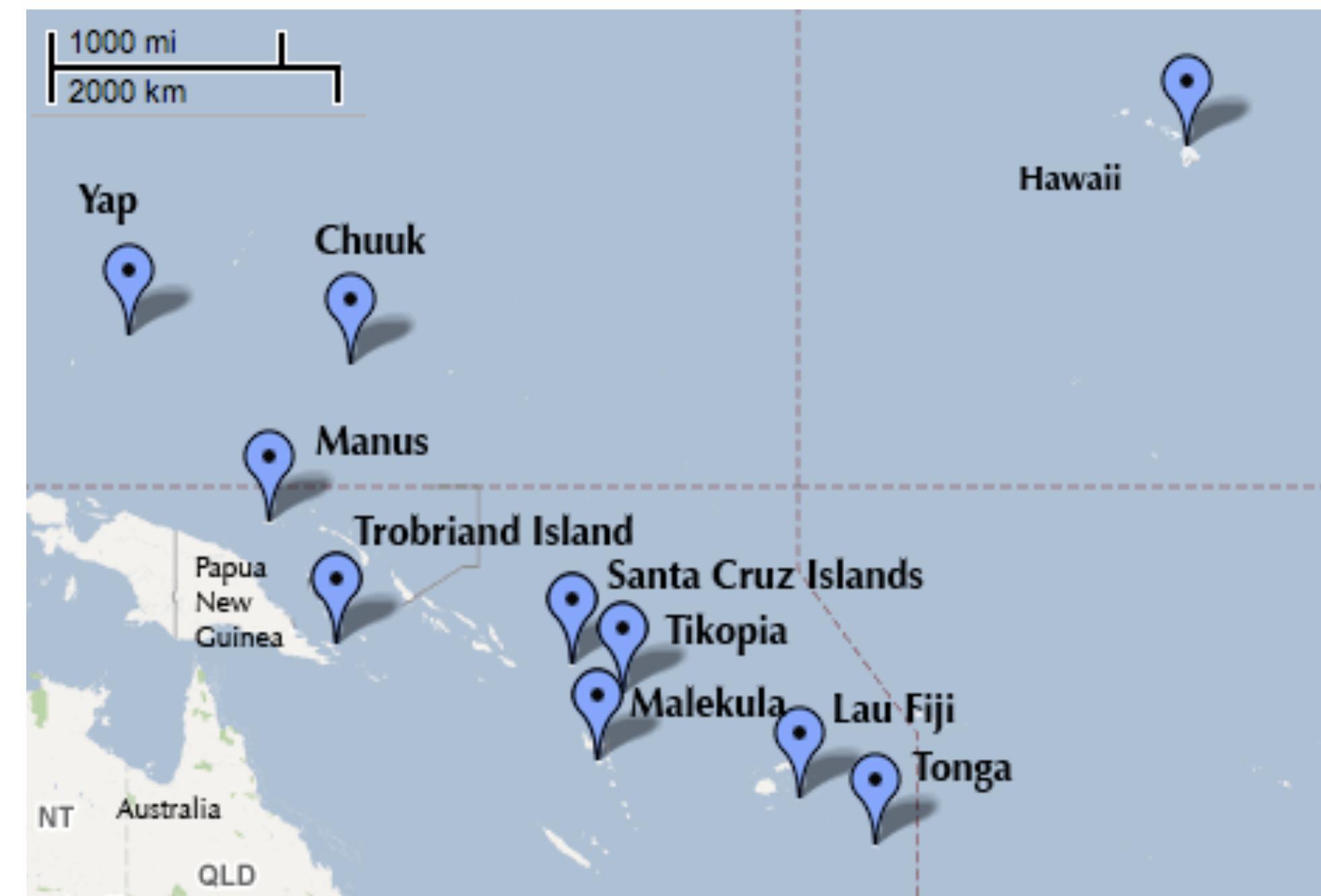
Number of **tool types** associated  
with **population** size

Spatial covariation: Islands close  
together share **unobserved**  
**confounds** and innovations



$$\Delta T = \alpha P^\beta - \gamma T$$

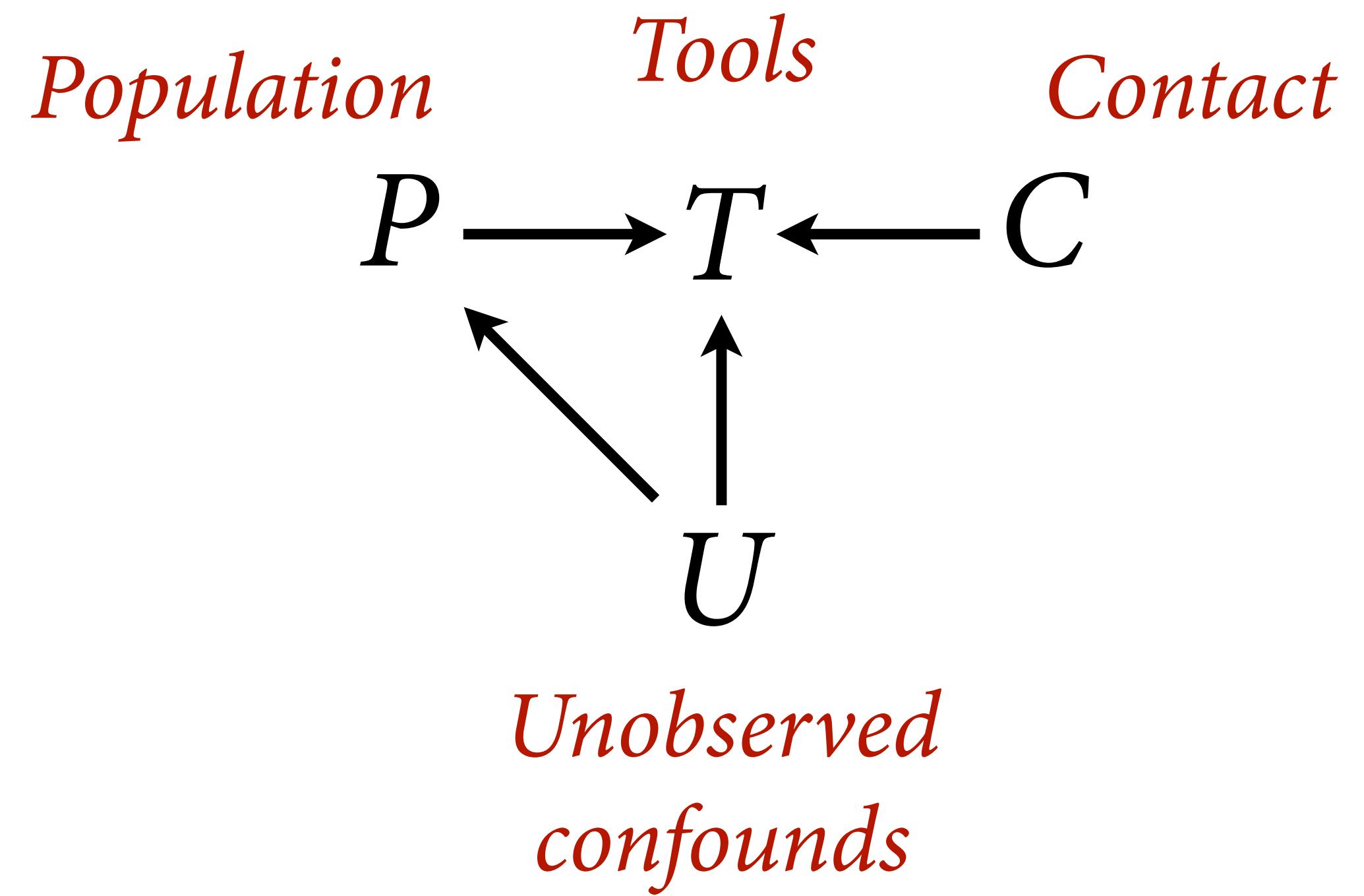
*change in tools*  
*innovation rate*  
*rate of loss*  
*diminishing returns (elasticity)*



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

$$\lambda_i = \hat{T}$$

$$\hat{T} = \frac{\alpha P^\beta}{\gamma}$$



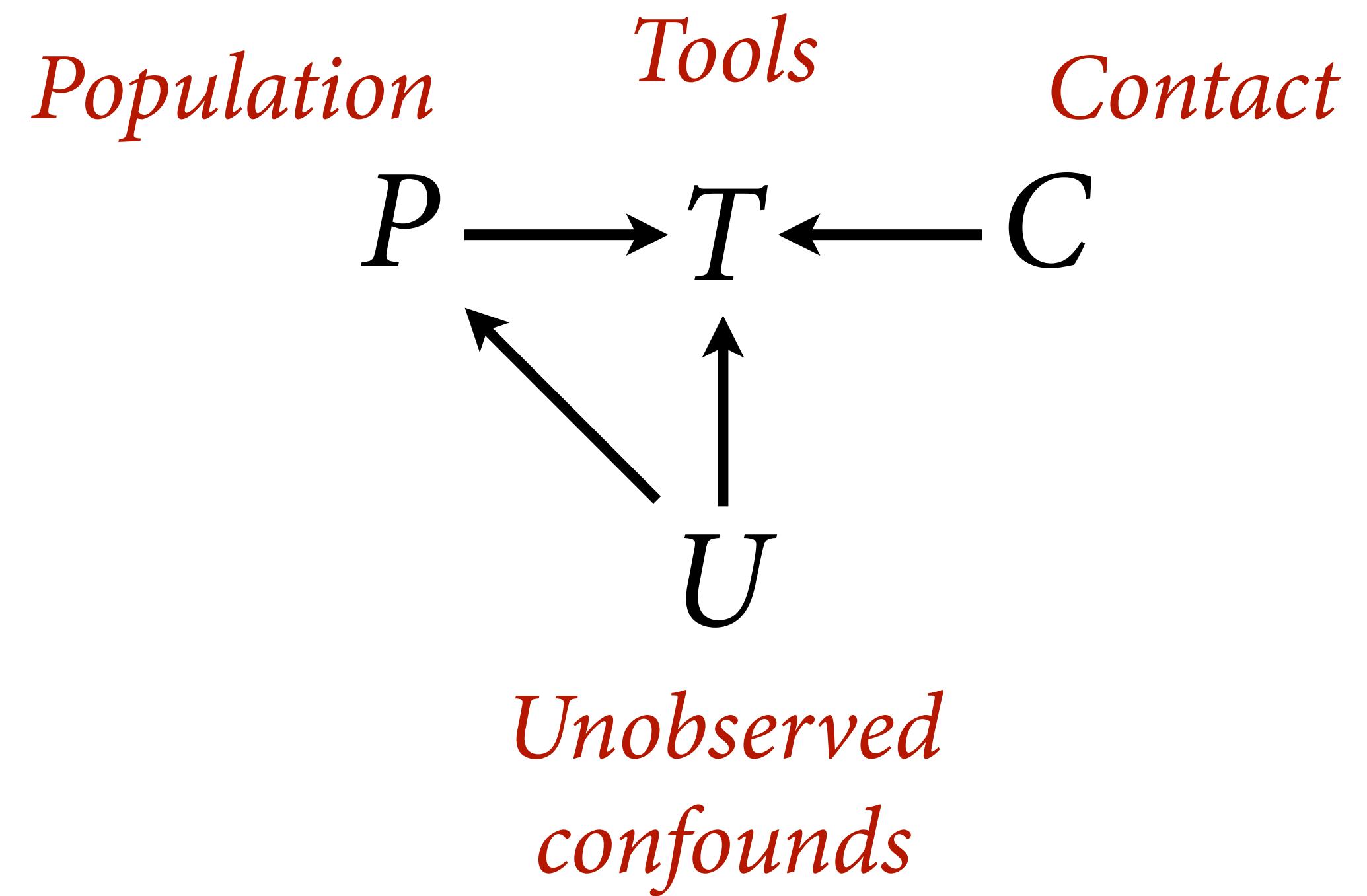
$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = \hat{T}$$

$$\hat{T} = \frac{\alpha P^\beta}{\gamma}$$

Spatial covariation: Islands close together share **unobserved confounds** and **innovations**

Effect of  $U$  is to make closer islands have more similar  $\hat{T}$



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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

*deviation log-tools  
in society  $i$*

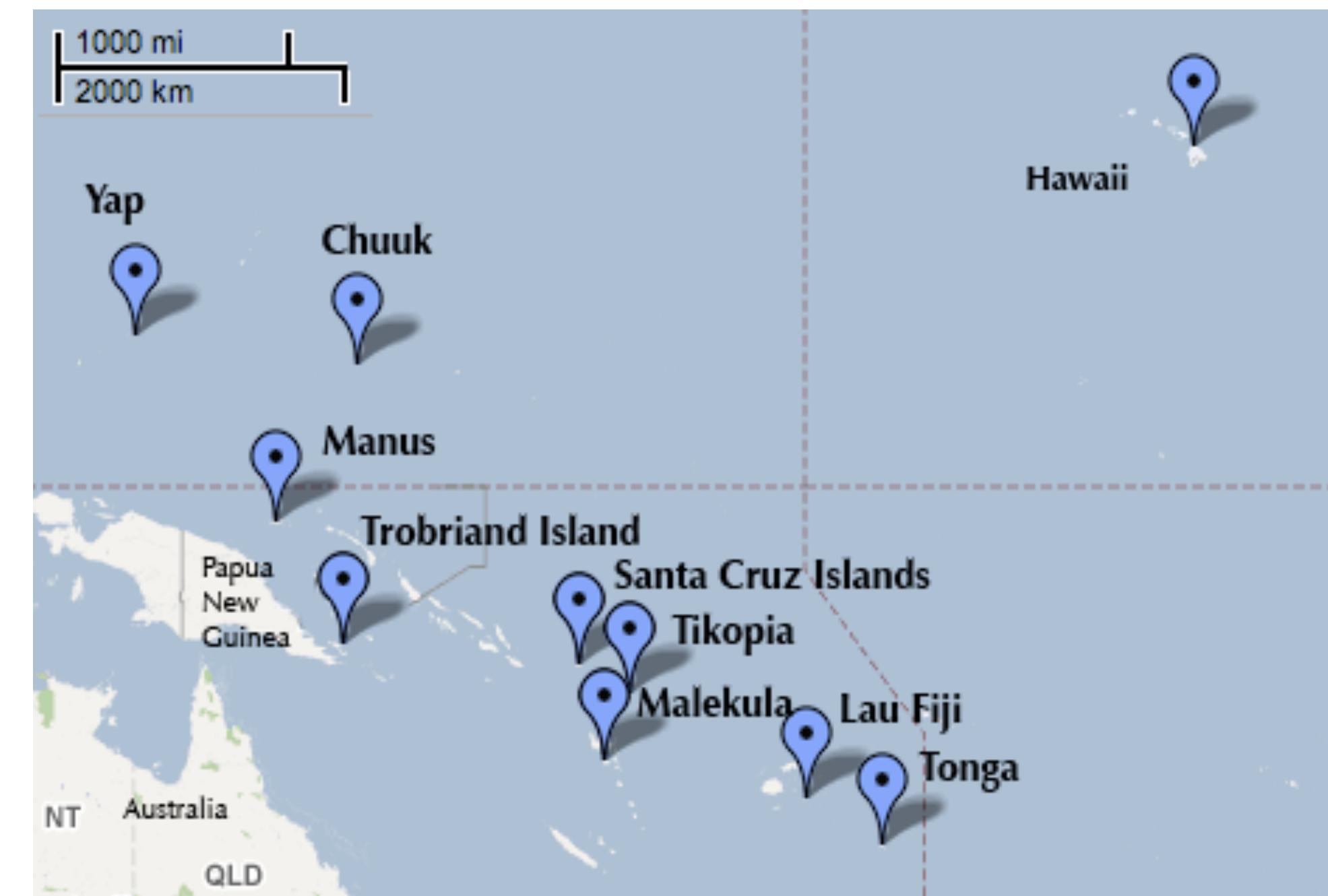


$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix}$$

*vector of all  
varying effects*



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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

*vector of all  
varying effects*

*vector of zeros*



*covariance matrix,  
the “Kernel”*

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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

$$\mathbf{K} = \begin{bmatrix} \sigma^2 & k_{1,2} & k_{1,3} & k_{1,4} & k_{1,5} & k_{1,6} & k_{1,7} & k_{1,8} & k_{1,9} & k_{1,10} \\ \sigma^2 & k_{2,3} & k_{2,4} & k_{2,5} & k_{2,6} & k_{2,7} & k_{2,8} & k_{2,9} & k_{2,10} & \\ \sigma^2 & k_{3,4} & k_{3,5} & k_{3,6} & k_{3,7} & k_{3,8} & k_{3,9} & k_{3,10} & & \\ \sigma^2 & k_{4,5} & k_{4,6} & k_{4,7} & k_{4,8} & k_{4,9} & k_{4,10} & & & \\ \sigma^2 & k_{5,6} & k_{5,7} & k_{5,8} & k_{5,9} & k_{5,10} & & & & \\ \sigma^2 & k_{6,7} & k_{6,8} & k_{6,9} & k_{6,10} & & & & & \\ \sigma^2 & k_{7,8} & k_{7,9} & k_{7,10} & & & & & & \\ \sigma^2 & k_{8,9} & k_{8,10} & & & & & & & \\ \sigma^2 & k_{9,10} & & & & & & & & \\ \sigma^2 & & & & & & & & & \end{bmatrix}$$

$$\mathbf{K} = \begin{bmatrix} & \text{Malekula} & & & & & & & & \\ & \text{Tikopia} & \text{Santa Cruz} & & & & & & & \\ & & & \text{Yap} & & \text{Fiji} & & & & \\ & & & & \text{Trobriand} & & & & & \\ & & & & & \text{Chuuk} & & & & \\ & & & & & & \text{Manus} & & & \\ & & & & & & & \text{Tonga} & & \\ & & & & & & & & \text{Hawaii} & \\ \sigma^2 & k_{1,2} & k_{1,3} & k_{1,4} & k_{1,5} & k_{1,6} & k_{1,7} & k_{1,8} & k_{1,9} & k_{1,10} \\ \sigma^2 & & k_{2,3} & k_{2,4} & k_{2,5} & k_{2,6} & k_{2,7} & k_{2,8} & k_{2,9} & k_{2,10} \\ \sigma^2 & & & k_{3,4} & k_{3,5} & k_{3,6} & k_{3,7} & k_{3,8} & k_{3,9} & k_{3,10} \\ \sigma^2 & & & & k_{4,5} & k_{4,6} & k_{4,7} & k_{4,8} & k_{4,9} & k_{4,10} \\ \sigma^2 & & & & & k_{5,6} & k_{5,7} & k_{5,8} & k_{5,9} & k_{5,10} \\ \sigma^2 & & & & & & k_{6,7} & k_{6,8} & k_{6,9} & k_{6,10} \\ \sigma^2 & & & & & & & k_{7,8} & k_{7,9} & k_{7,10} \\ \sigma^2 & & & & & & & & k_{8,9} & k_{8,10} \\ \sigma^2 & & & & & & & & & k_{9,10} \\ \sigma^2 & & & & & & & & & \end{bmatrix} \begin{array}{l} \text{Malekula} \\ \text{Tikopia} \\ \text{Santa Cruz} \\ \text{Yap} \\ \text{Fiji} \\ \text{Trobriand} \\ \text{Chuuk} \\ \text{Manus} \\ \text{Tonga} \\ \text{Hawaii} \end{array}$$

45 covariances

# Gaussian Processes

A Gaussian Process is “*an infinite-dimensional generalization of multivariate normal distributions*”

What does this mean?

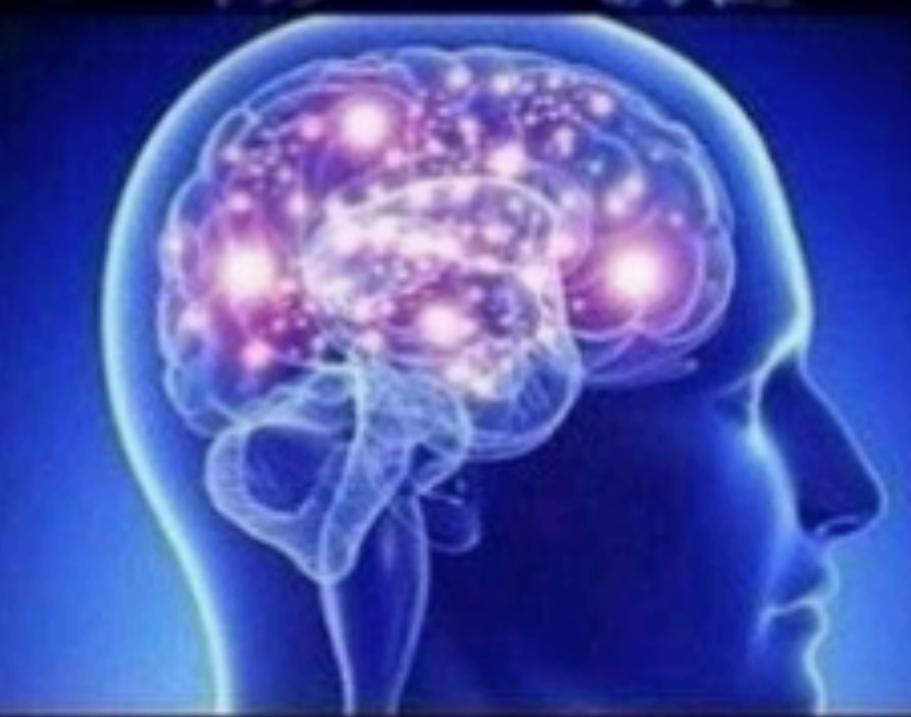
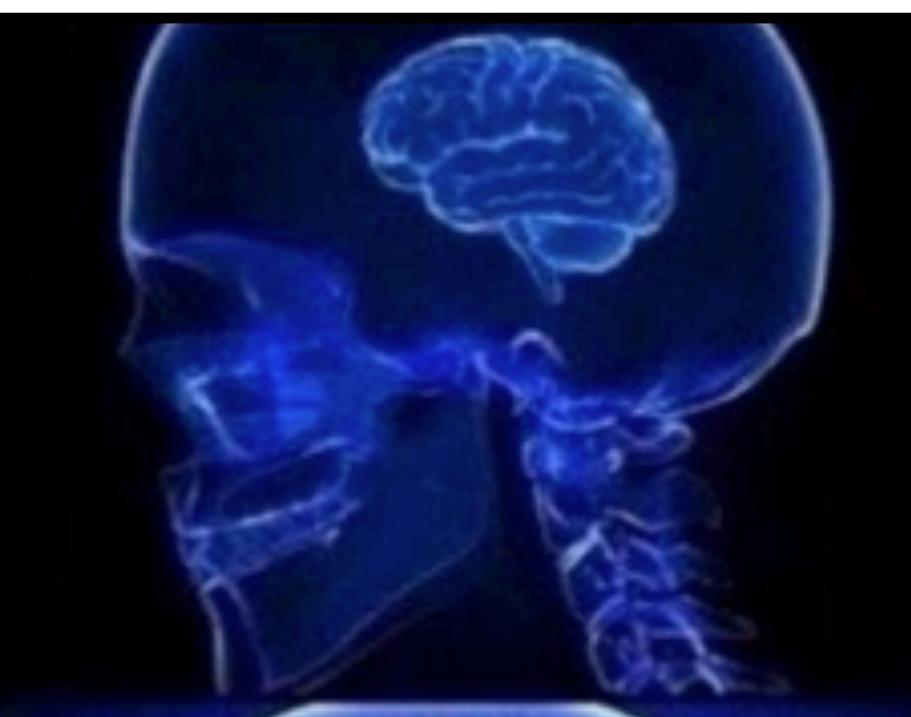
Instead of conventional covariance matrix, use a kernel function that generalizes to infinite dimensions/observations/predictions

**A NUMBER**

**A NORMAL  
DISTRIBUTION**

**A MULTIVARIATE  
NORMAL  
DISTRIBUTION**

**AN INFINITE  
DIMENSIONAL  
MULTIVARIATE  
NORMAL  
DISTRIBUTION**



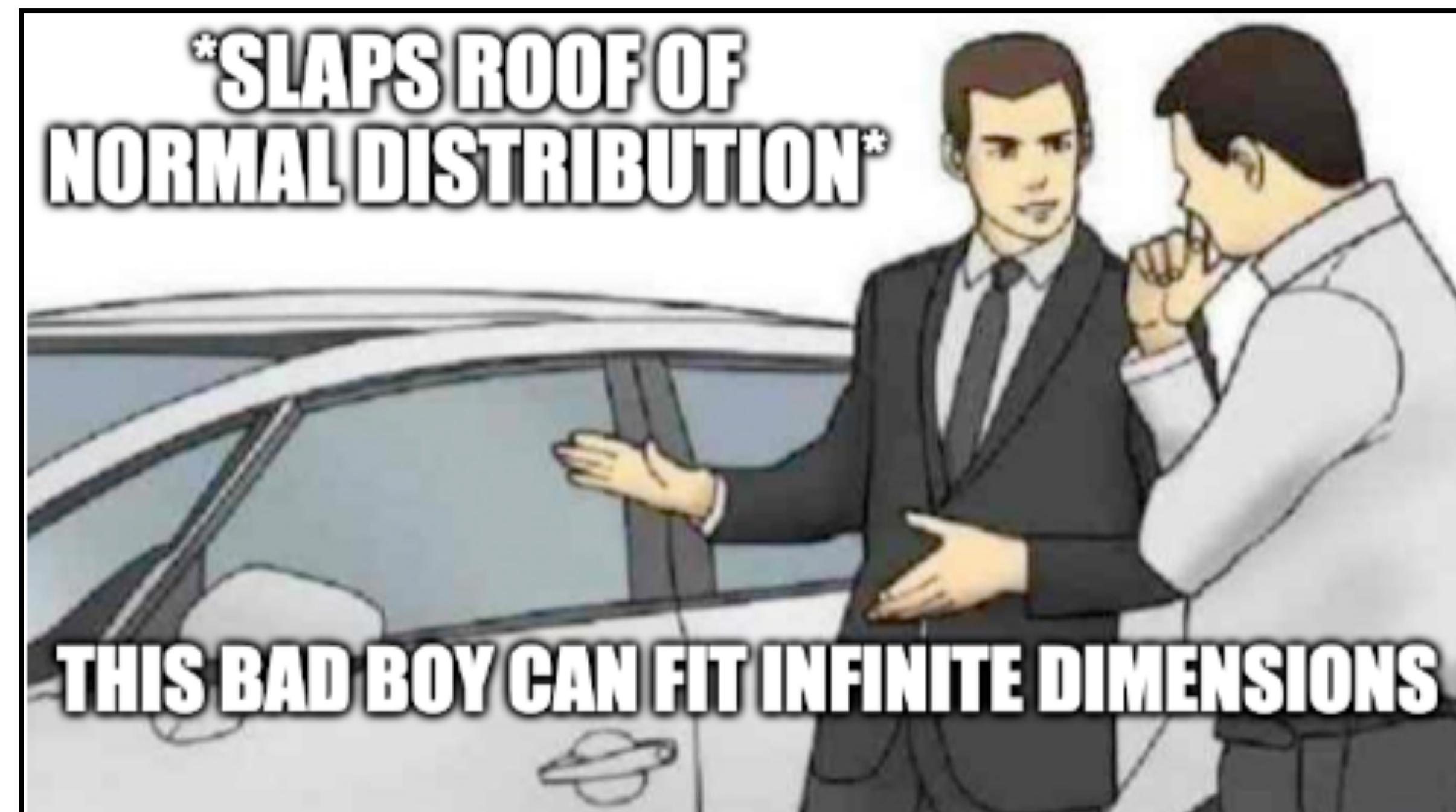
# Gaussian Processes

Instead of conventional covariance matrix, use a **kernel function**

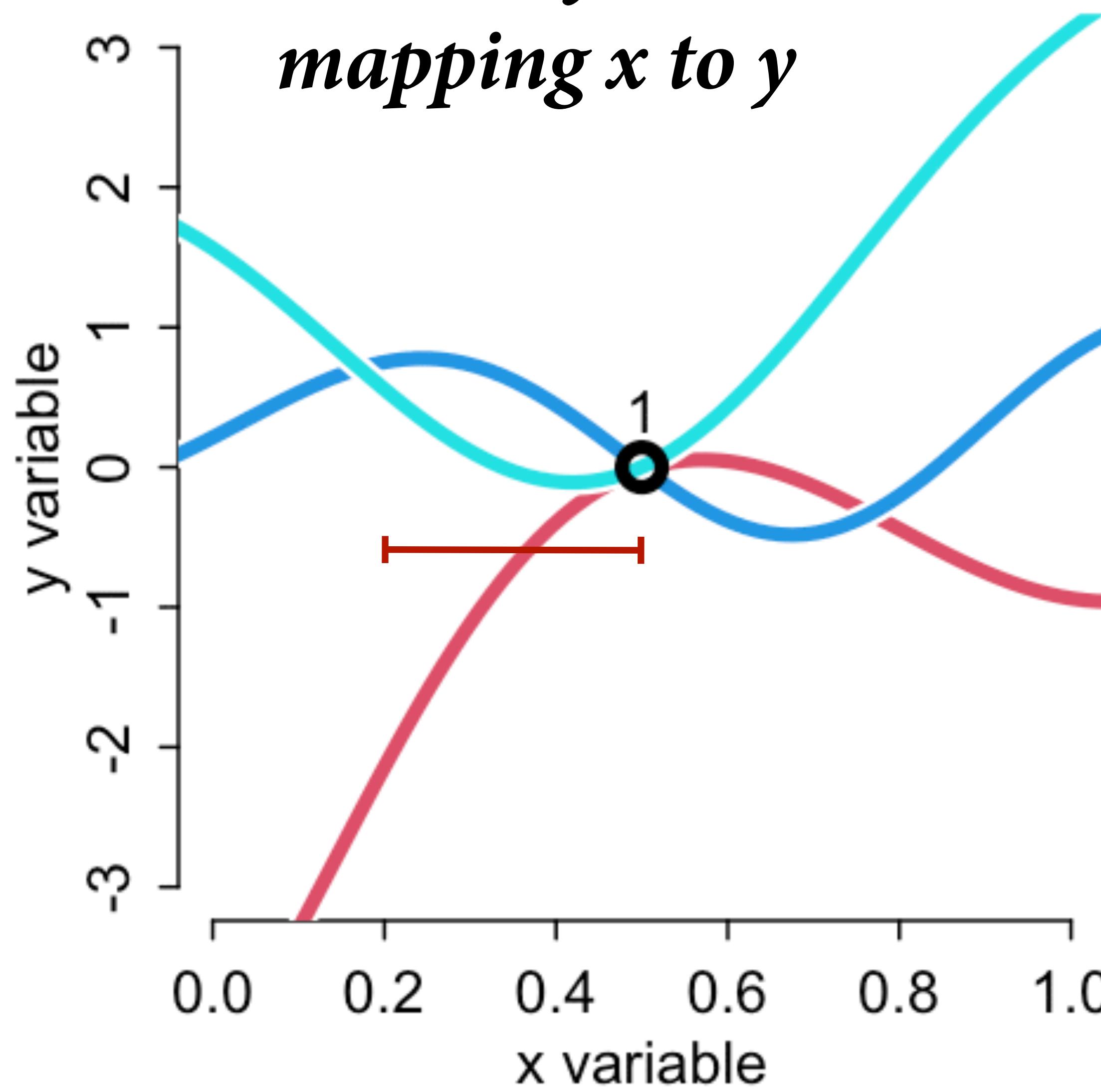
The kernel gives the covariance between any pair of points as a function of their **distance**

Distance can be difference, space, time, etc

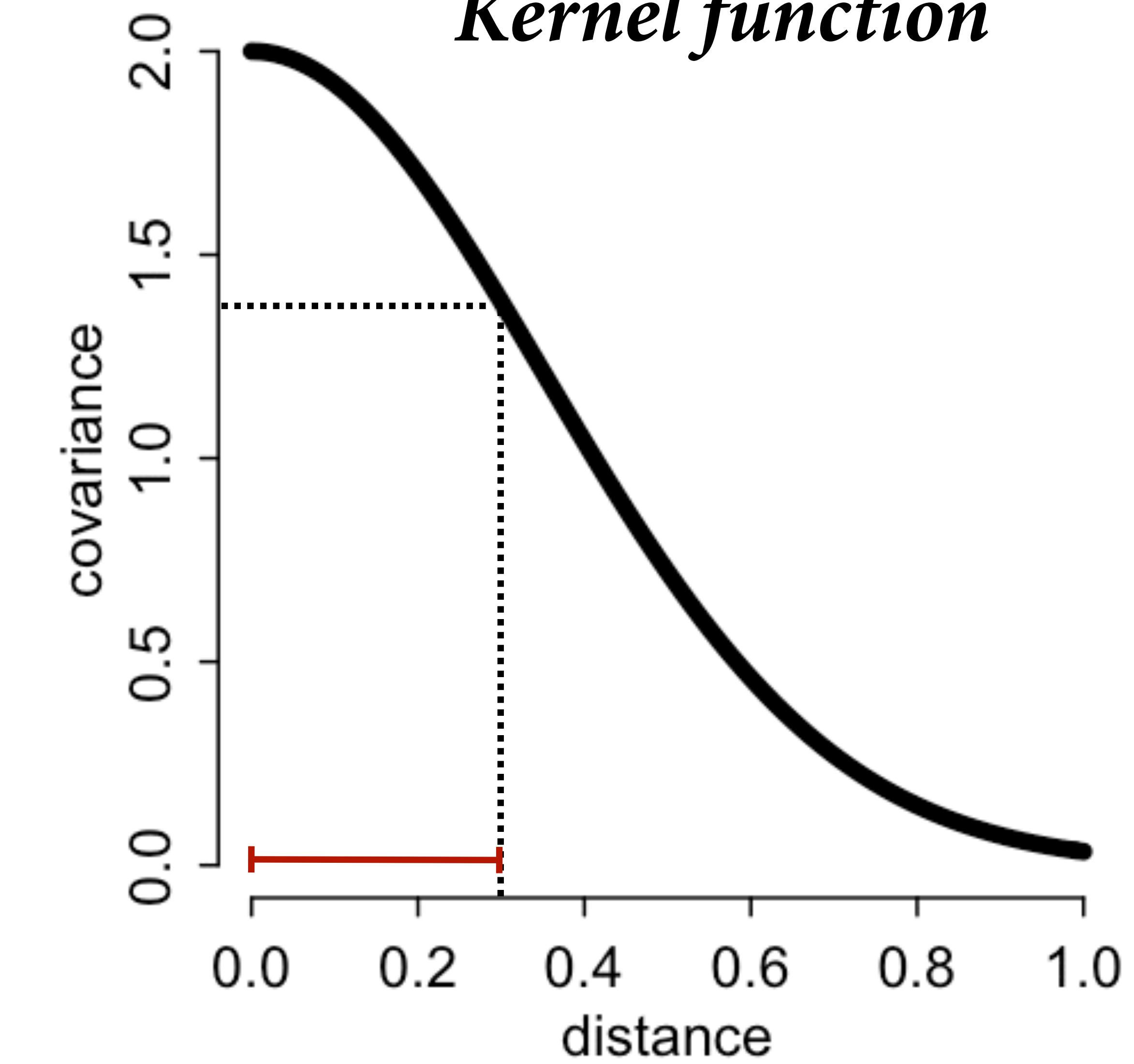
Continuous, ordered categories



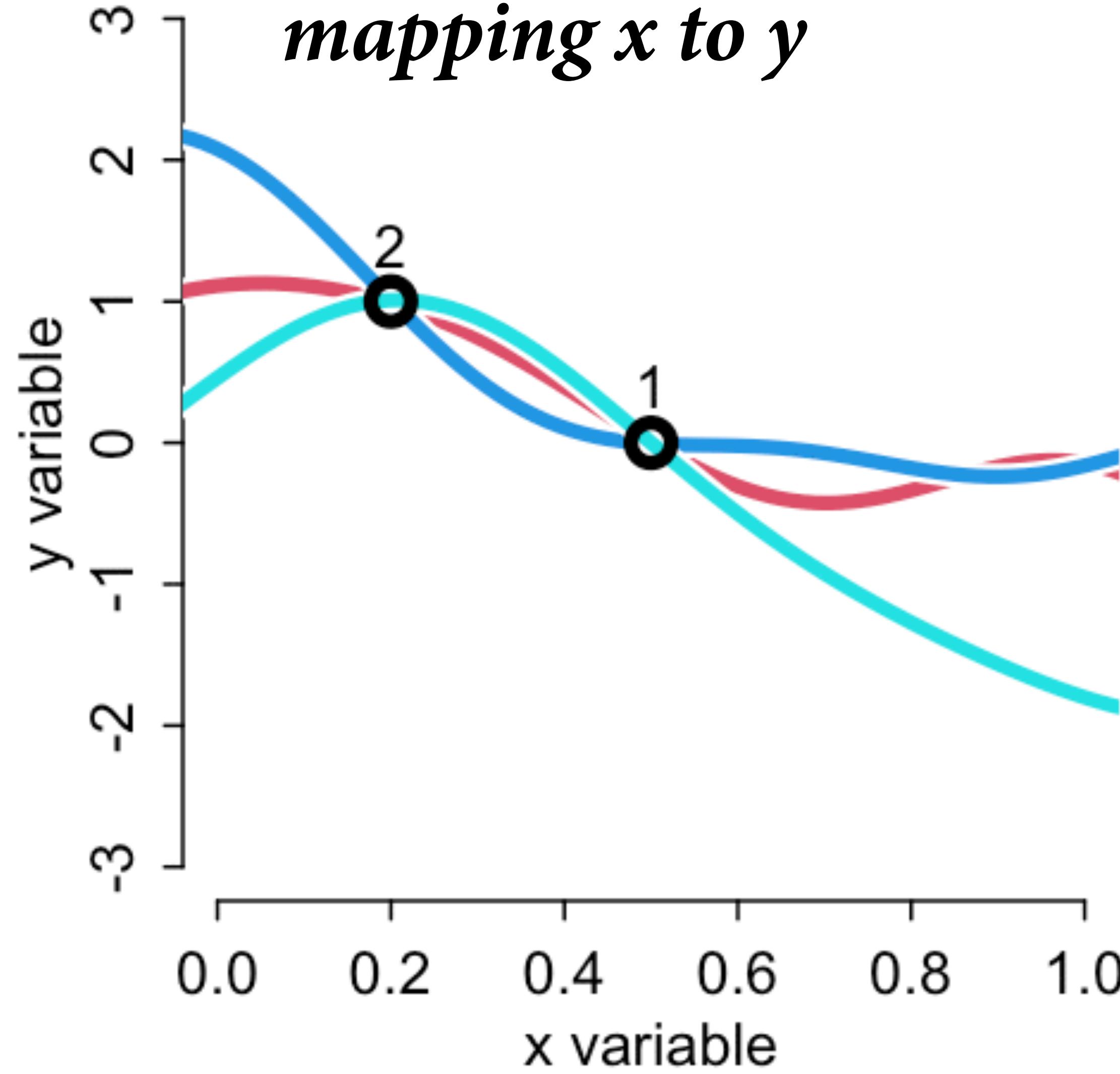
*Possible functions  
mapping  $x$  to  $y$*



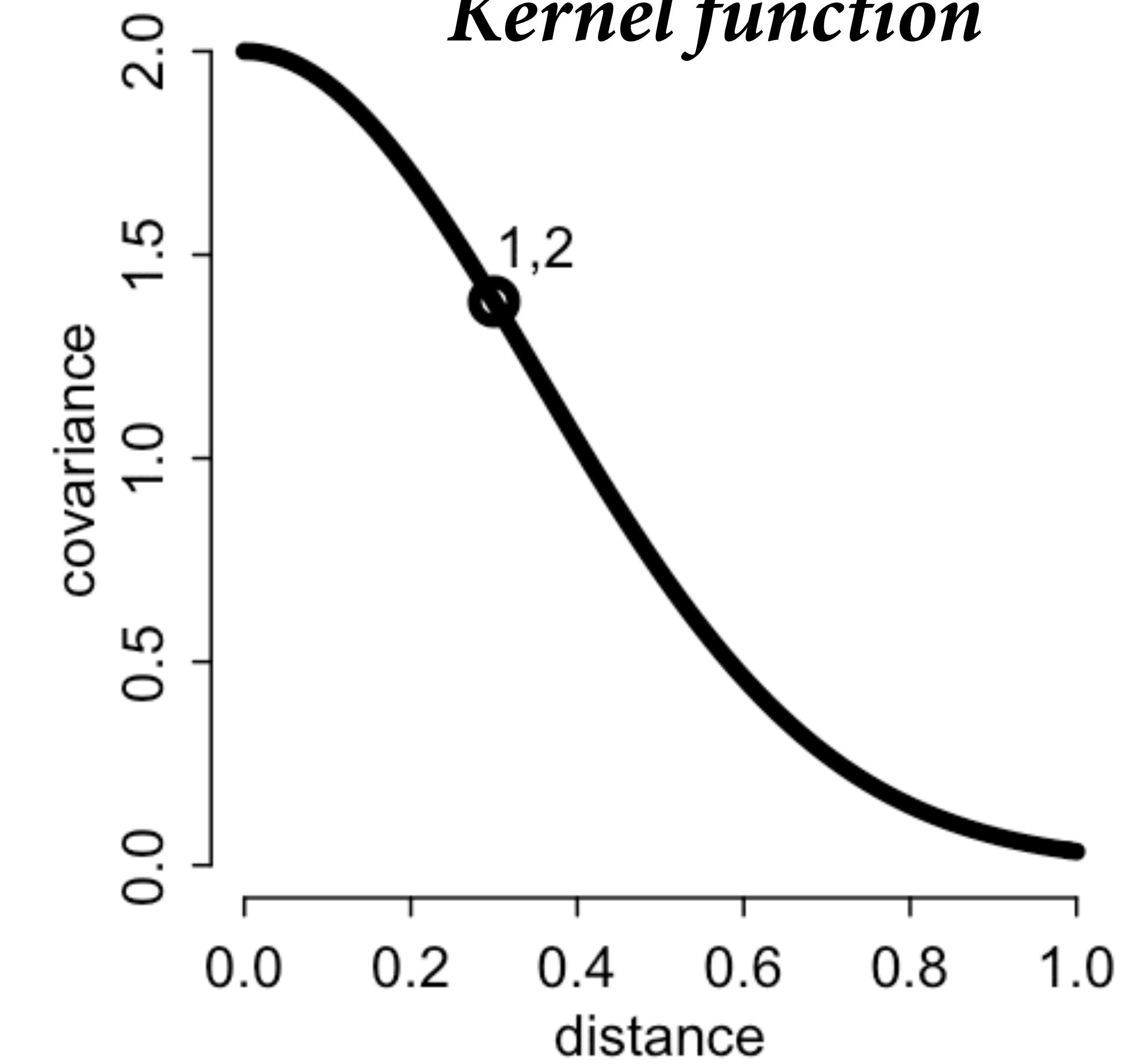
*Kernel function*



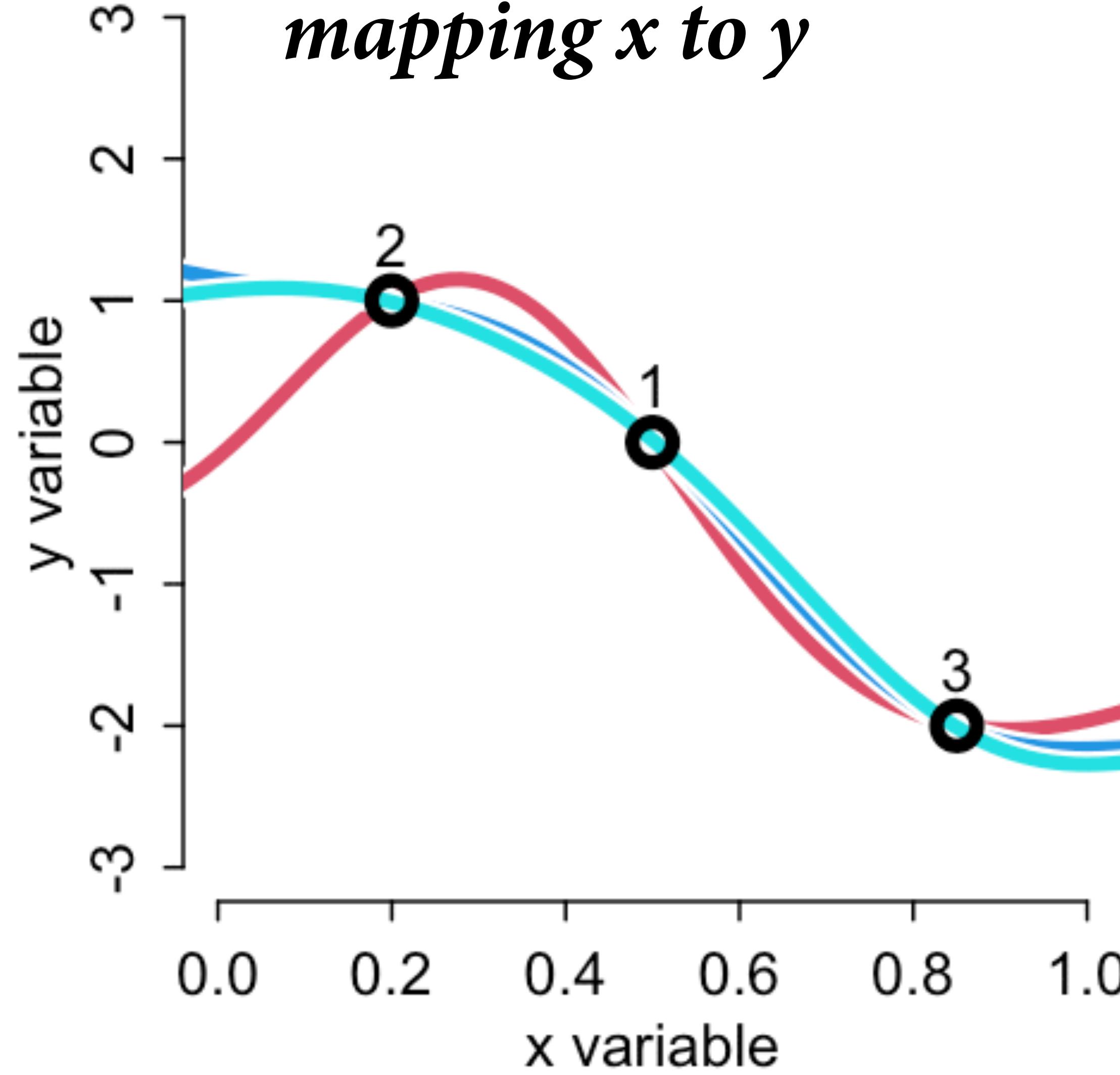
*Possible functions  
mapping  $x$  to  $y$*



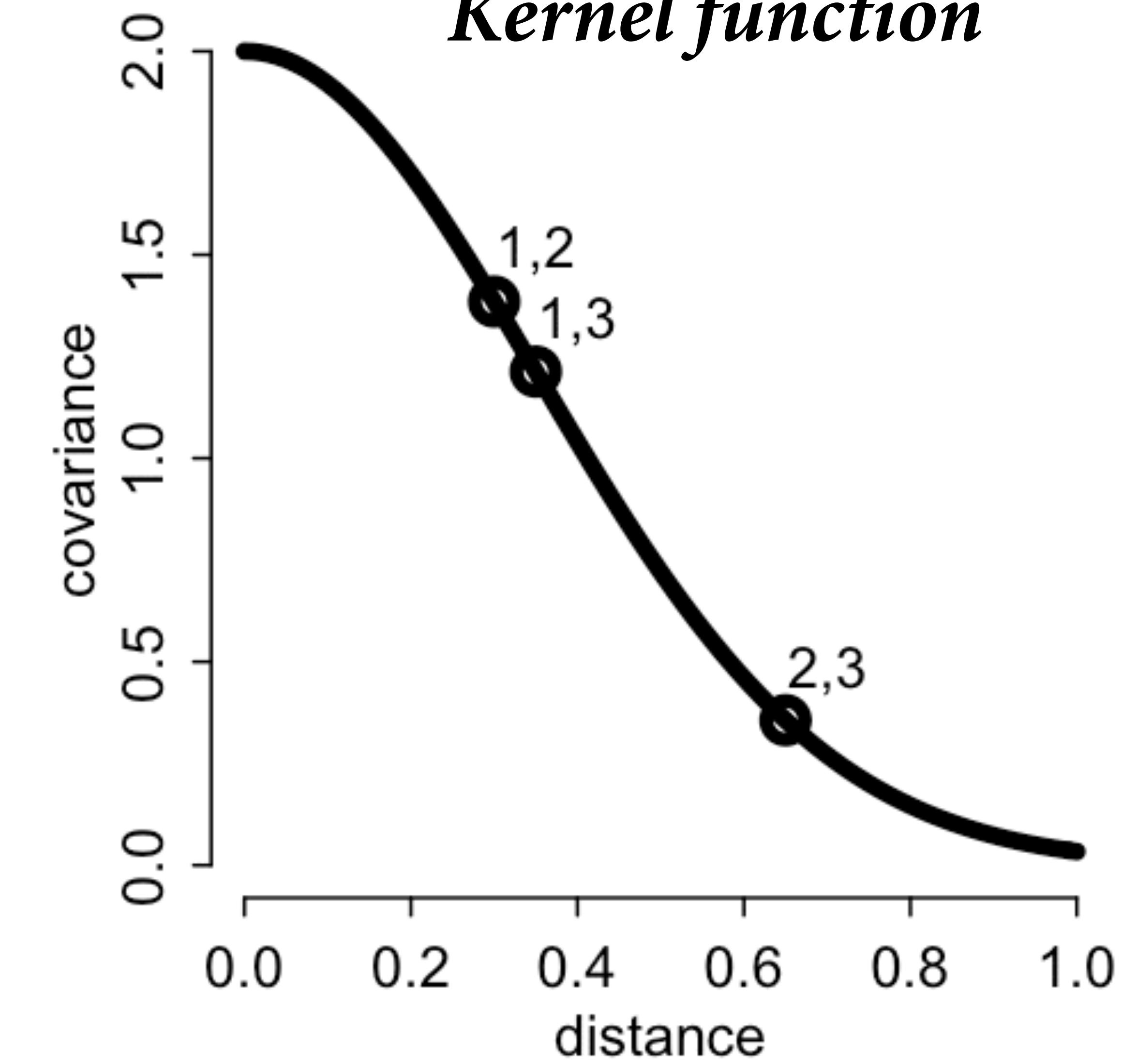
*Kernel function*



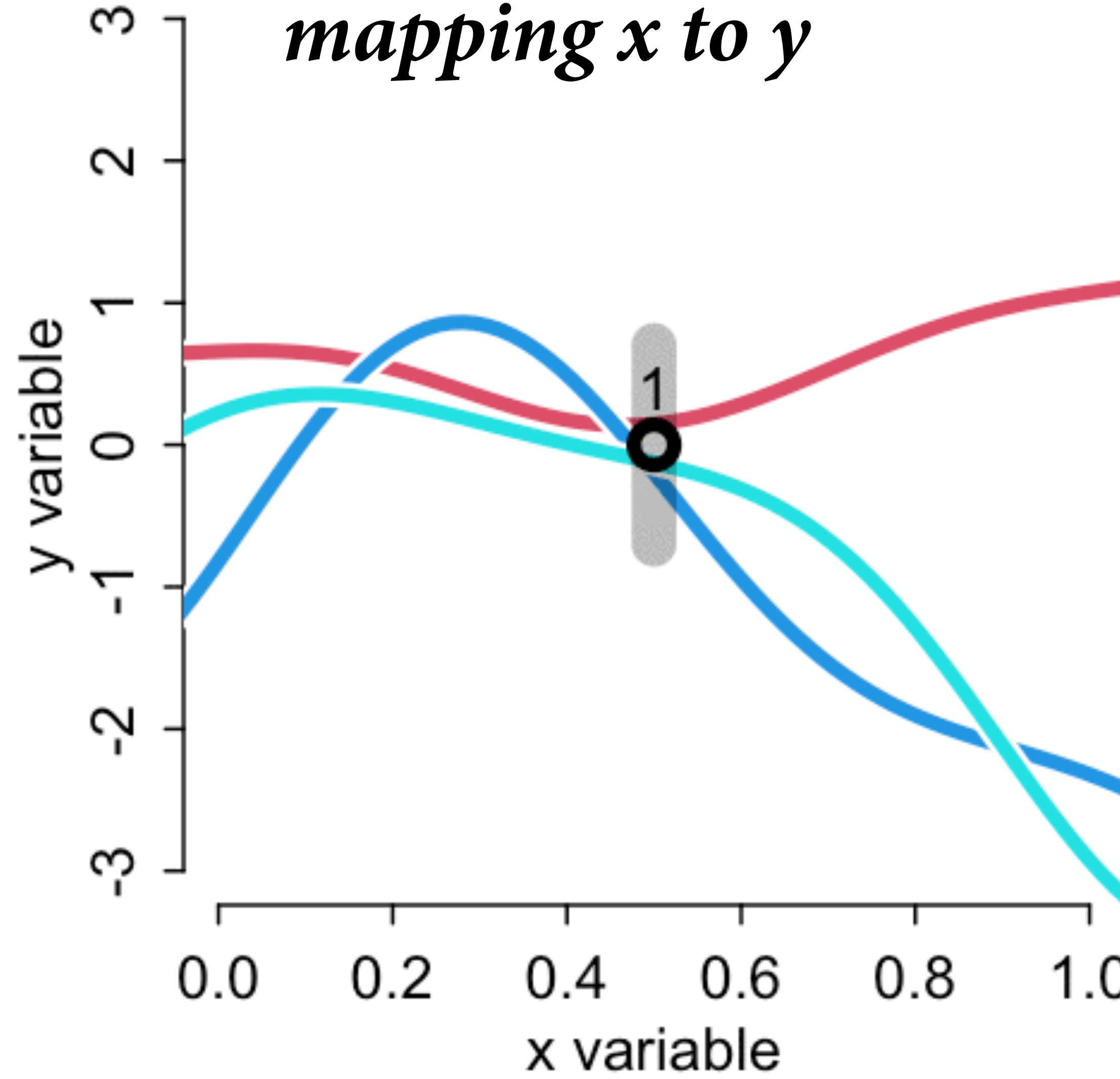
*Possible functions  
mapping  $x$  to  $y$*



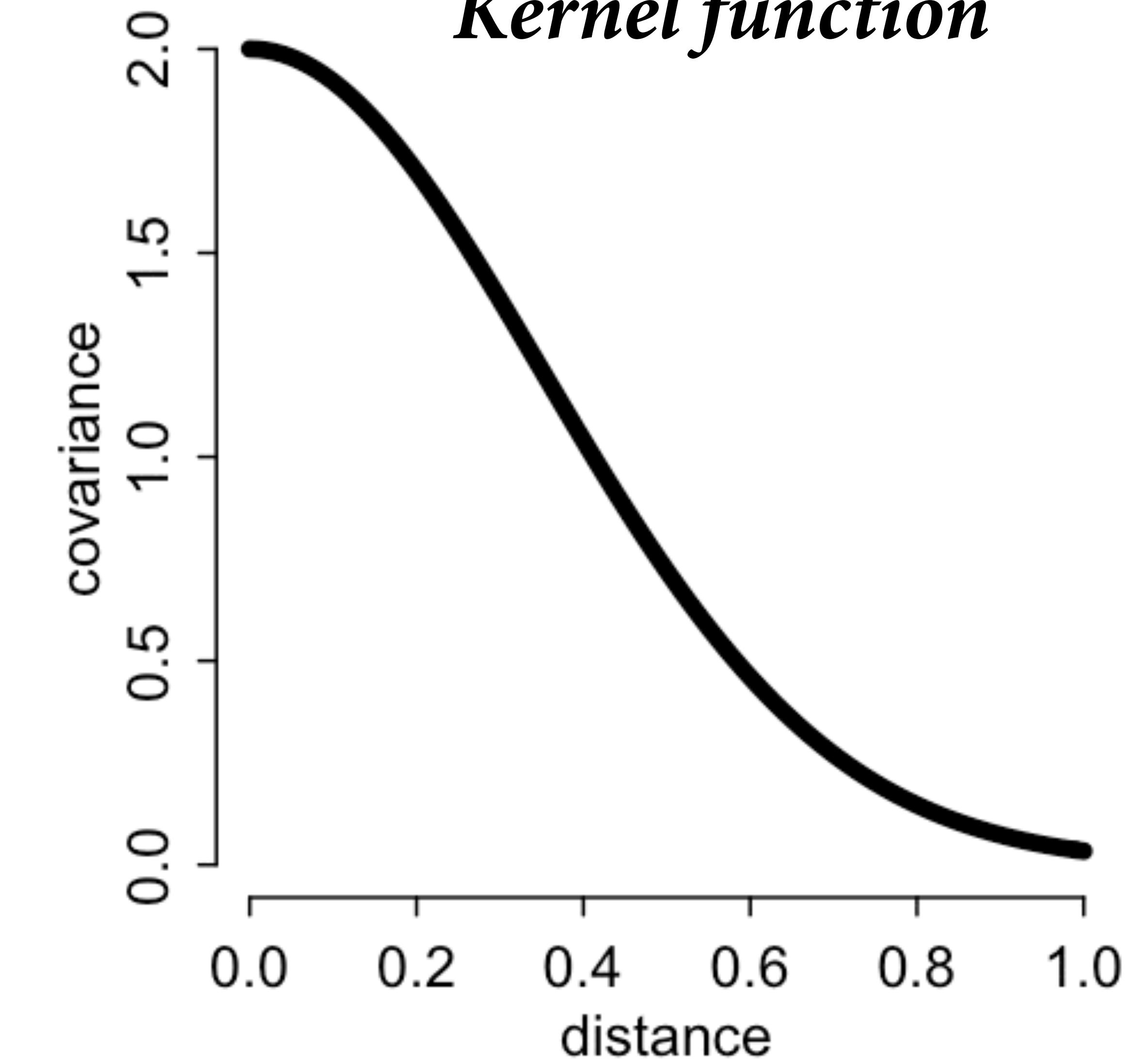
*Kernel function*

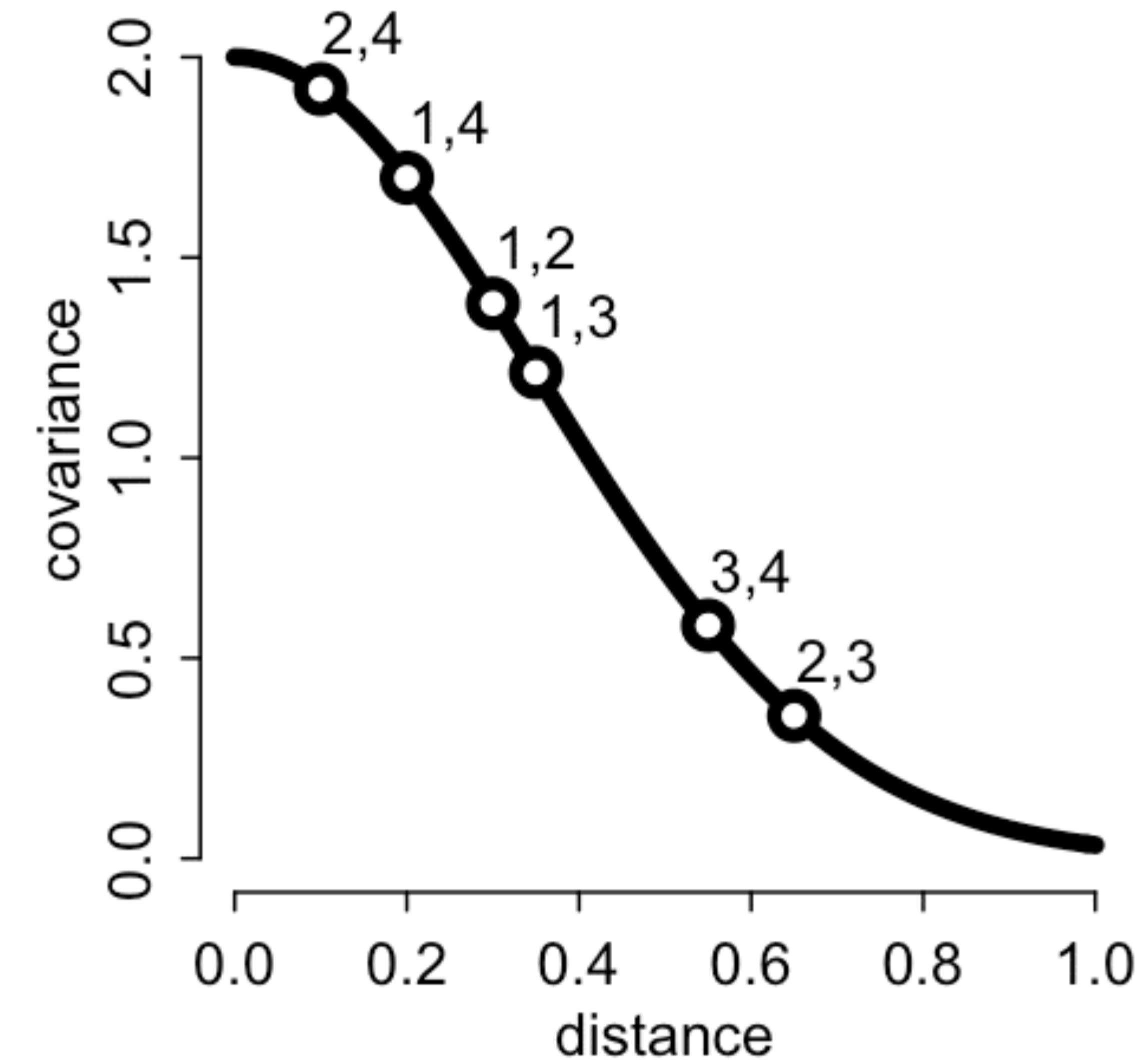
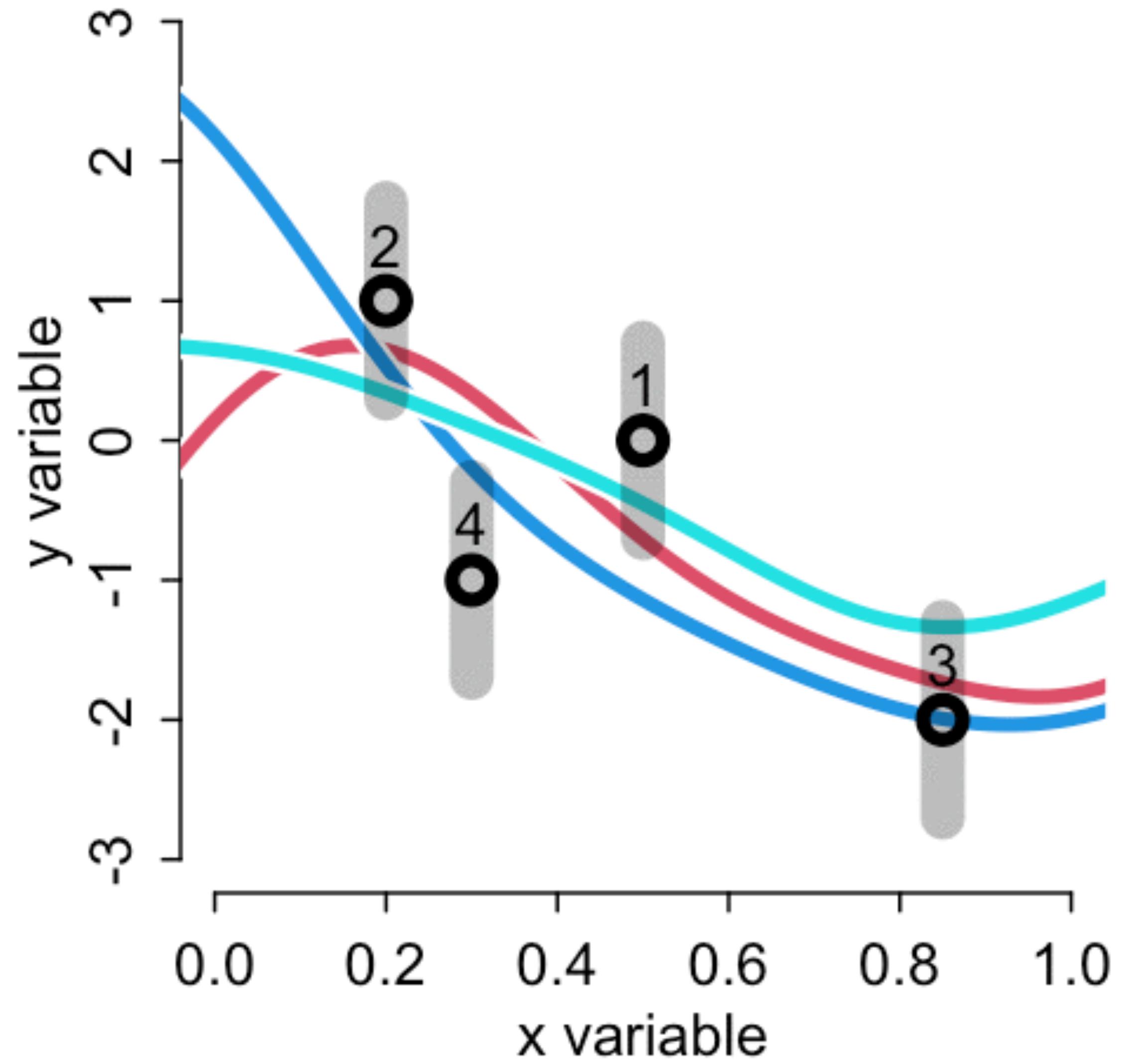


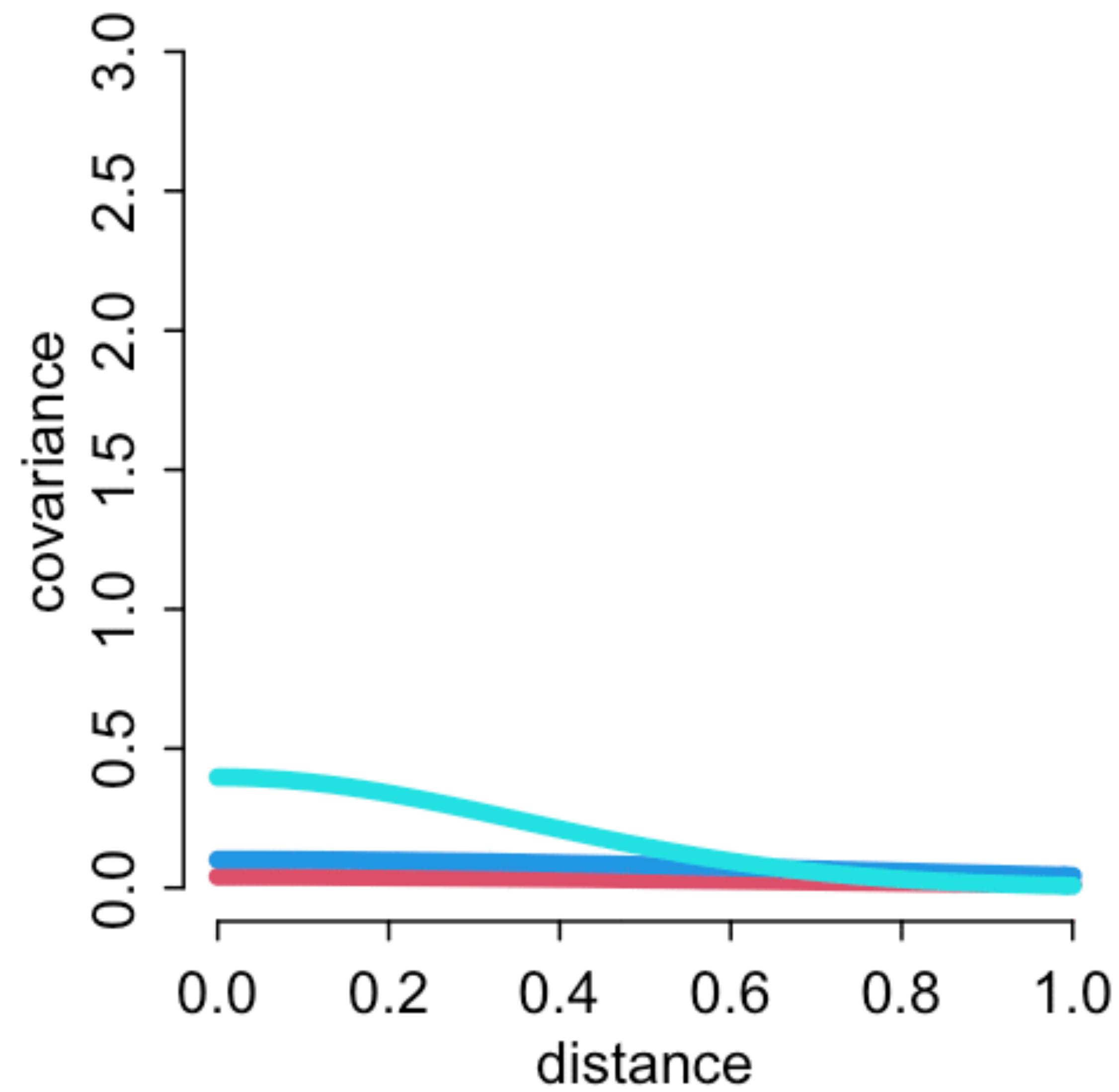
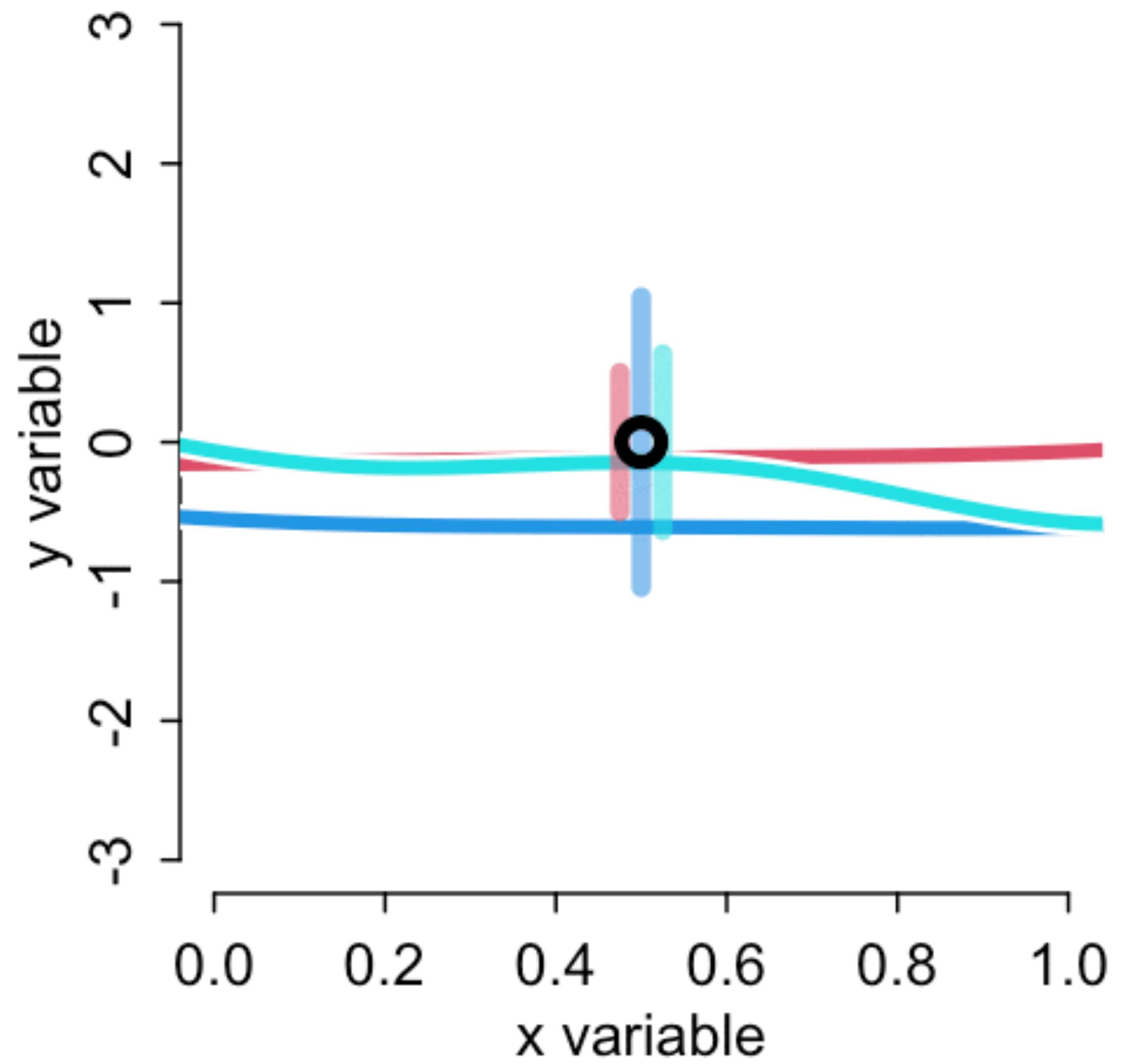
*Possible functions  
mapping  $x$  to  $y$*

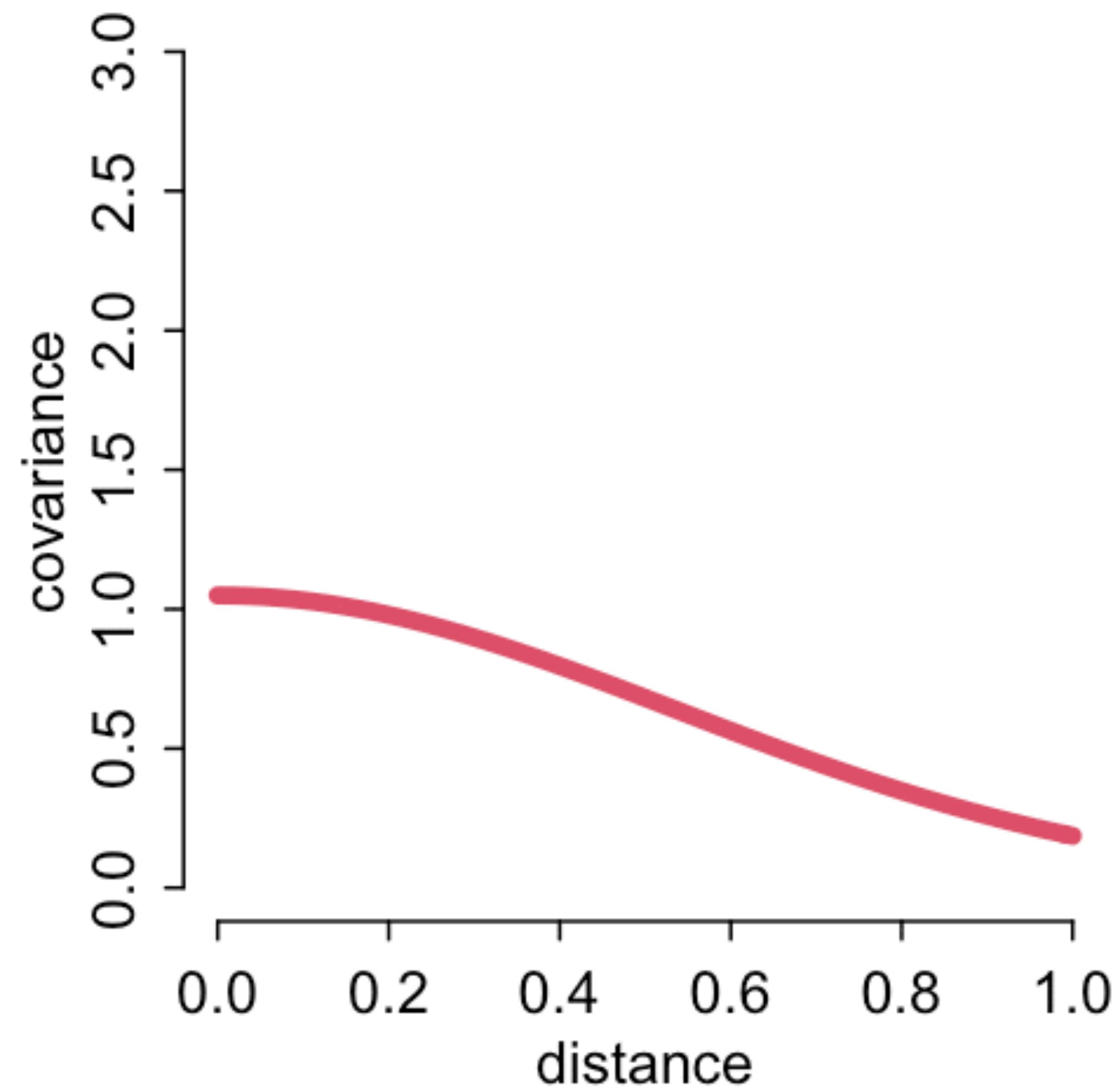
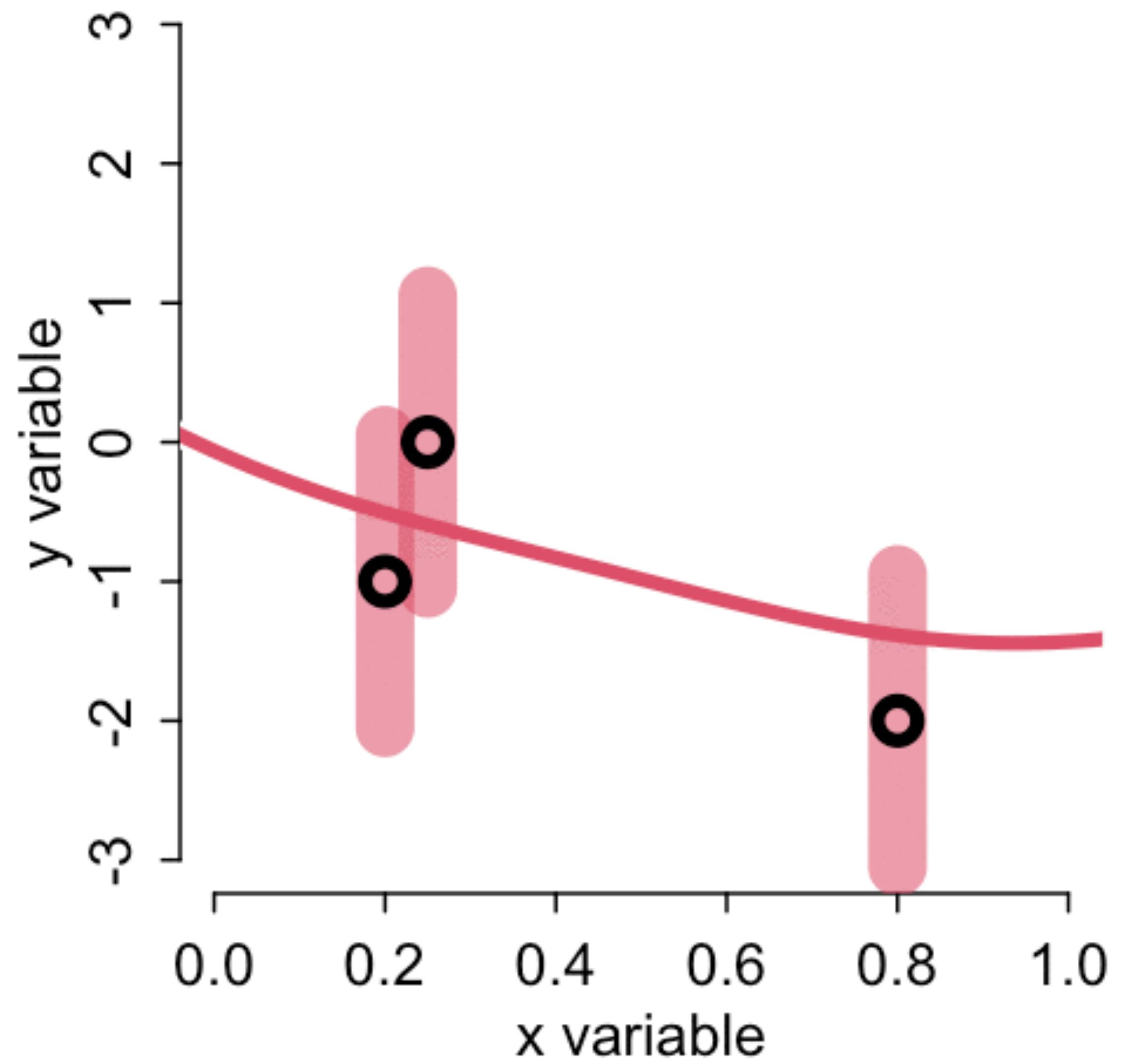


*Kernel function*



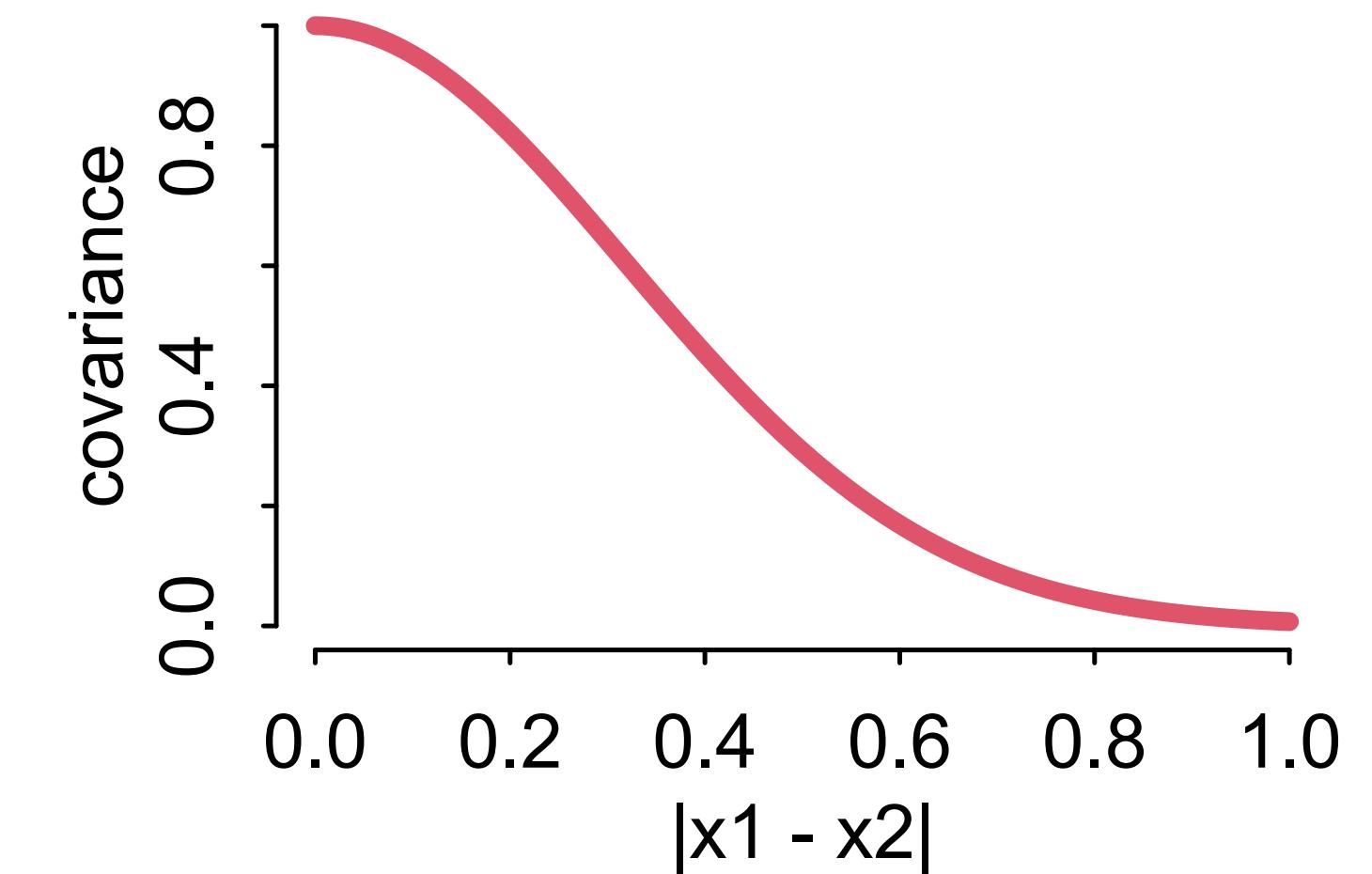






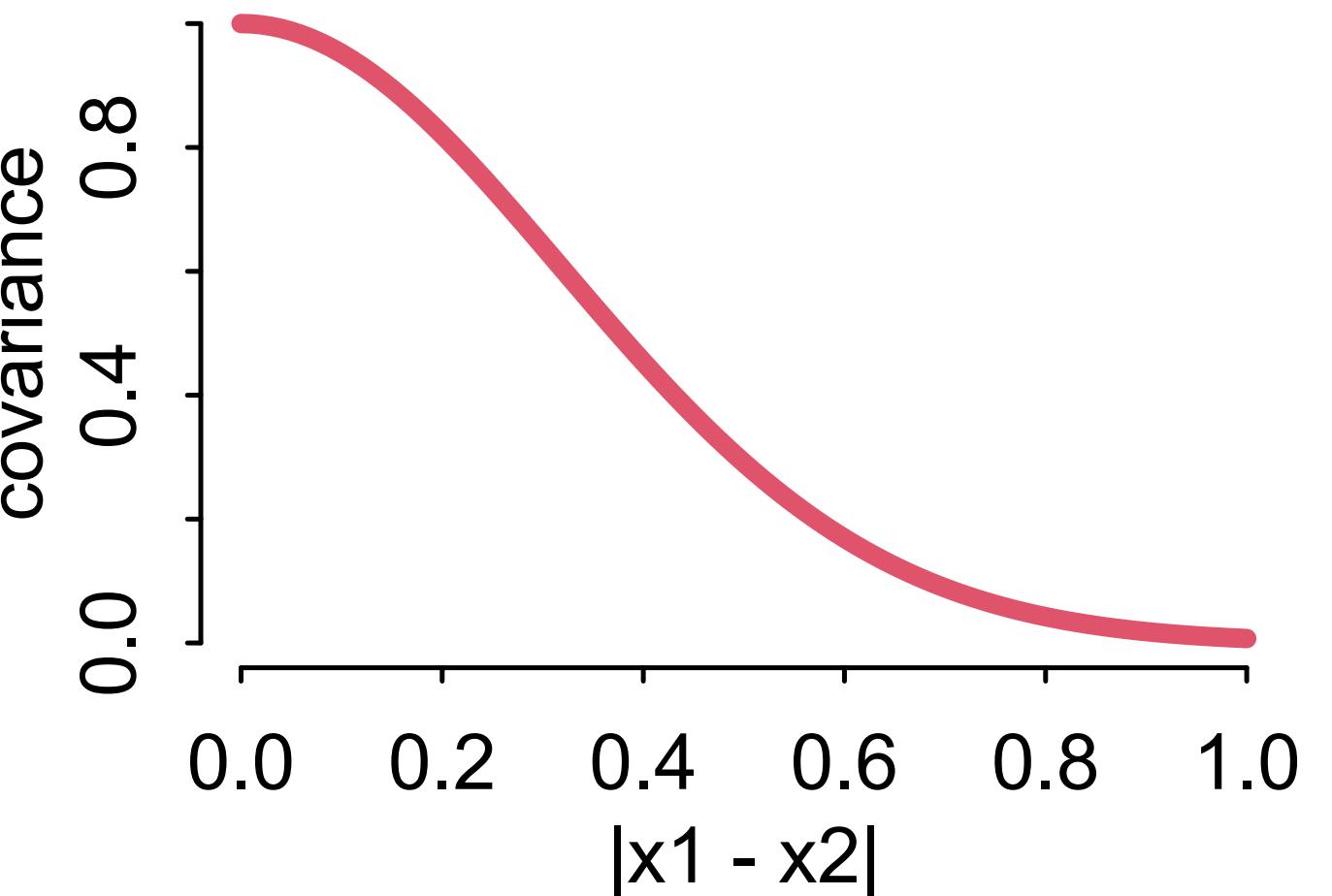
## Quadratic (L2)

$$k(x_1, x_2) = \alpha^2 \exp\left(-\frac{(x_1 - x_2)^2}{\sigma^2}\right)$$



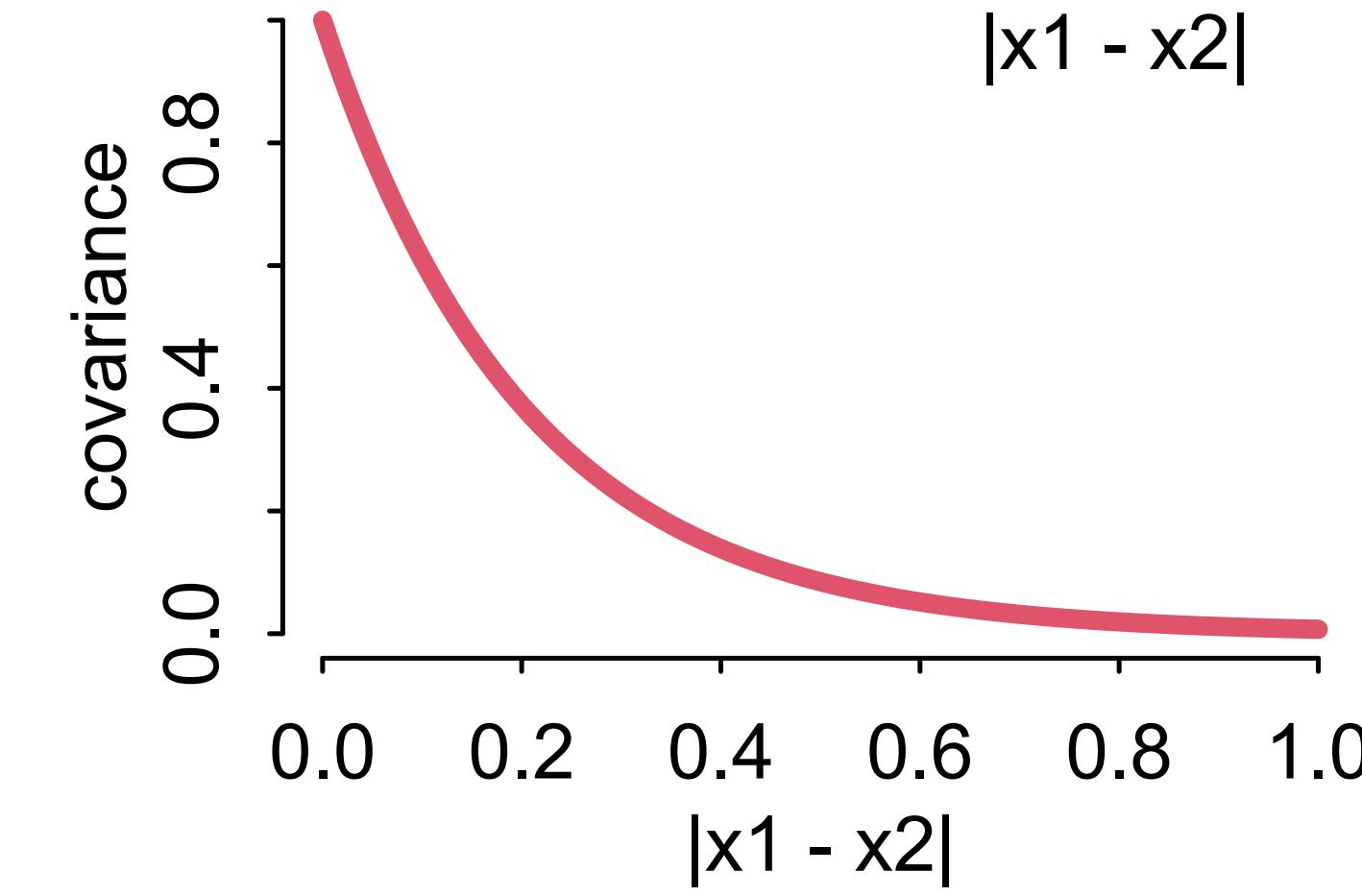
## Quadratic (L2)

$$k(x_1, x_2) = \alpha^2 \exp\left(-\frac{(x_1 - x_2)^2}{\sigma^2}\right)$$



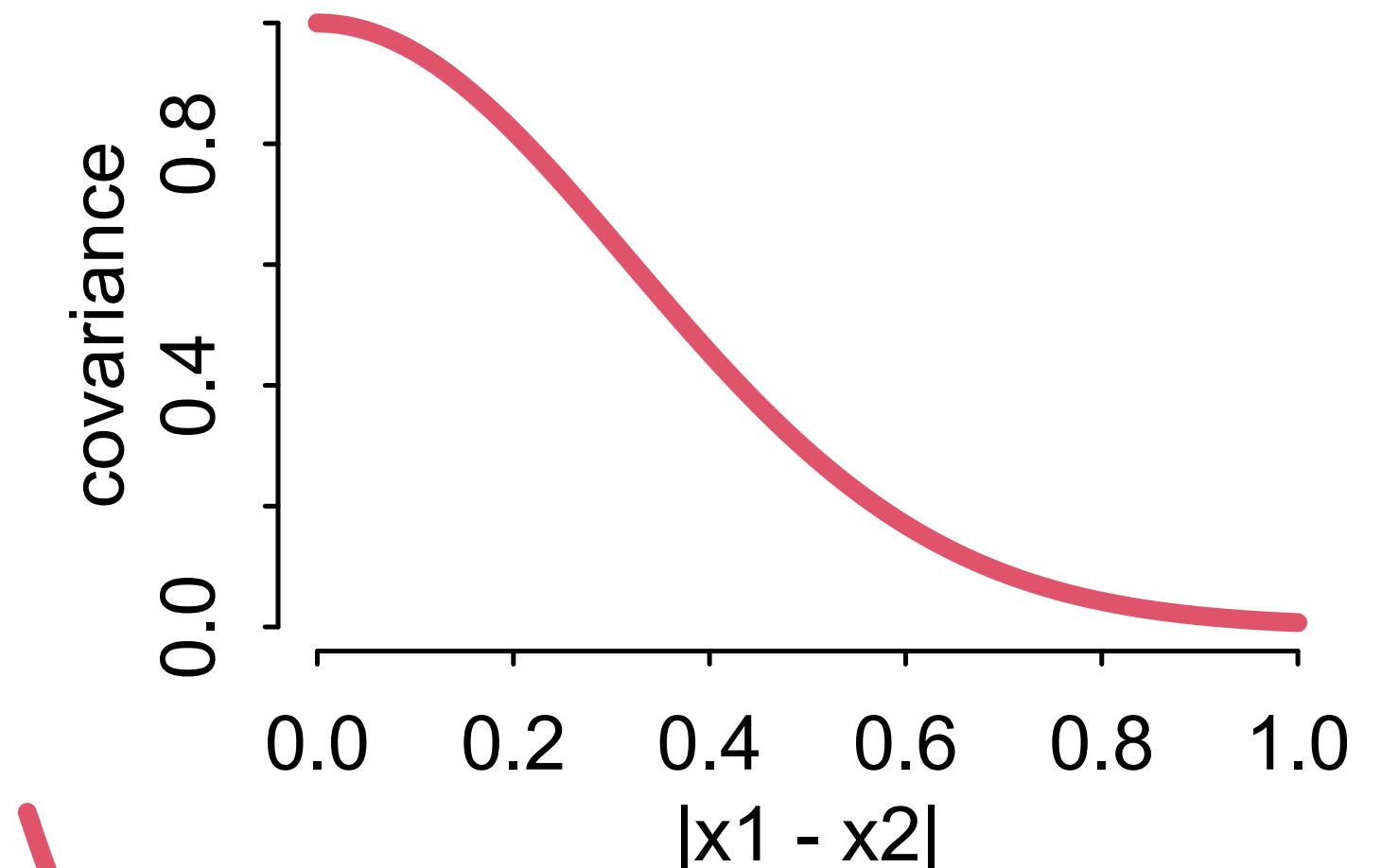
## Ornstein-Uhlenbeck

$$k(x_1, x_2) = \alpha^2 \exp\left(-\frac{|x_1 - x_2|}{\sigma}\right)$$



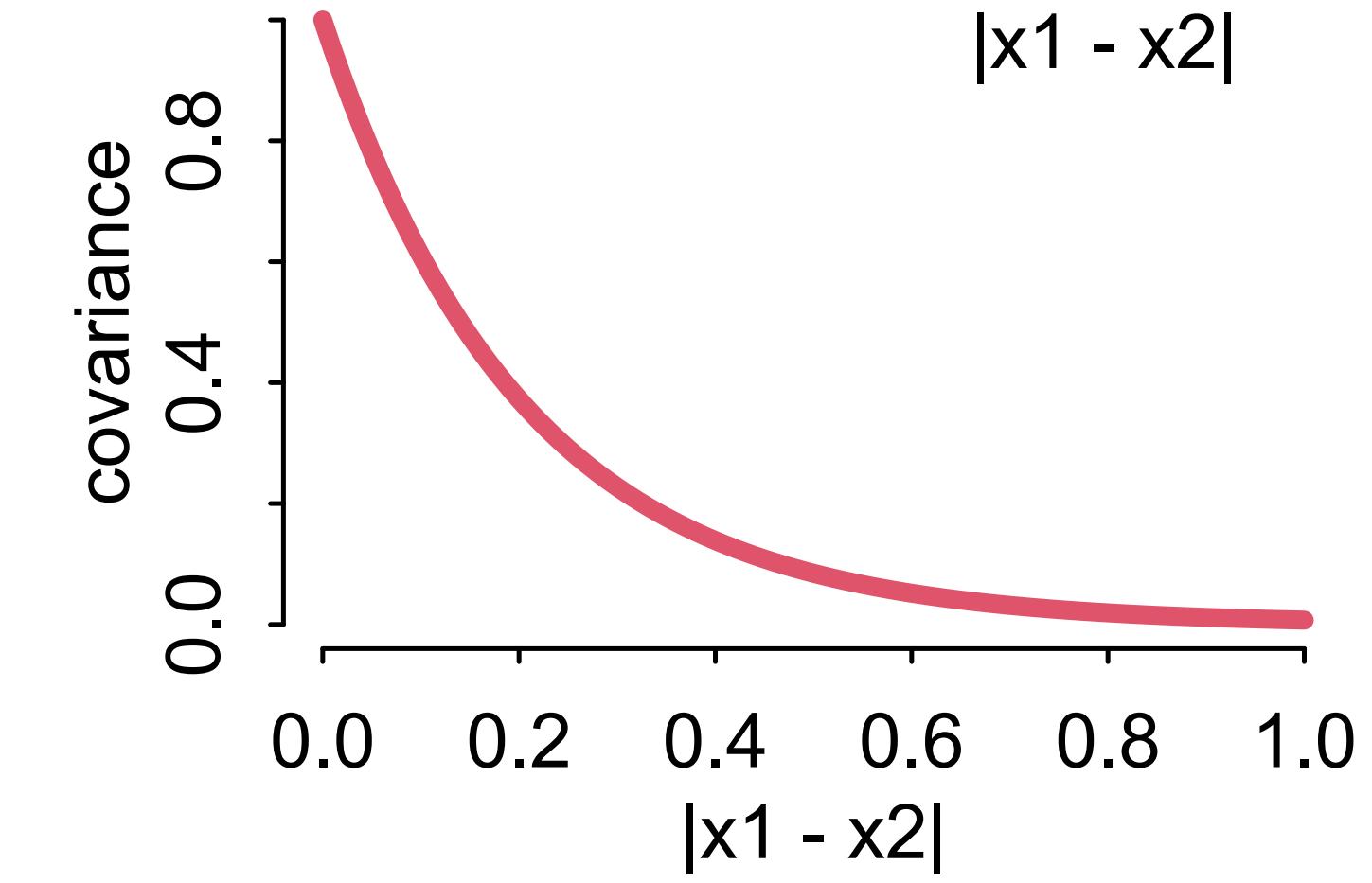
Quadratic (L2)

$$k(x_1, x_2) = \alpha^2 \exp\left(-\frac{(x_1 - x_2)^2}{\sigma^2}\right)$$



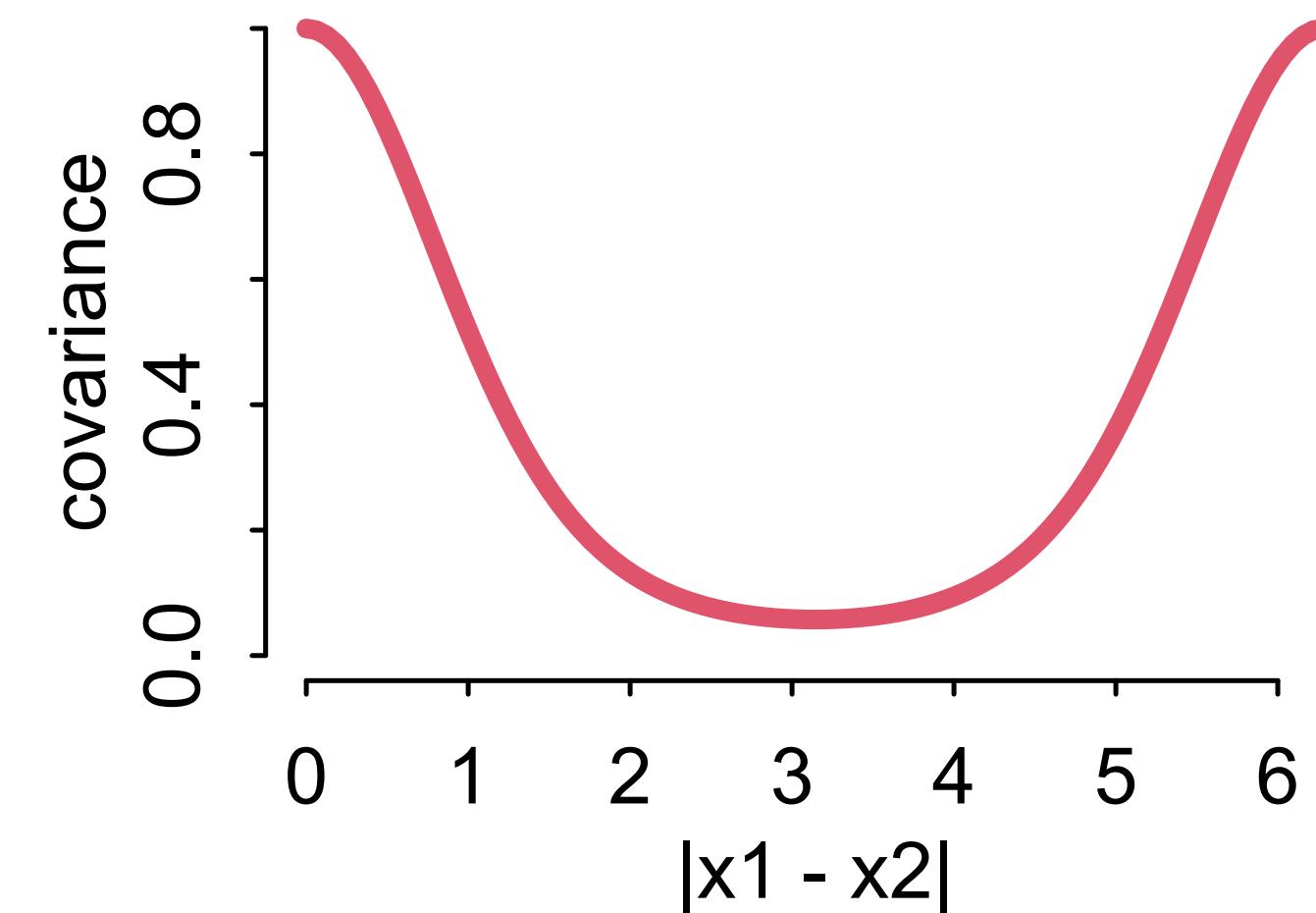
Ornstein-Uhlenbeck

$$k(x_1, x_2) = \alpha^2 \exp\left(-\frac{|x_1 - x_2|}{\sigma}\right)$$



Periodic

$$k(x_1, x_2) = \alpha^2 \exp\left(-\frac{2 \sin^2((x_1 - x_2)/2)}{\sigma^2}\right)$$



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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

$$k_{i,j} = \eta^2 \exp \left( -\rho^2 d_{i,j}^2 \right)$$

*covariance*      *maximum covariance*      *rate of decline*  
*distance  $i,j$*

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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

$$k_{i,j} = \eta^2 \exp(-\rho^2 d_{i,j}^2)$$

$$\bar{\alpha} \sim \text{Normal}(3, 0.5)$$

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

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

Distance matrix (thousand 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 |

*What do these priors imply?*

$$k_{i,j} = \eta^2 \exp\left(-\rho^2 d_{i,j}^2\right)$$

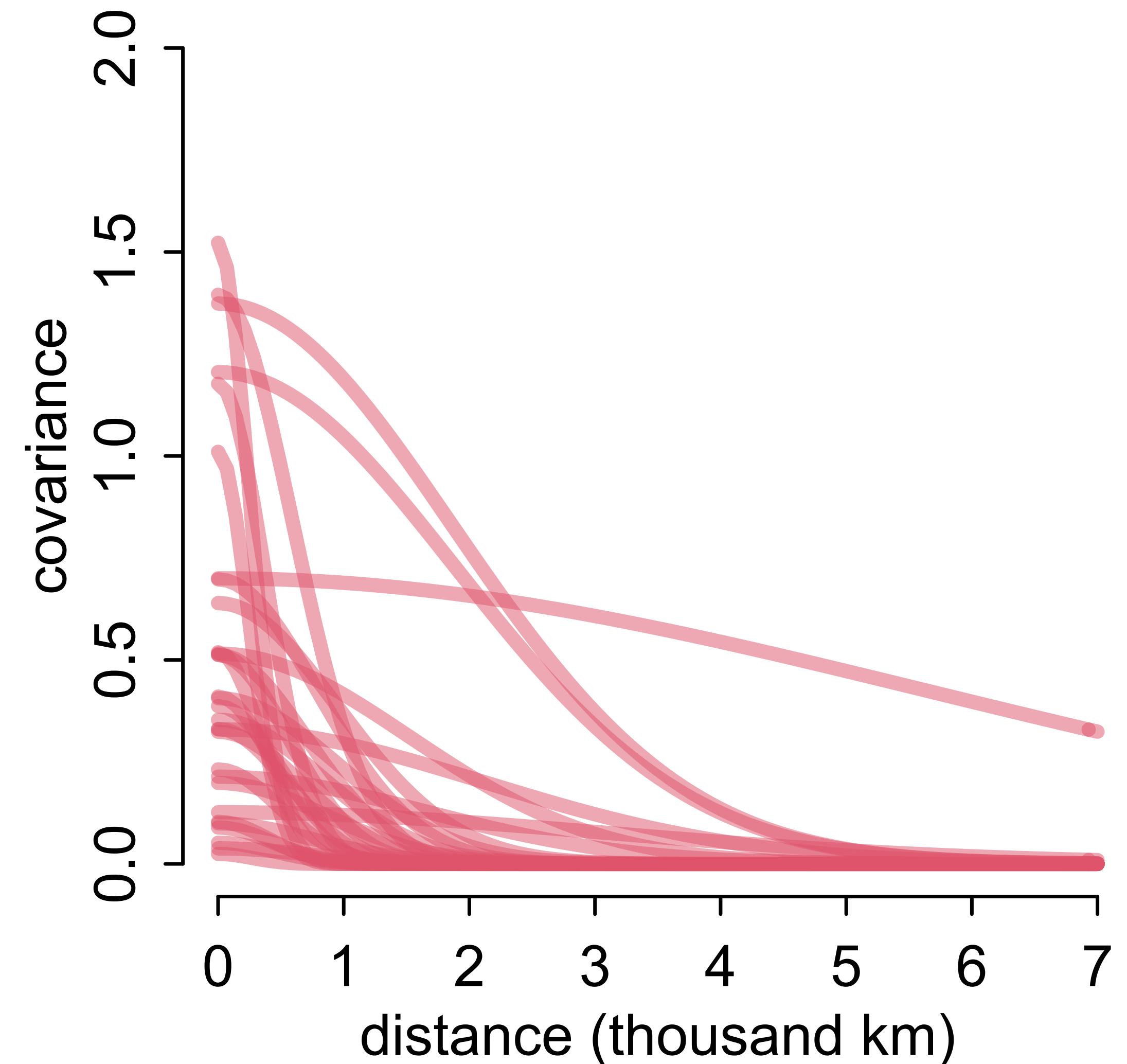
$$\eta^2 \sim \text{Exponential}(2)$$

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

```
# sim priors for distance model
n <- 30
etasq <- rexp(n,2)
rhosq <- rexp(n,0.5)

plot( NULL , xlim=c(0,7) , ylim=c(0,2) ,
      xlab="distance (thousand km)" ,
      ylab="covariance" )

for ( i in 1:n )
  curve( etasq[i]*exp(-rhosq[i]*x^2) ,
         add=TRUE , lwd=4 ,
         col=col.alpha(2,0.5) )
```



## Distance matrix (thousand km)

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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

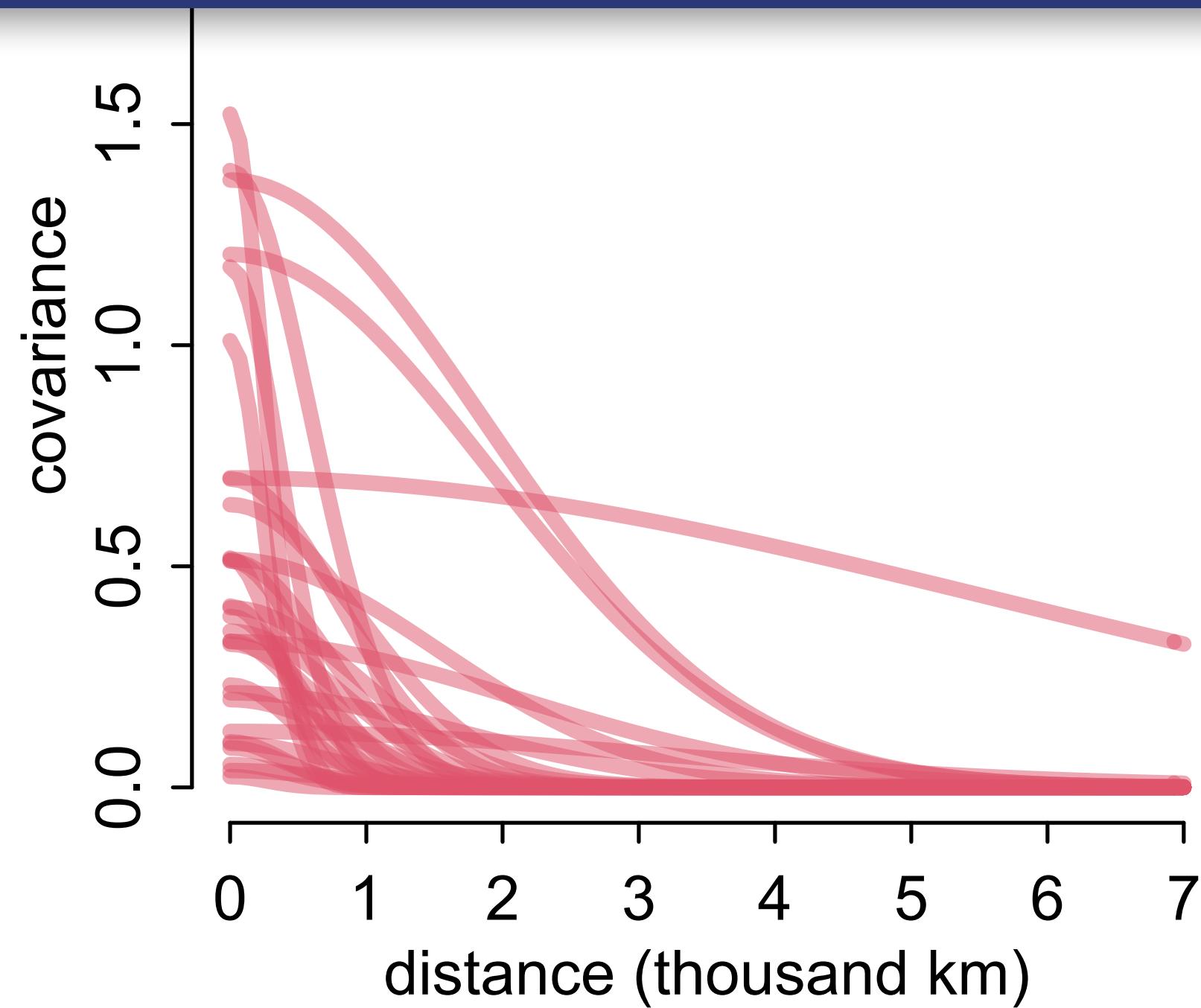
$$k_{i,j} = \eta^2 \exp \left( -\rho^2 d_{i,j}^2 \right)$$

$$\bar{\alpha} \sim \text{Normal}(3, 0.5)$$

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

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

|            | 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 |



```

data(Kline2)
d <- Kline2
data(islandsDistMatrix)

dat_list <- list(
  T = d$total_tools,
  S = 1:10,
  D = islandsDistMatrix )

mTdist <- ulam(
  alist(
    T ~ dpois(lambda),
    log(lambda) <- abar + a[S],
    vector[10]:a ~ multi_normal( 0 , K ),
    matrix[10,10]:K <- cov_GPL2(D,etasq,rhosq,0.01),
    abar ~ normal(3,0.5),
    etasq ~ dexp( 2 ),
    rhosq ~ dexp( 0.5 )
  ), data=dat_list , chains=4 , cores=4 , iter=4000 )

```

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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

$$k_{i,j} = \eta^2 \exp \left( -\rho^2 d_{i,j}^2 \right)$$

$$\bar{\alpha} \sim \text{Normal}(3,0.5)$$

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

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

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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

```

data(Kline2)
d <- Kline2
data(islandsDistMatrix)

dat_list <- list(
  T = d$total_tools,
  S = 1:10,
  D = islandsDistMatrix )

mTdist <- ulam(
  alist(
    T ~ dpois(lambda),
    log(lambda) <- abar + a[S],
    vector[10]:a ~ multi_normal( 0 , K ),
    matrix[10,10]:K <- cov_GPL2(D,etasq,rhosq,0.01),
    abar ~ normal(3,0.5),
    etasq ~ dexp( 2 ),
    rhosq ~ dexp( 0.5 )
  ), data=dat_list , chains=4 , cores=4 , iter=4000 )

```

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

$$k_{i,j} = \eta^2 \exp \left( -\rho^2 d_{i,j}^2 \right)$$

$$\bar{\alpha} \sim \text{Normal}(3,0.5)$$

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

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

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

$$\log \lambda_i = \bar{\alpha} + \alpha_{S[i]}$$

```

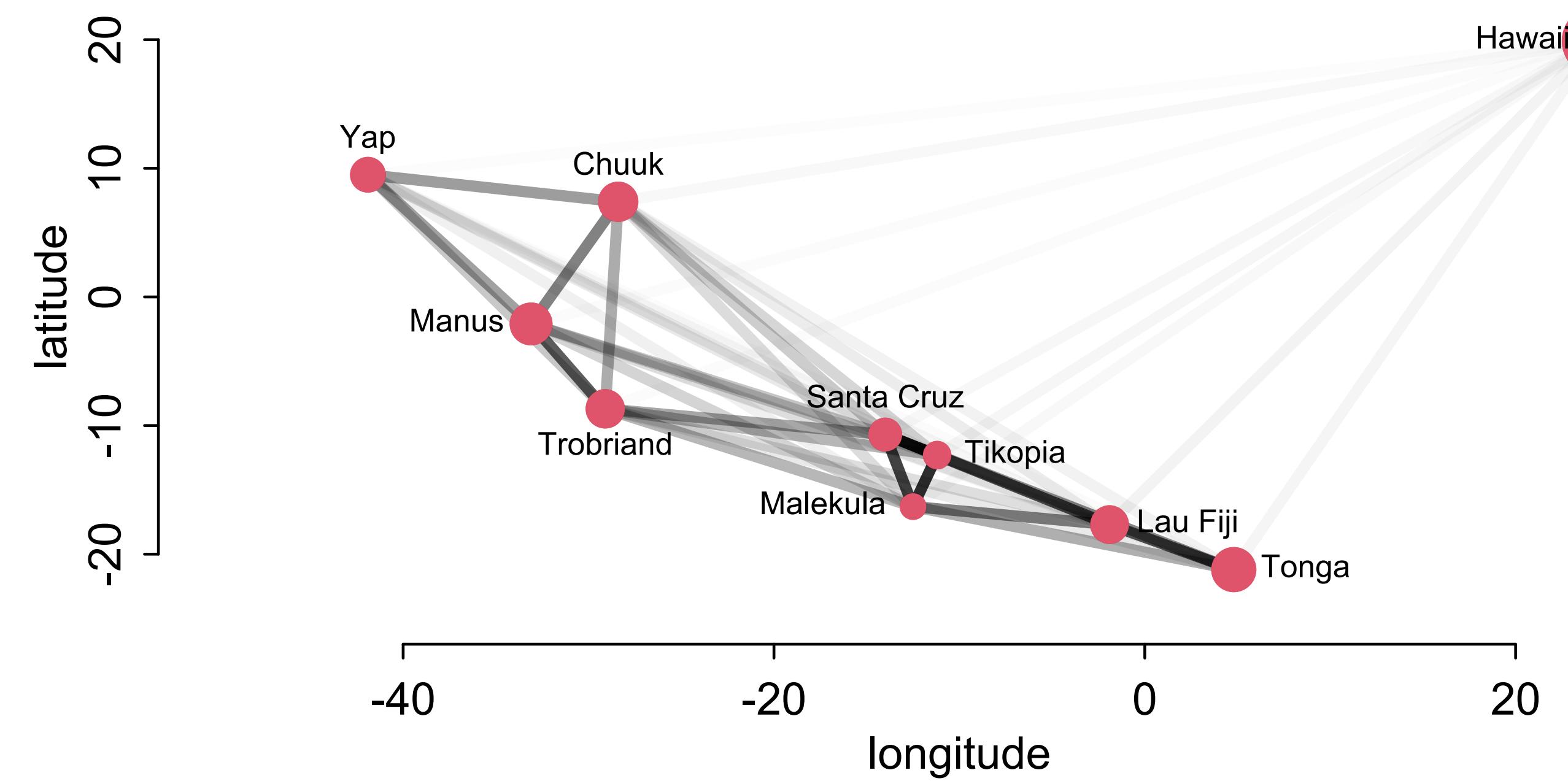
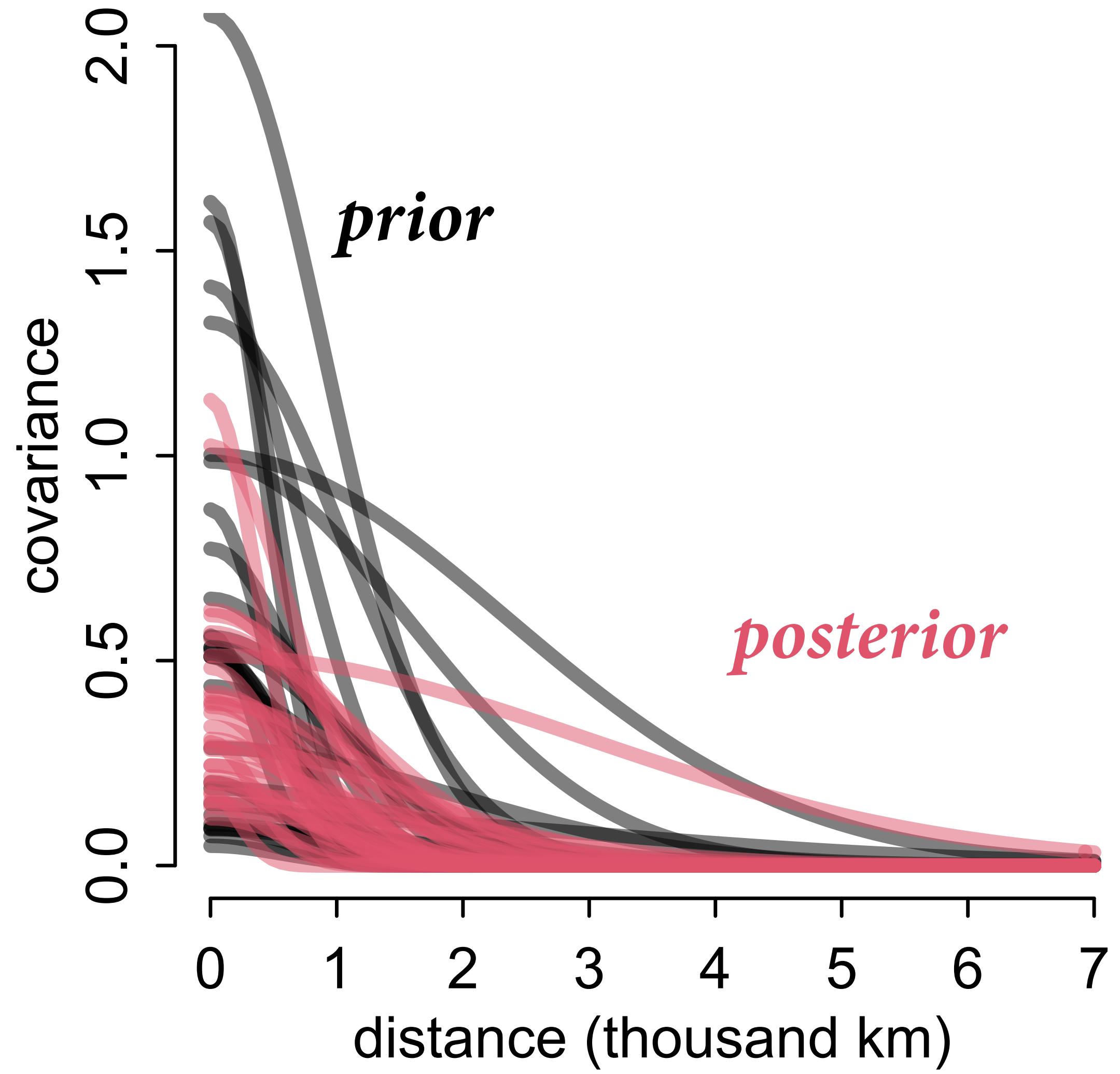
data(Kline2)
d <- Kline2
data(islandsDistMatrix)

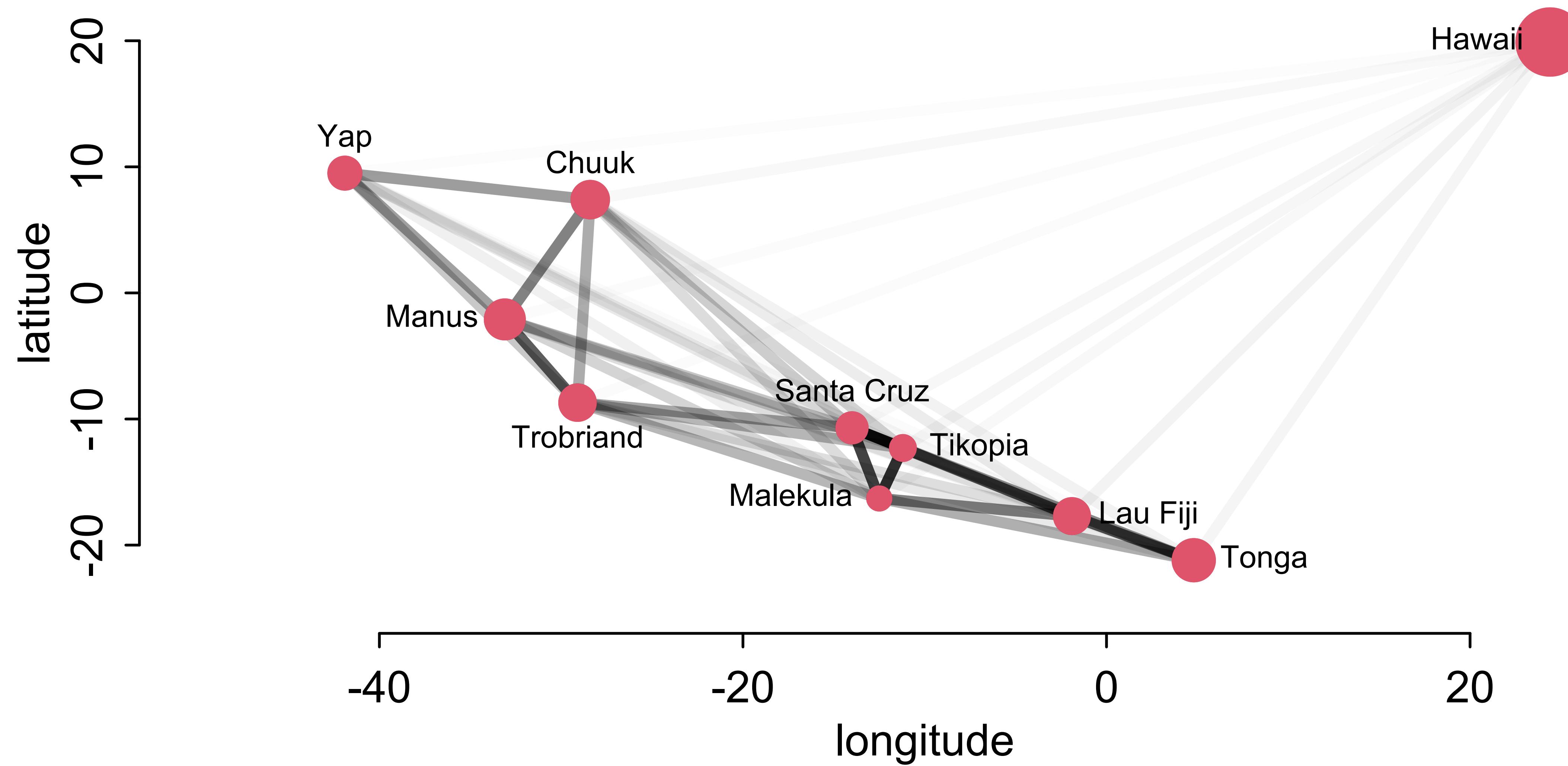
dat_list <- list(
  T = d$total_tools,
  S = 1:10,
  D = islandsDistMatrix )

mTdist <- ulam(
  alist(
    T ~ dpois(lambda),
    log(lambda) <- abar + a[S],
    vector[10]:a ~ multi_normal( 0
    matrix[10,10]:K <- cov_GPL2(D,epsilon),
    abar ~ normal(3,0.5),
    etasq ~ dexp( 2 ),
    rhosq ~ dexp( 0.5 )
  ), data=dat_list , chains=4 , cores=4 )

```

|       | > precis(mTdist,2) |      |       |       |       |       |
|-------|--------------------|------|-------|-------|-------|-------|
|       | mean               | sd   | 5.5%  | 94.5% | n_eff | Rhat4 |
| a[1]  | -0.67              | 0.31 | -1.16 | -0.18 | 1385  | 1     |
| a[2]  | -0.43              | 0.30 | -0.89 | 0.06  | 1075  | 1     |
| a[3]  | -0.38              | 0.30 | -0.87 | 0.10  | 1137  | 1     |
| a[4]  | 0.24               | 0.28 | -0.20 | 0.71  | 1133  | 1     |
| a[5]  | 0.02               | 0.29 | -0.44 | 0.49  | 1124  | 1     |
| a[6]  | -0.48              | 0.30 | -0.96 | -0.01 | 1303  | 1     |
| a[7]  | 0.16               | 0.29 | -0.28 | 0.63  | 1154  | 1     |
| a[8]  | -0.17              | 0.30 | -0.65 | 0.29  | 1164  | 1     |
| a[9]  | 0.45               | 0.28 | 0.01  | 0.90  | 1053  | 1     |
| a[10] | 0.73               | 0.27 | 0.30  | 1.18  | 1027  | 1     |
| abar  | 3.49               | 0.25 | 3.08  | 3.88  | 896   | 1     |
| etasq | 0.38               | 0.27 | 0.12  | 0.87  | 1989  | 1     |
| rhosq | 1.09               | 1.39 | 0.08  | 3.68  | 3302  | 1     |





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

```

dat_list <- list(
  T = d$total_tools,
  P = d$population,
  S = 1:10,
  D = islandsDistMatrix )

mTDP <- ulam(
  alist(
    T ~ dpois(lambda),
    lambda <- (abar * P^b / g) * exp(a[S]),
    vector[10]:a ~ multi_normal( 0 , K ),
    transpars> matrix[10,10]:K <-
cov_GPL2(D,etasq,rhosq,0.01),
    c(abar,b,g) ~ dexp( 1 ),
    etasq ~ dexp( 2 ),
    rhosq ~ dexp( 0.5 )
  ), data=dat_list , chains=4 , cores=4 , iter=4000 )

```

$$\lambda_i = \frac{\bar{\alpha} P^\beta}{\gamma} \exp(\alpha_{S[i]})$$

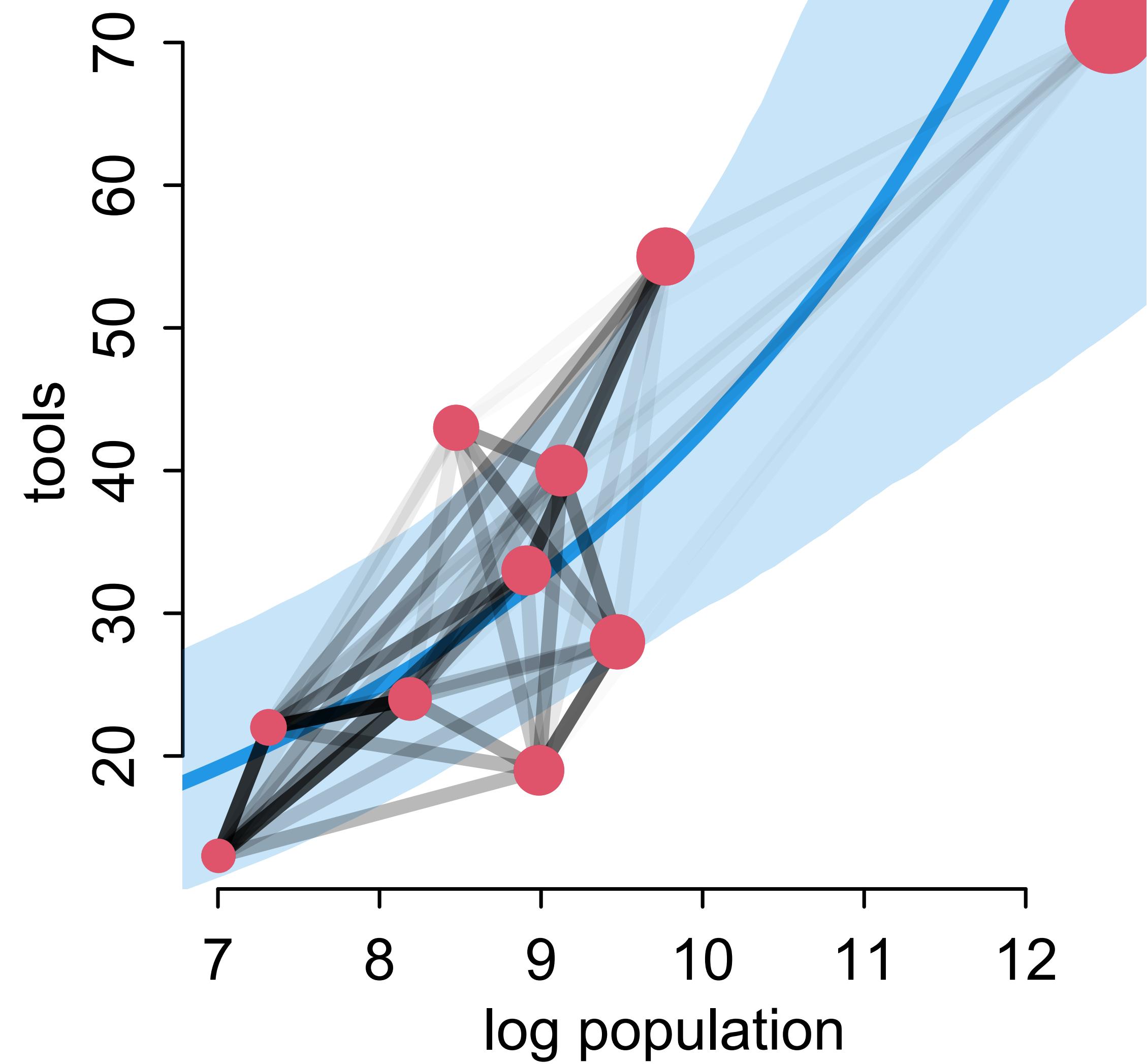
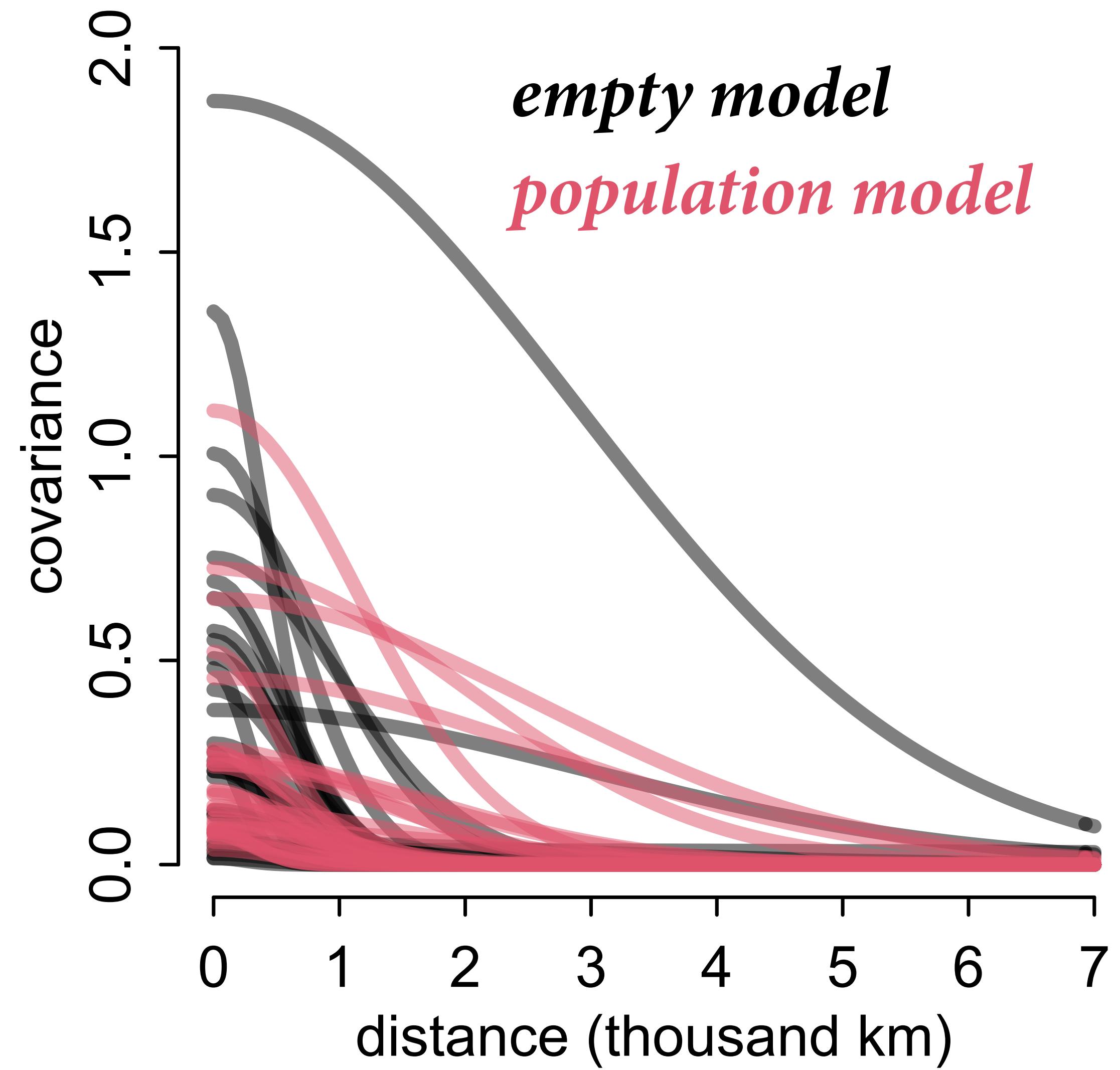
$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{K} \right)$$

$$k_{i,j} = \eta^2 \exp \left( -\rho^2 d_{i,j}^2 \right)$$

$$\bar{\alpha}, \beta, \gamma \sim \text{Exponential}(1)$$

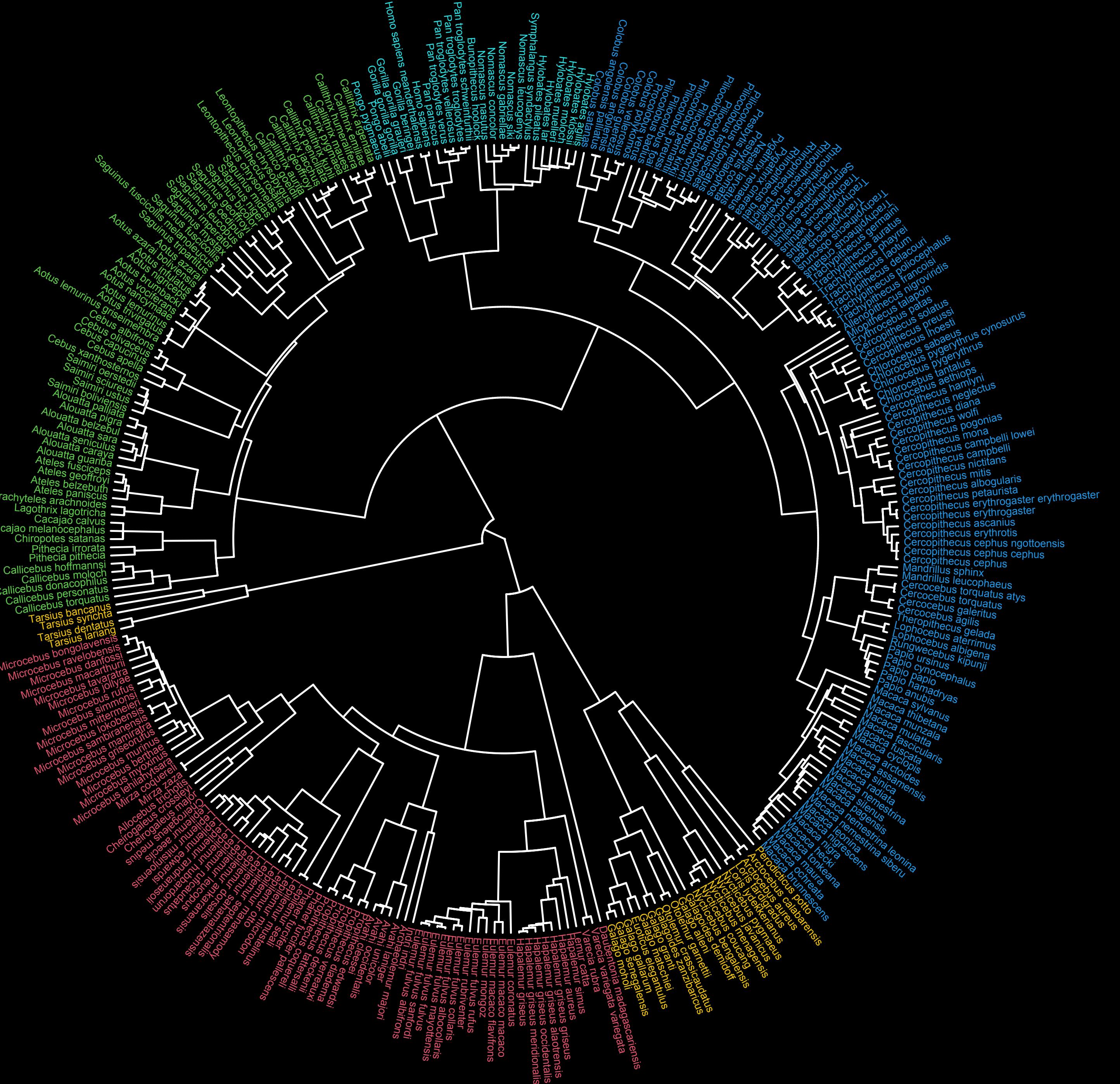
$$\eta^2 \sim \text{Exponential}(2)$$

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



**PAUSE**

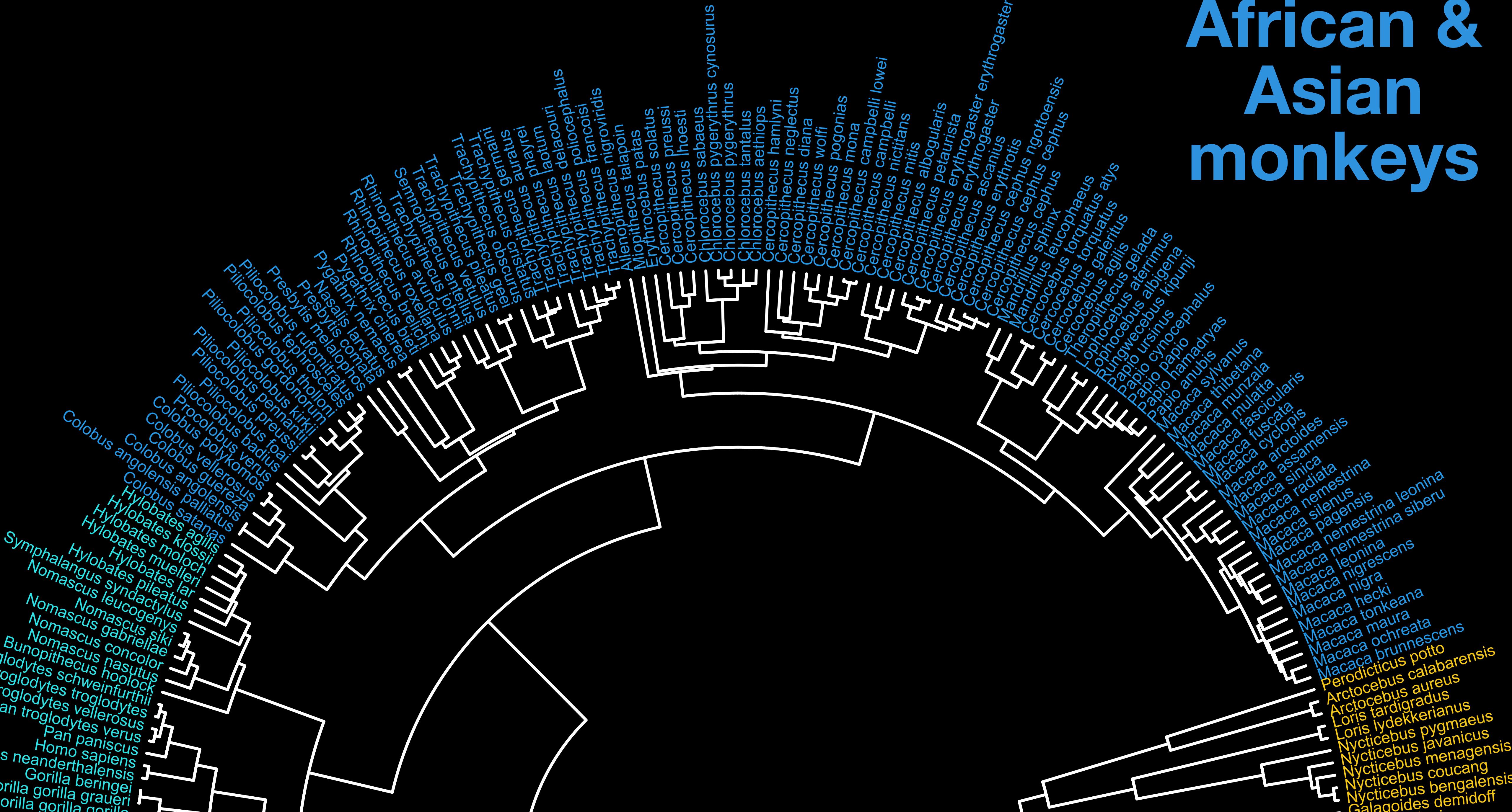
# Primates



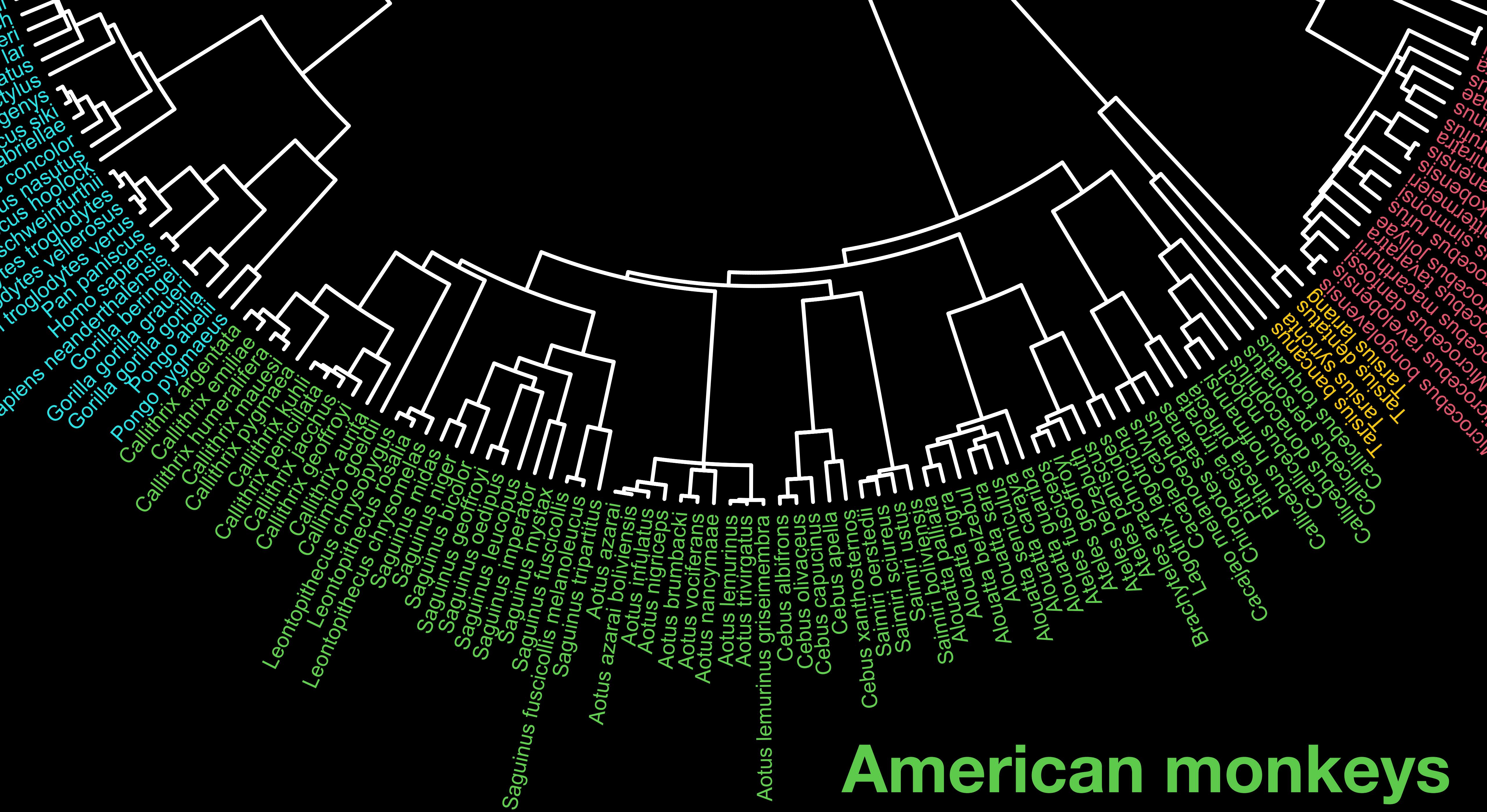
# Apes



# African & Asian monkeys

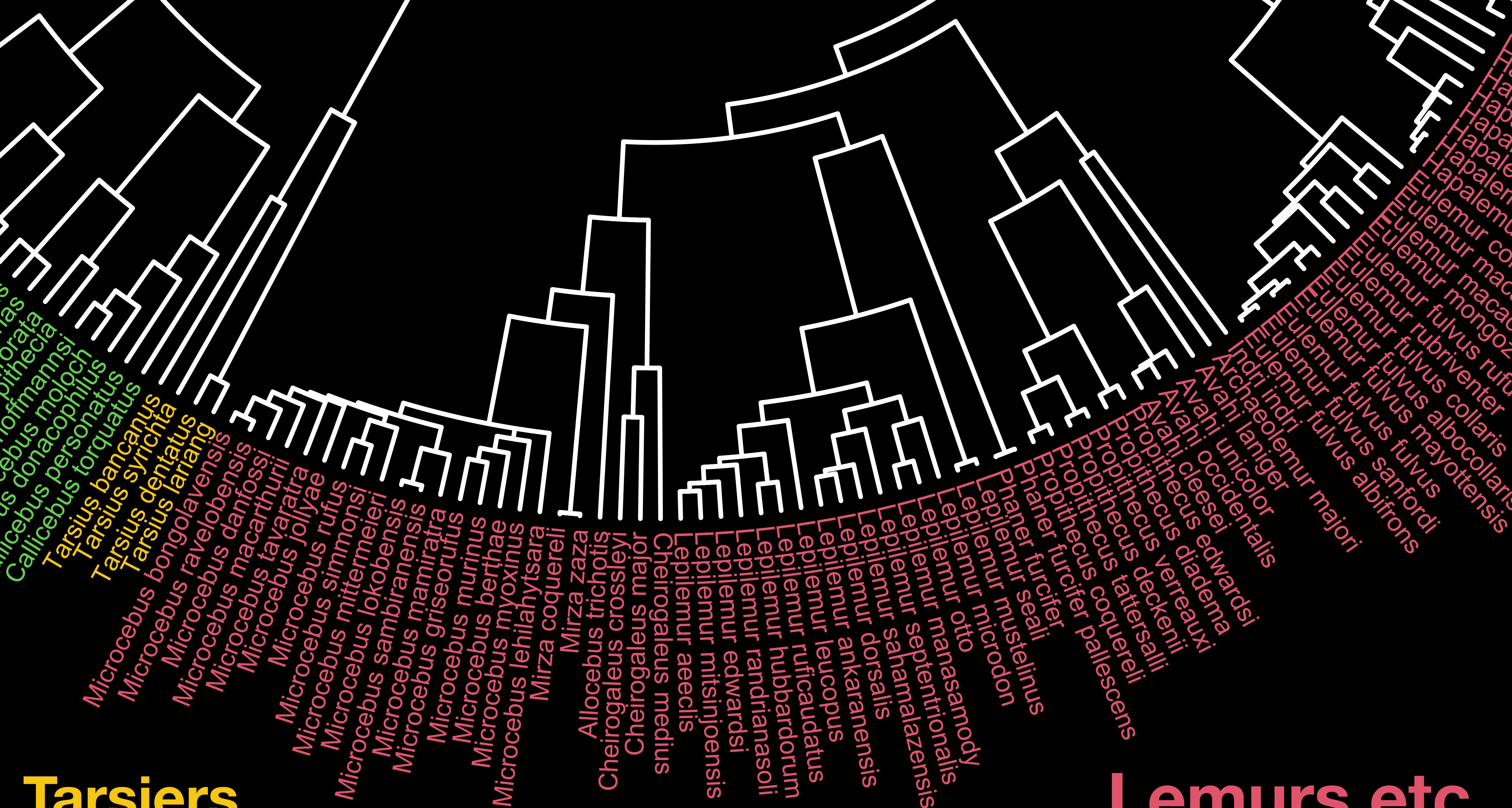


# American monkeys

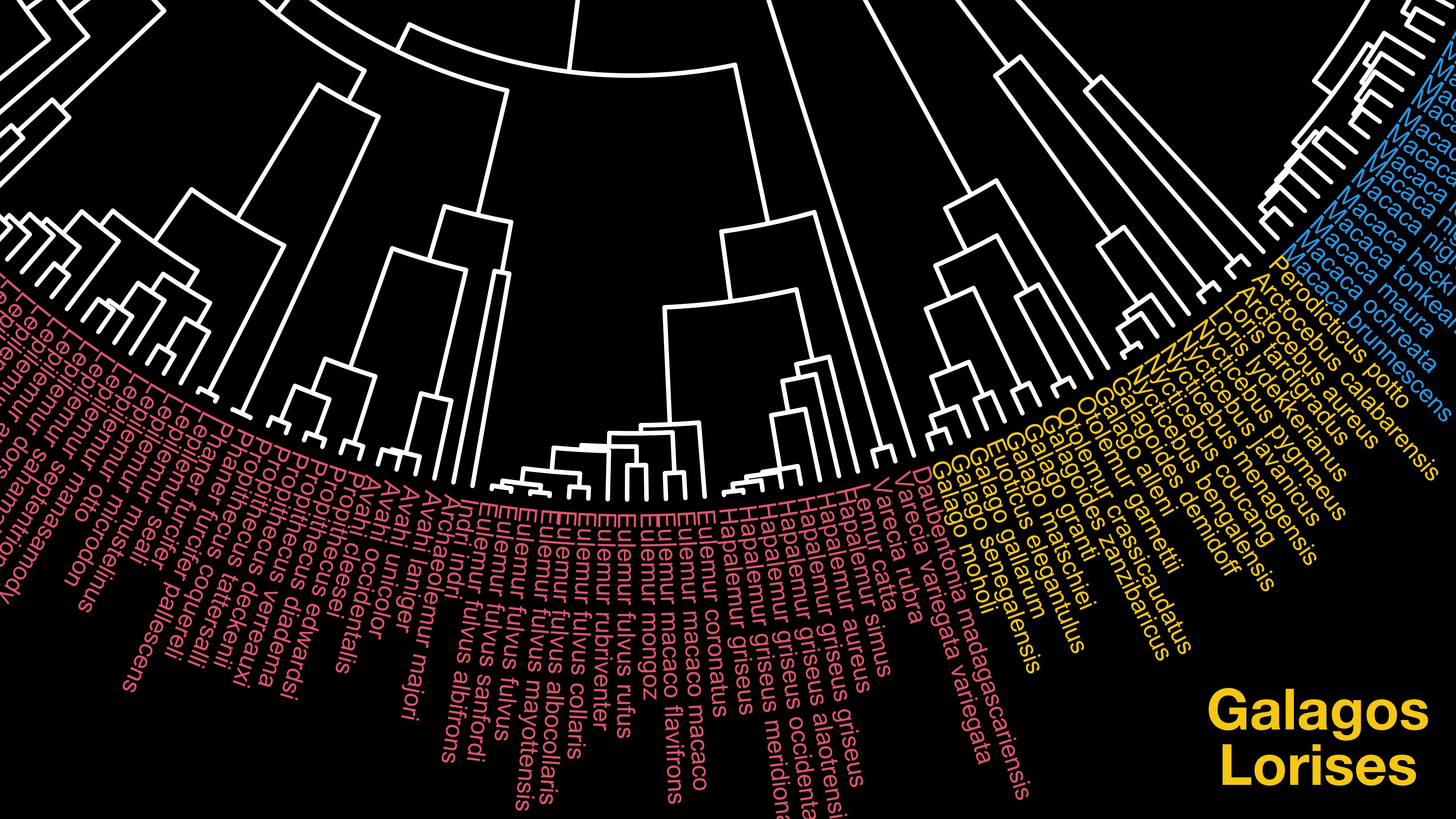


# Tarsiers

# Lemurs etc



# Galagos Lorises





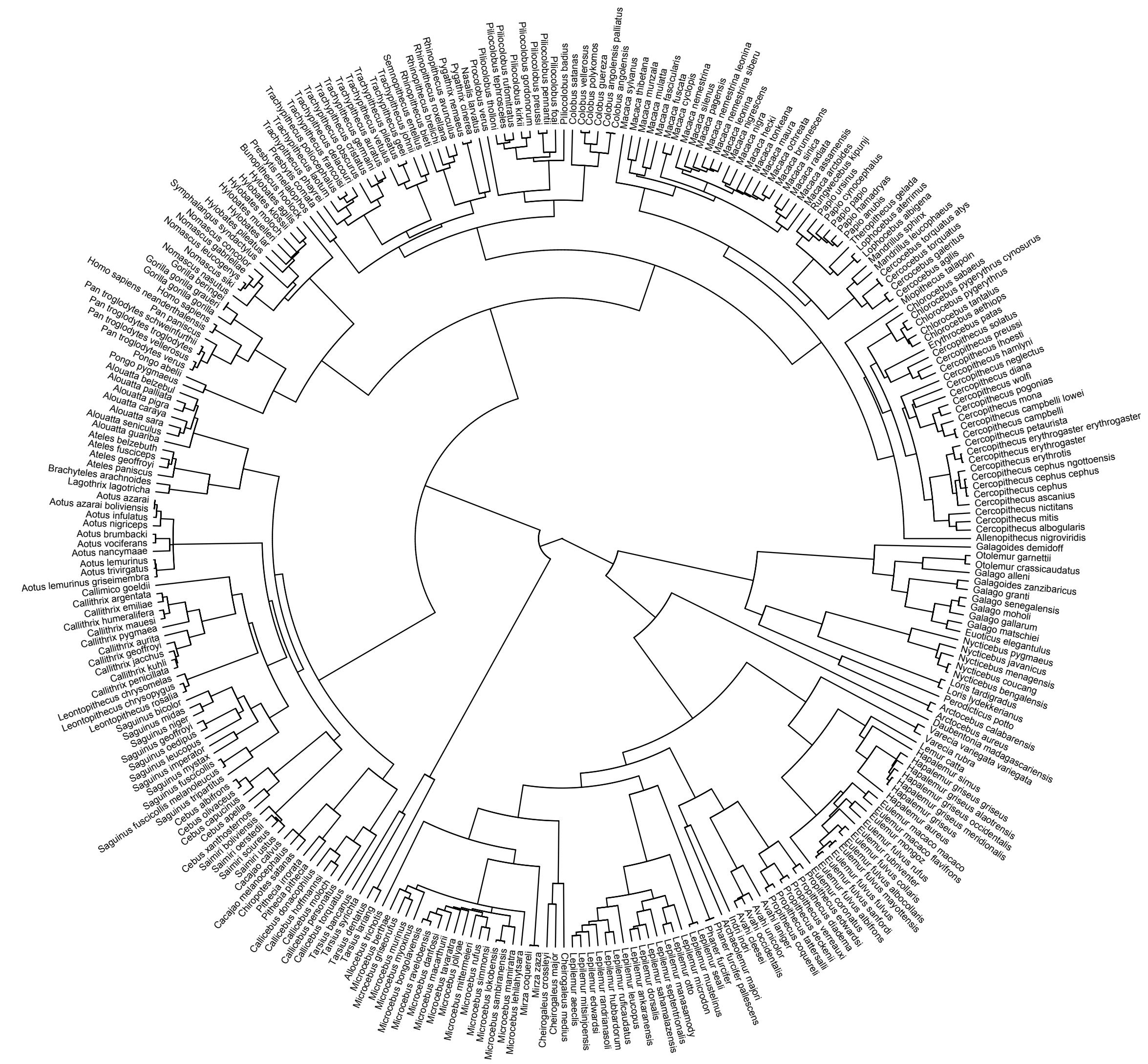
# Phylogenetic regression

```
data(Primates301)
```

# Life history traits

# Mass g, brain cc, group size

Much missing data,  
measurement error, unobserved  
confounding

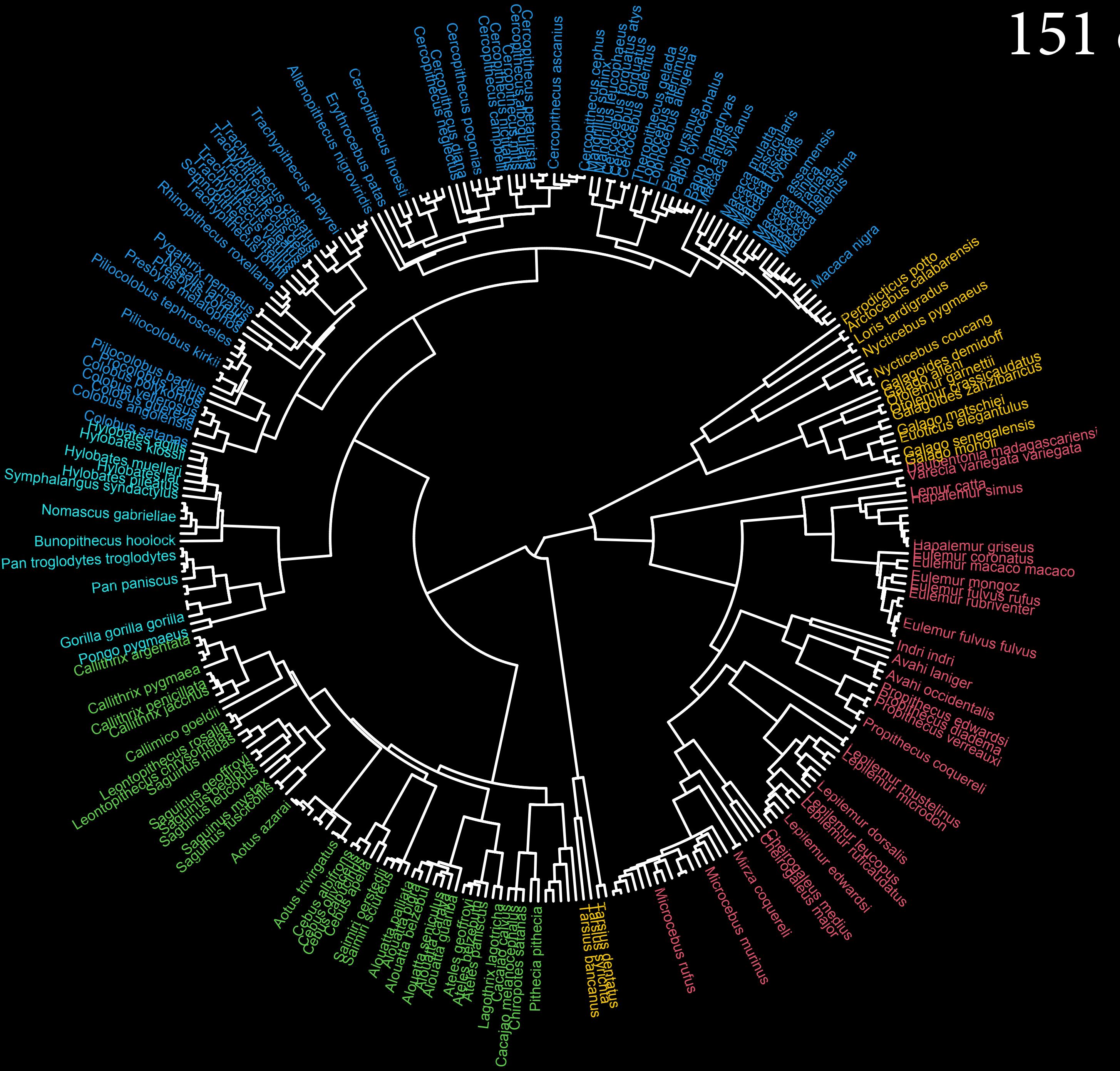


# 301 species

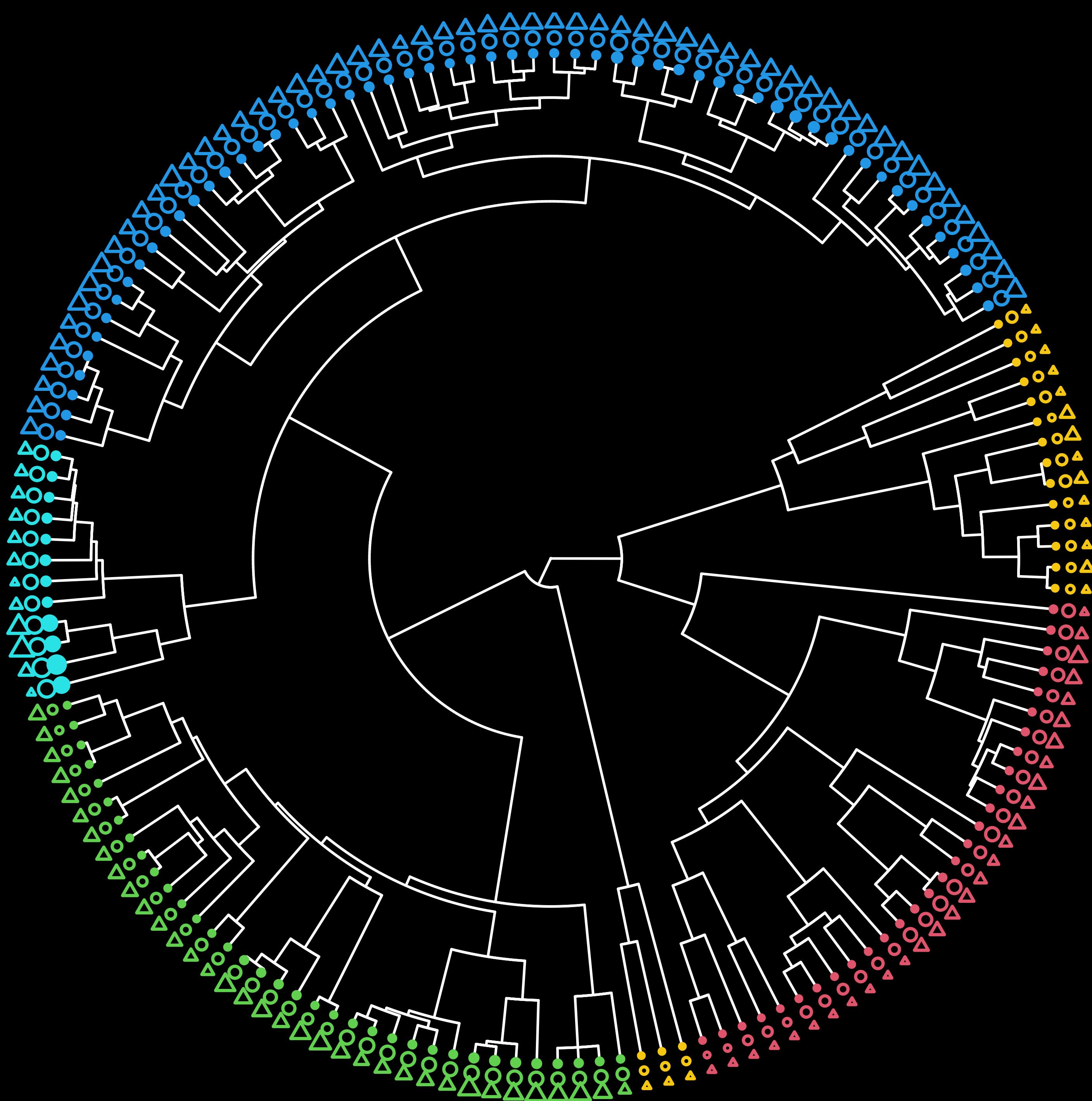


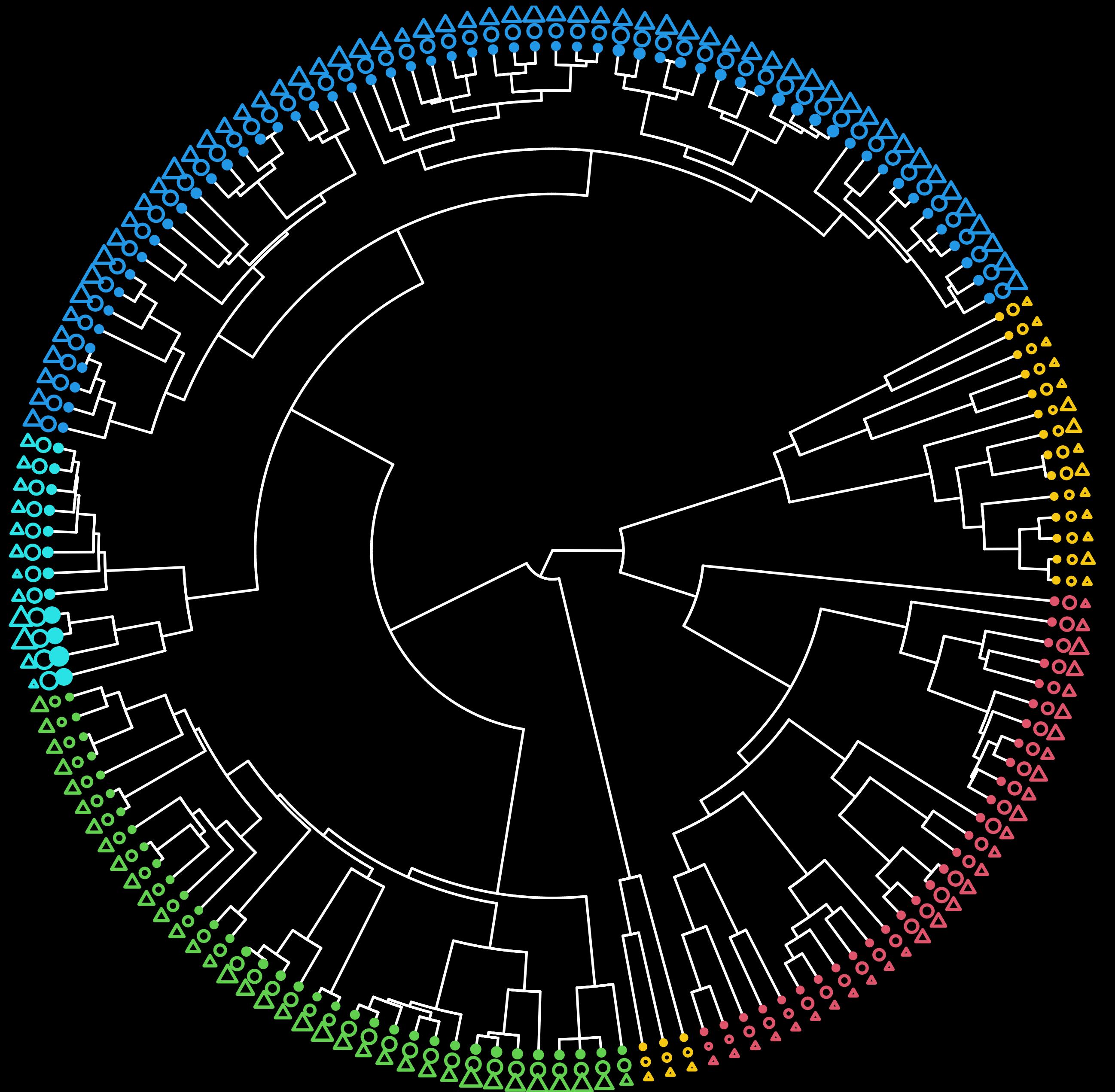
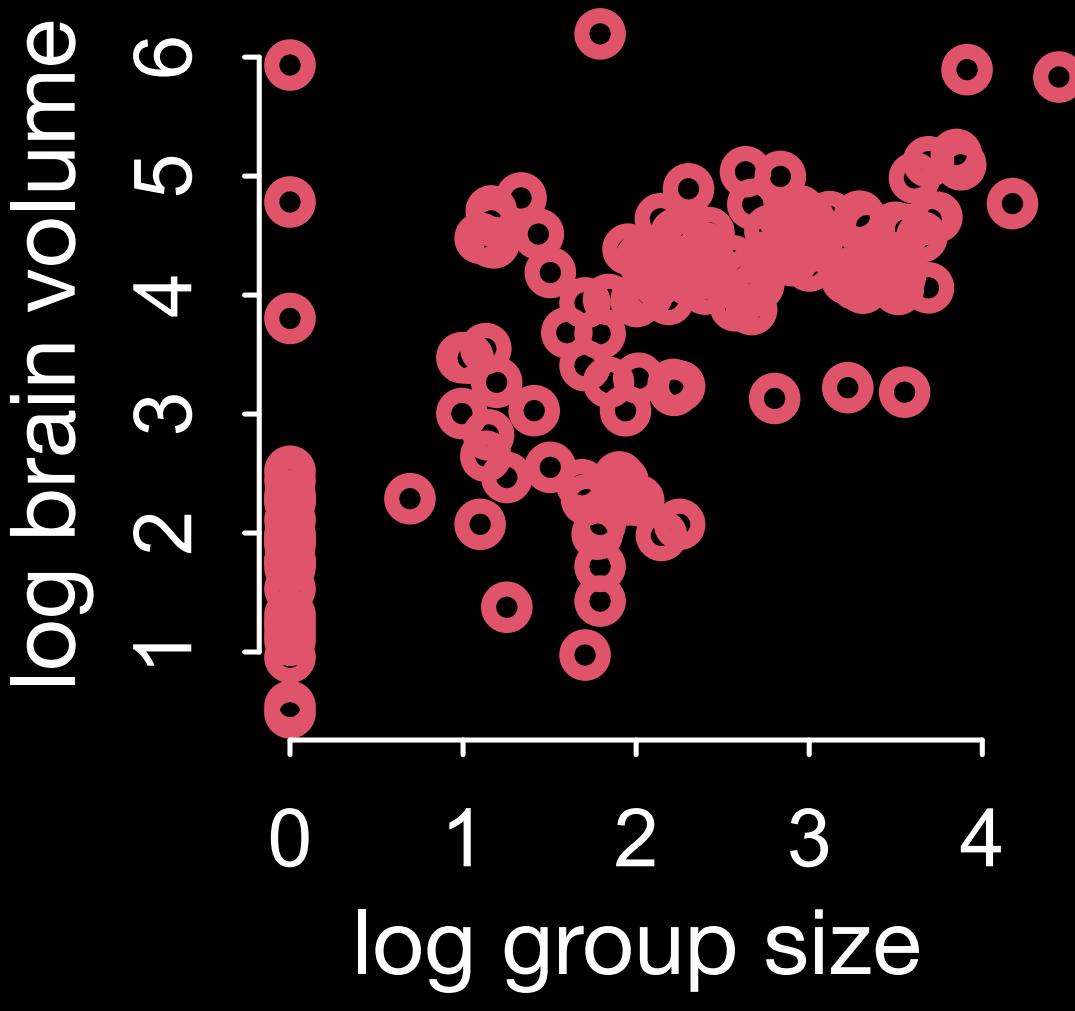
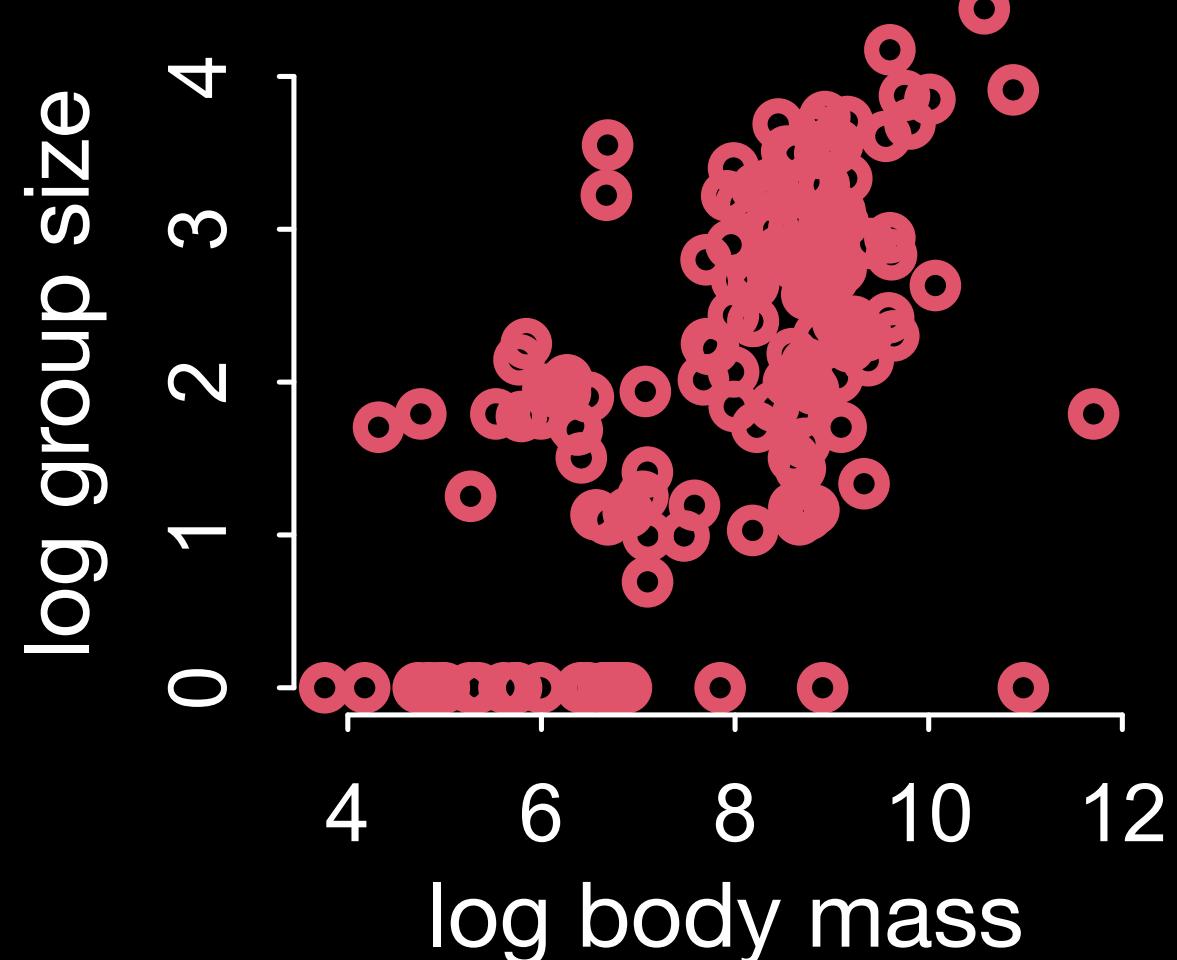
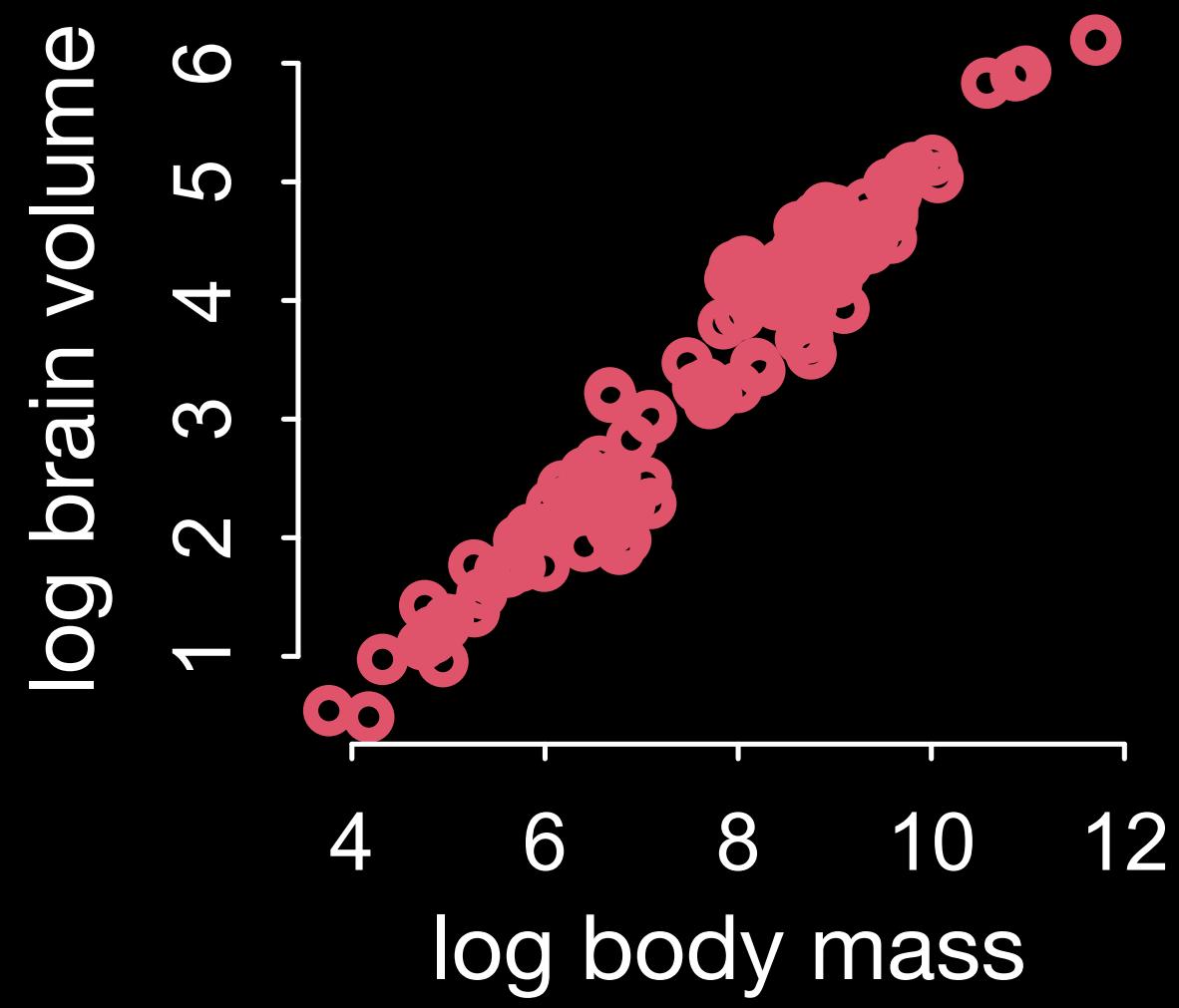
# 301 species

# 151 complete cases



- Brain vol (cc)
- Mass (log kg)
- △ Group size (log)





# Causal Salad in Evolutionary Ecology

Phylogenetic comparative methods  
dominated by causal salad

*Causal salad*: Tossing factors into  
regression and interpreting every  
coefficient as causal

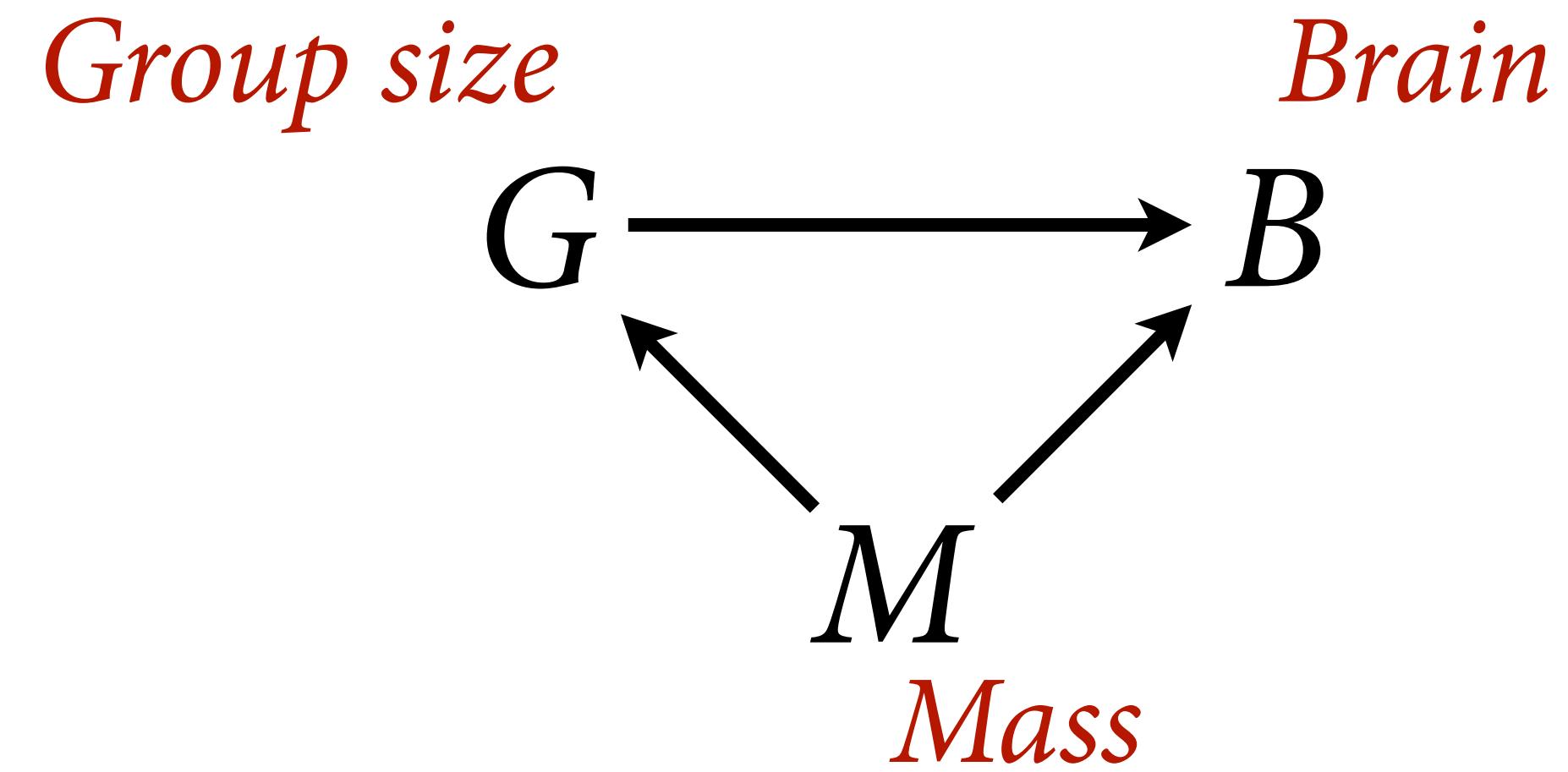
“Controlling for phylogeny”: Required  
but mindless

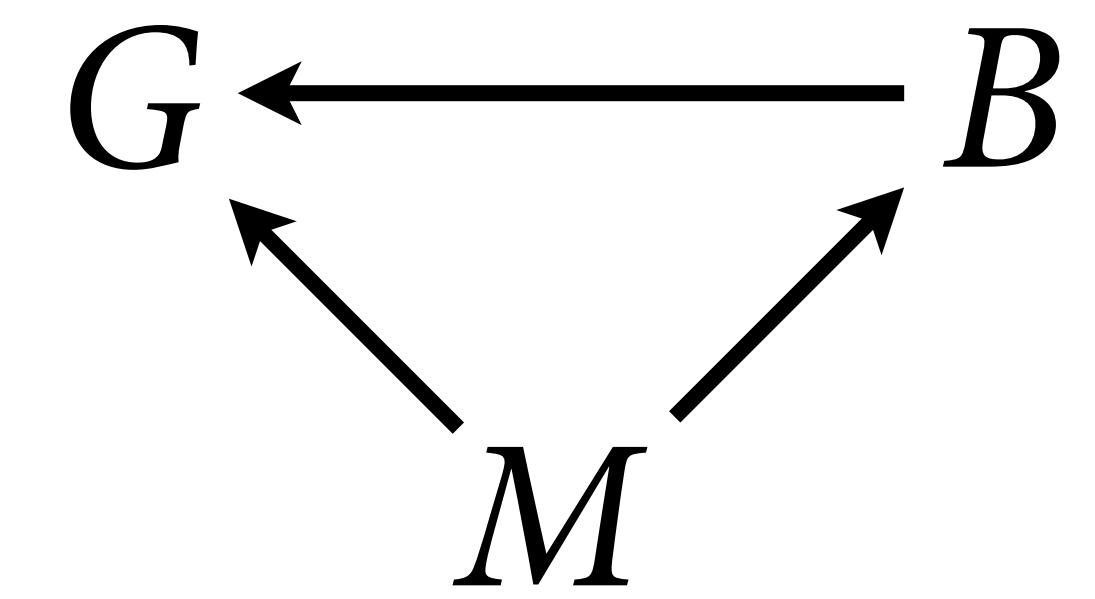
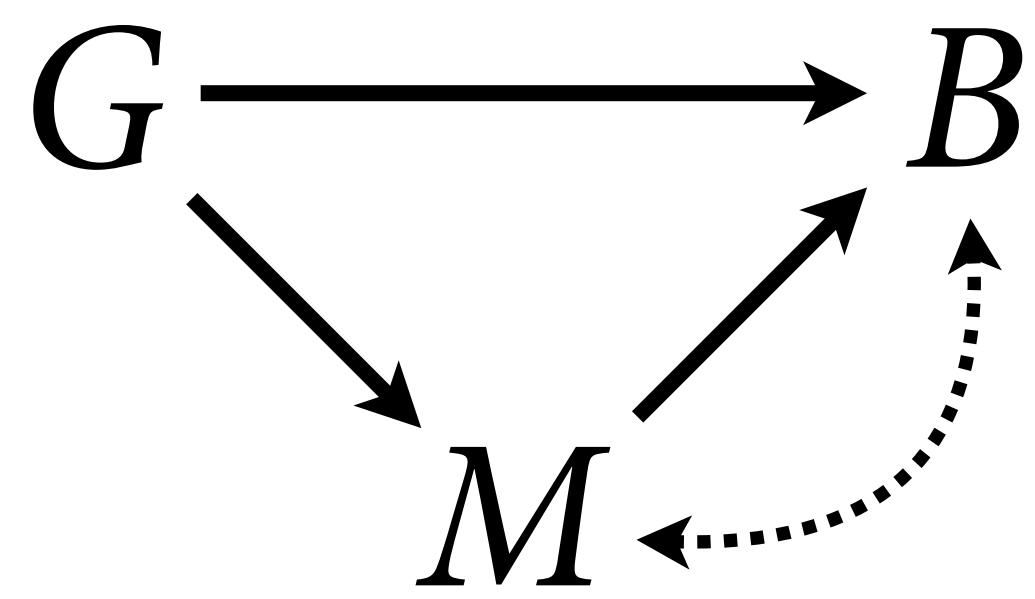
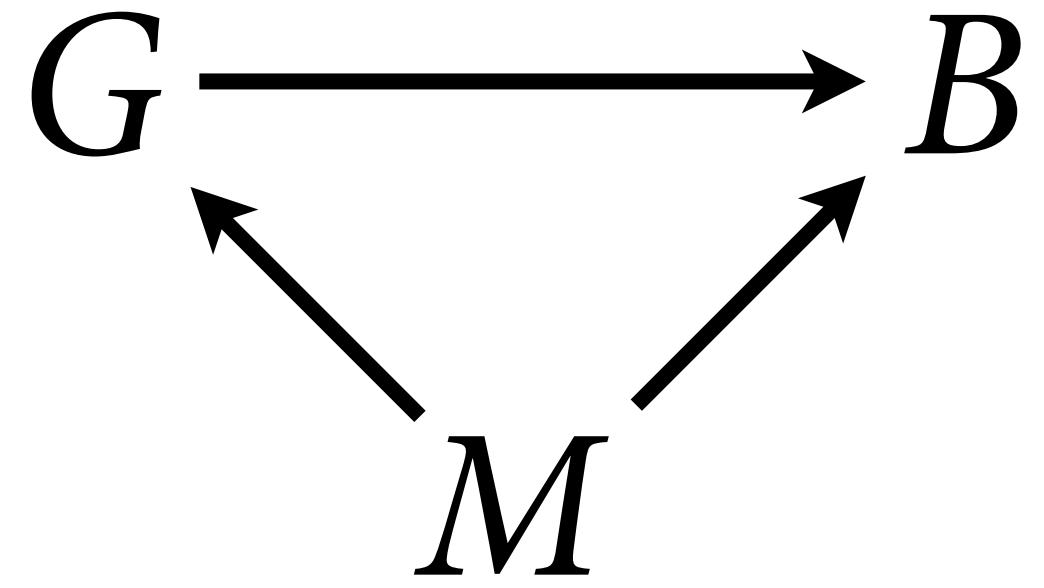
Regression + phylogeny still requires  
causal model



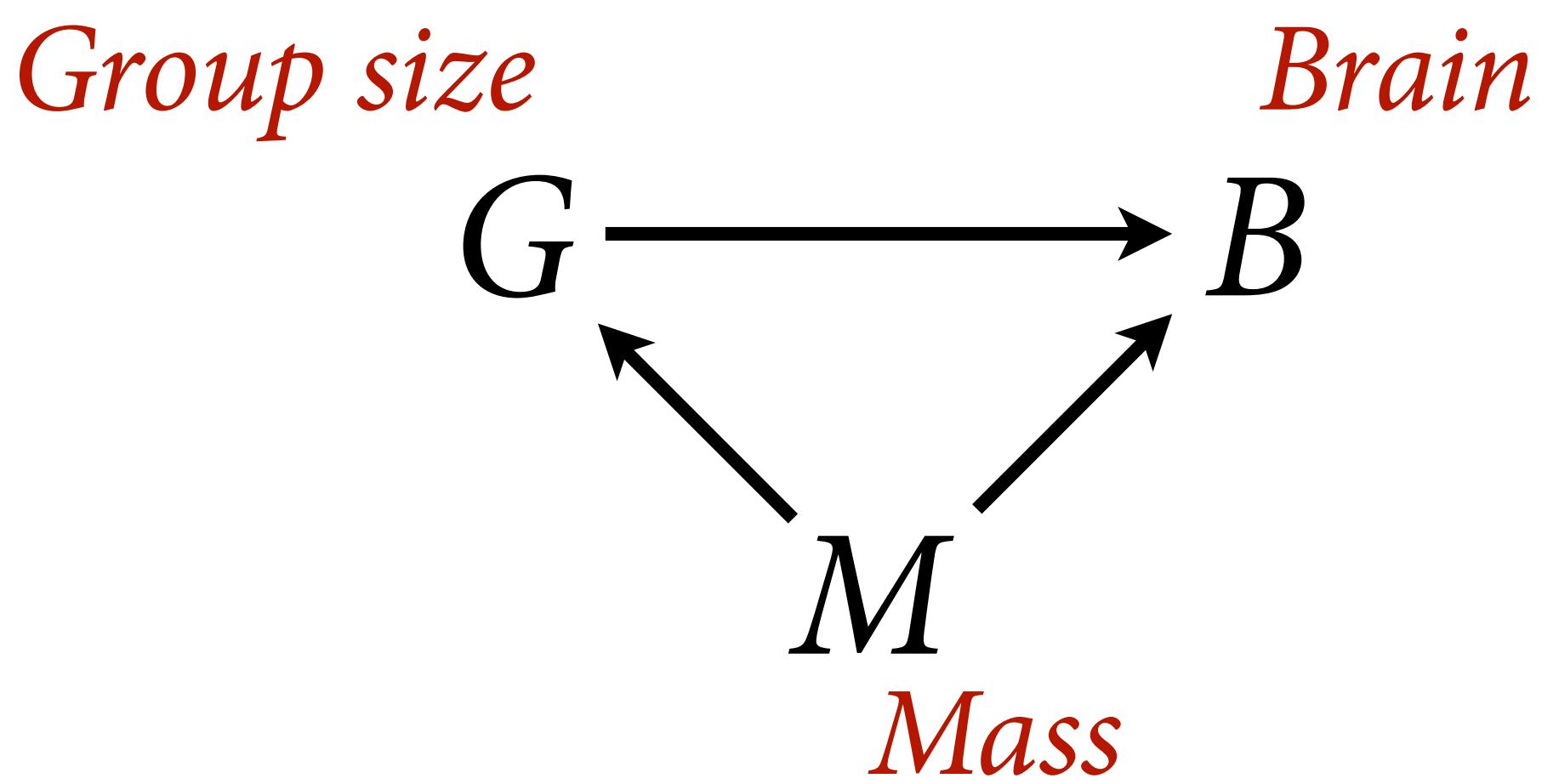
Illustration by Julia Suits

# Social brain hypothesis

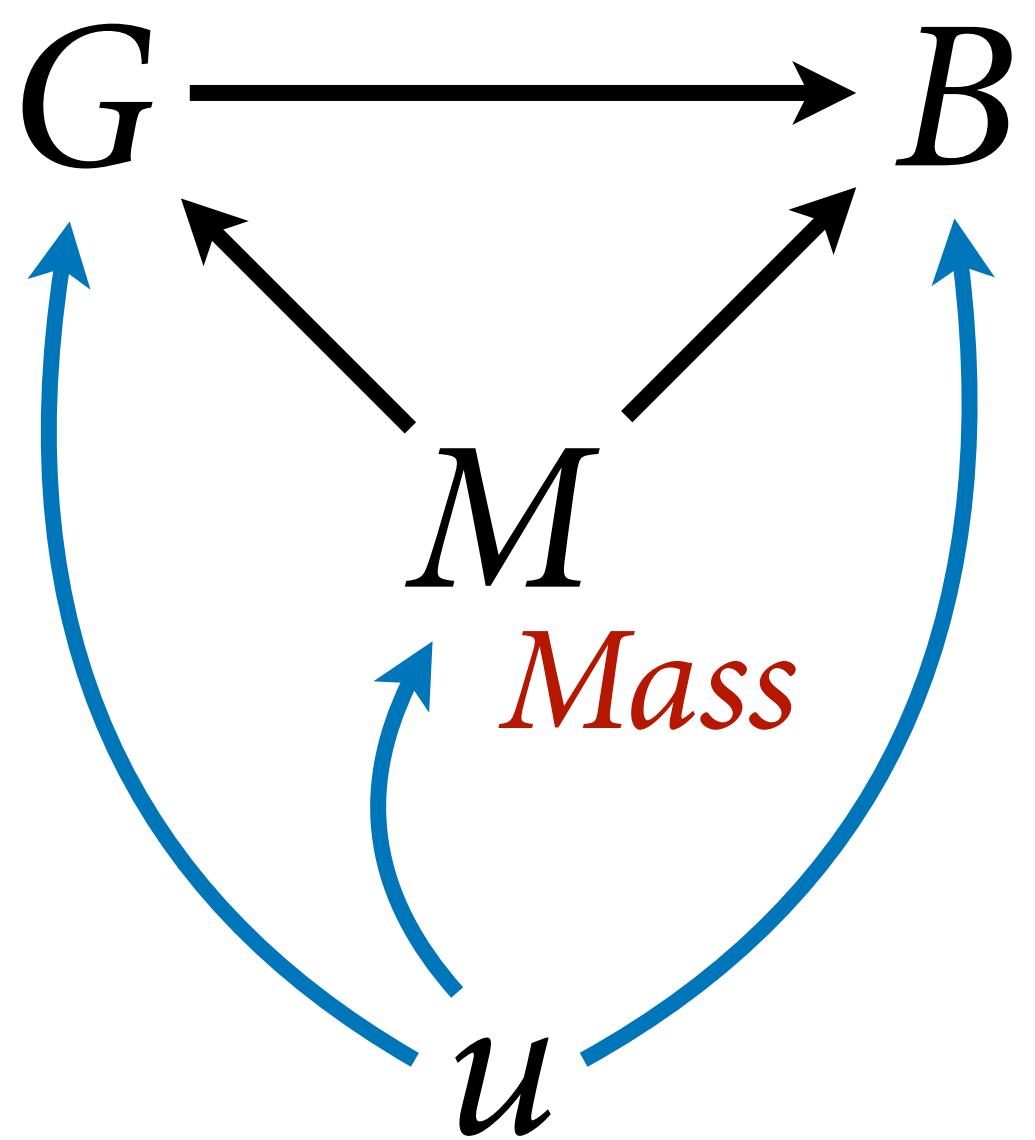


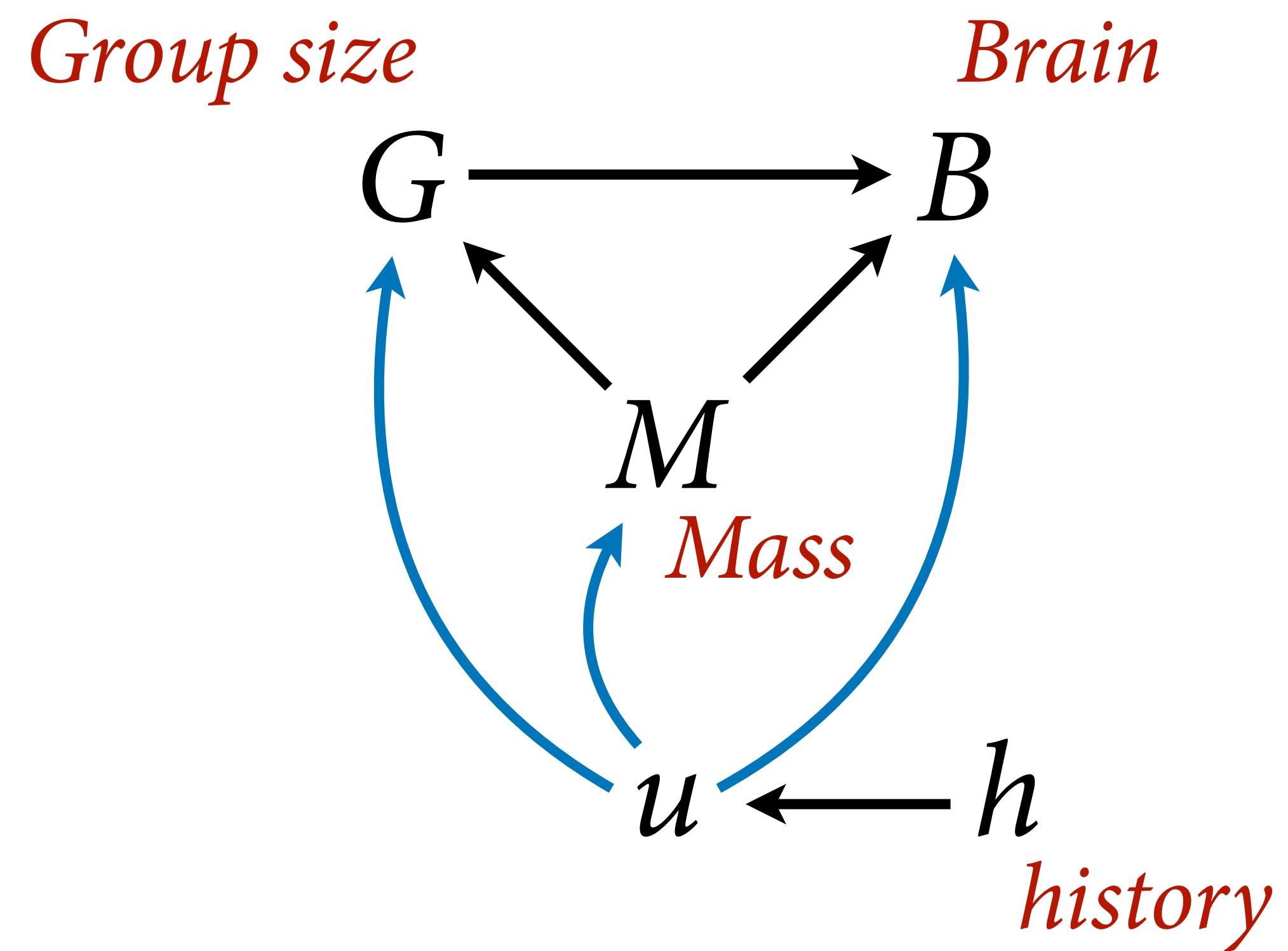


*No interpretation without causal representation*



*Group size*      *Brain*

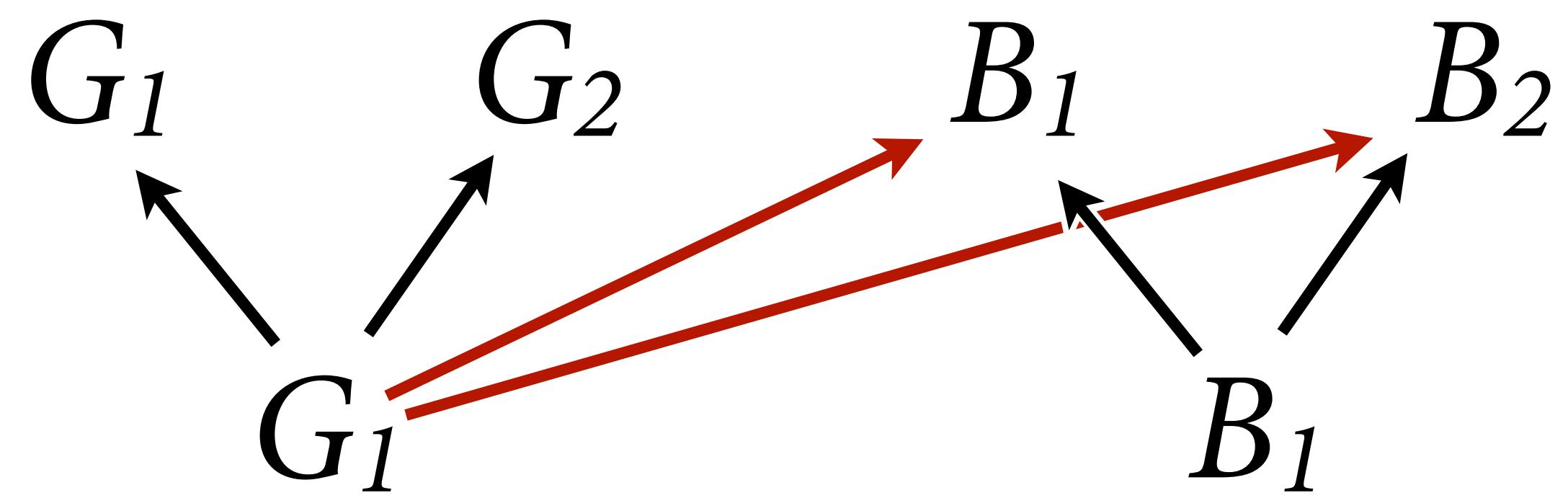


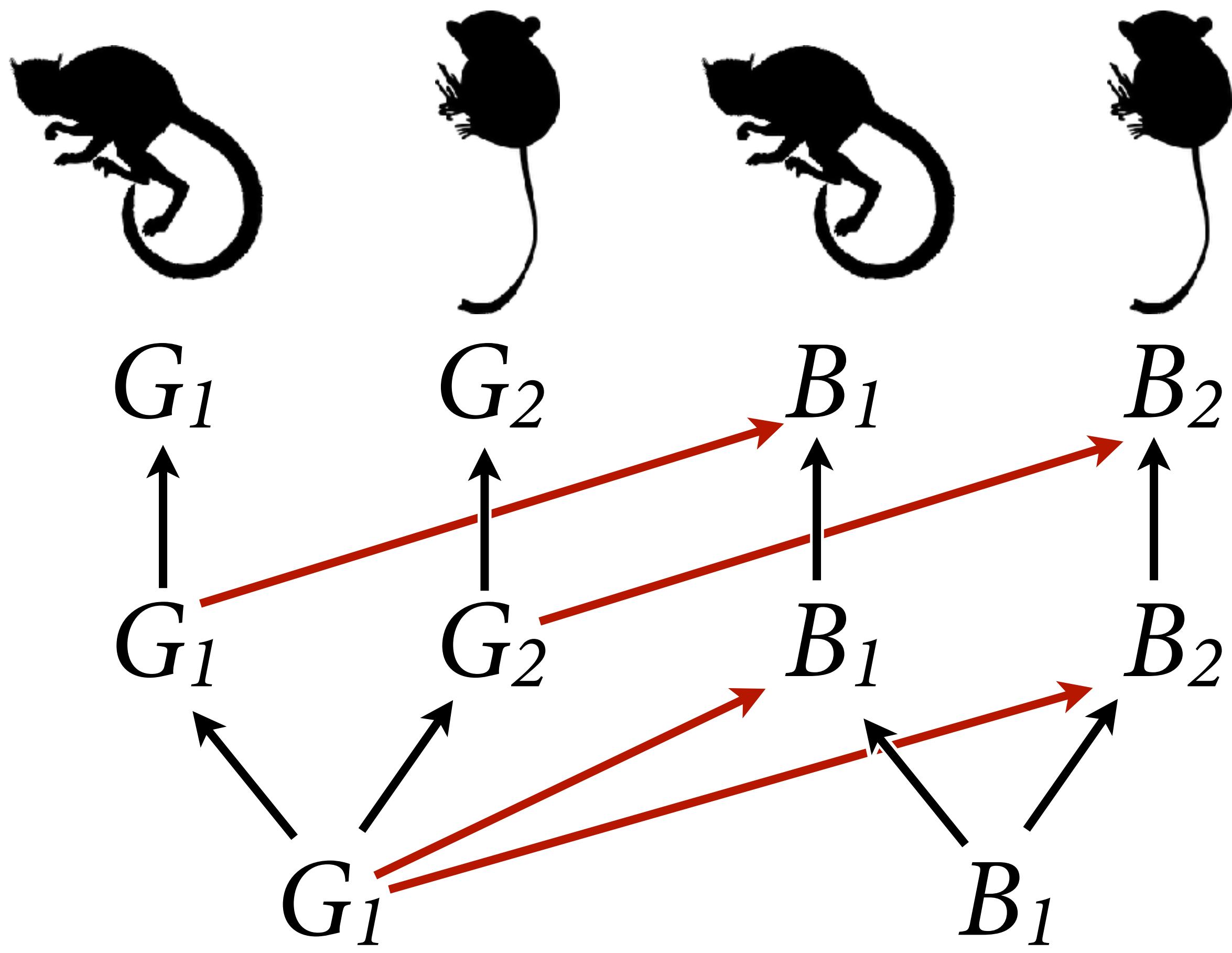


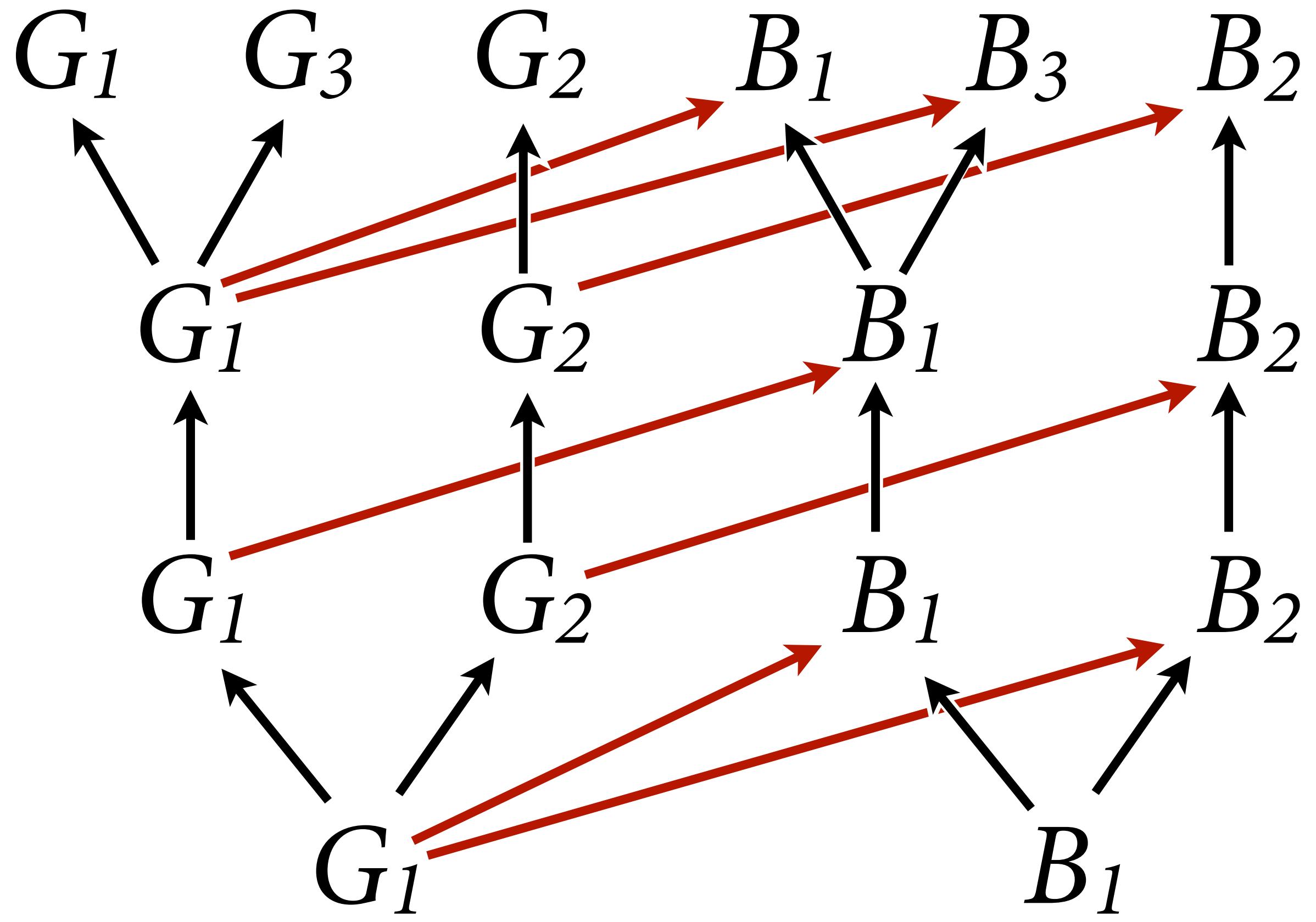
$G_1$

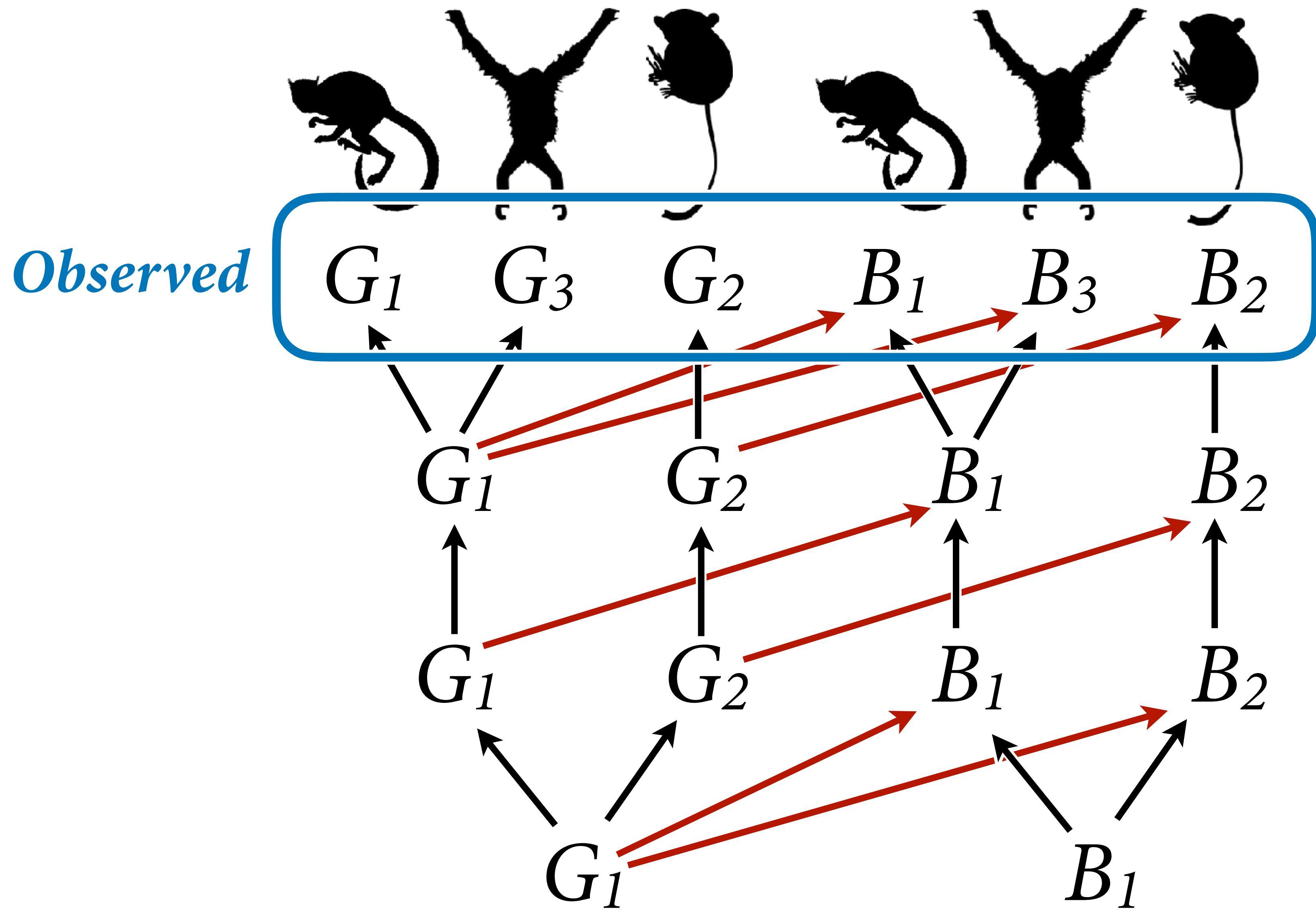


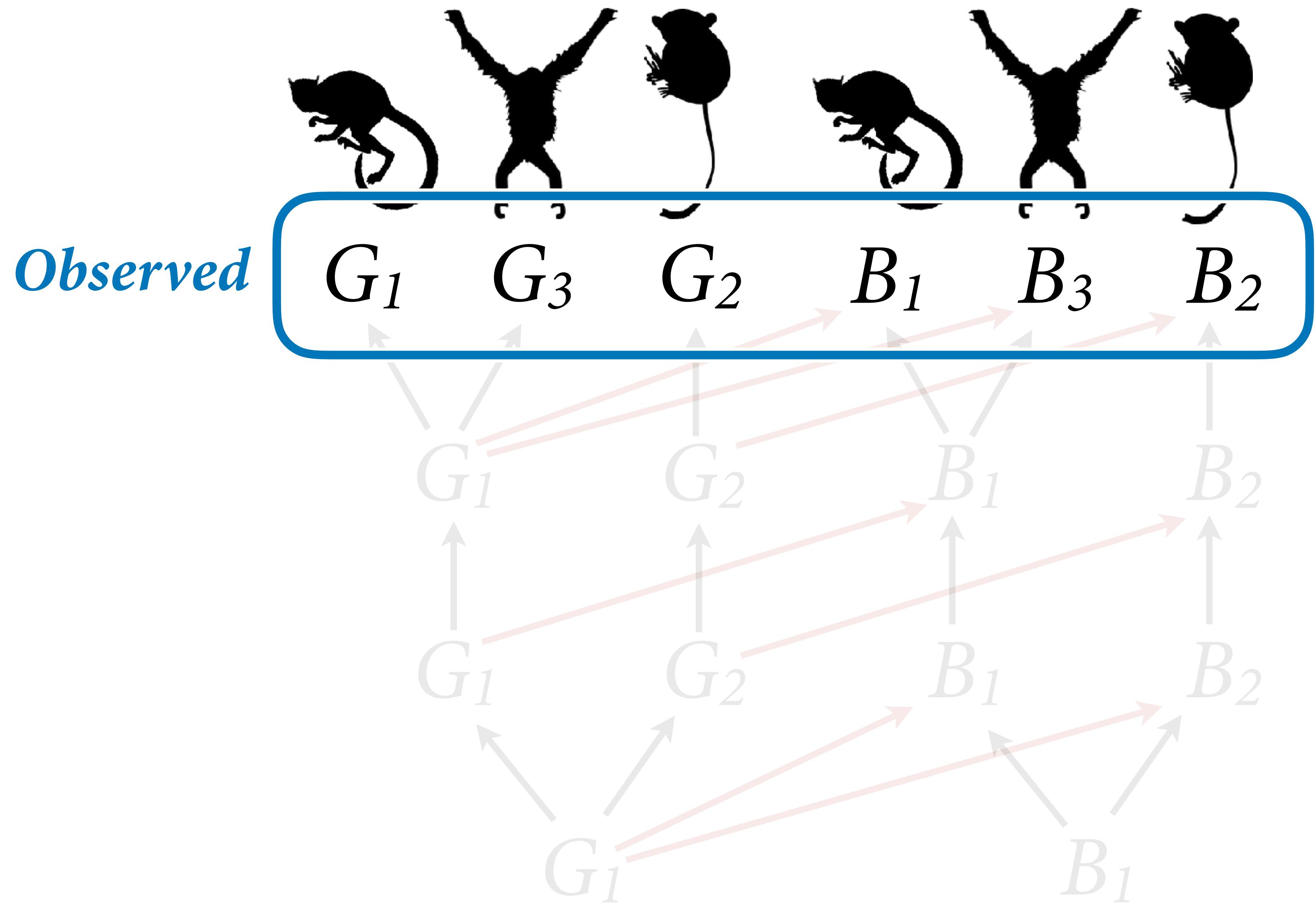
$B_1$







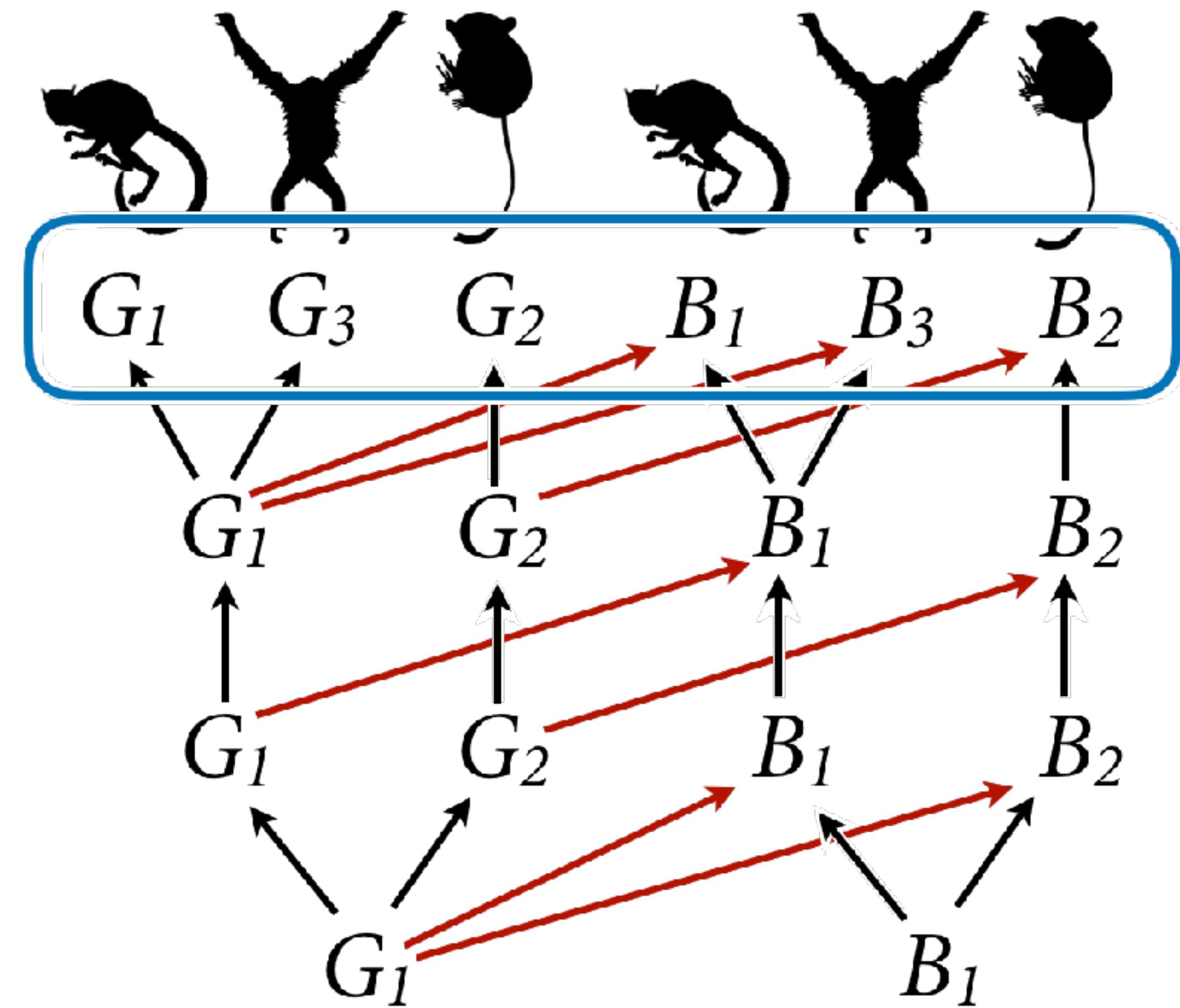




# Phylogenetic regression

Two conjoint problems

- (1) What is the history (phylogeny)?
- (2) How to use it to model causes?



# Phylogenetic regression

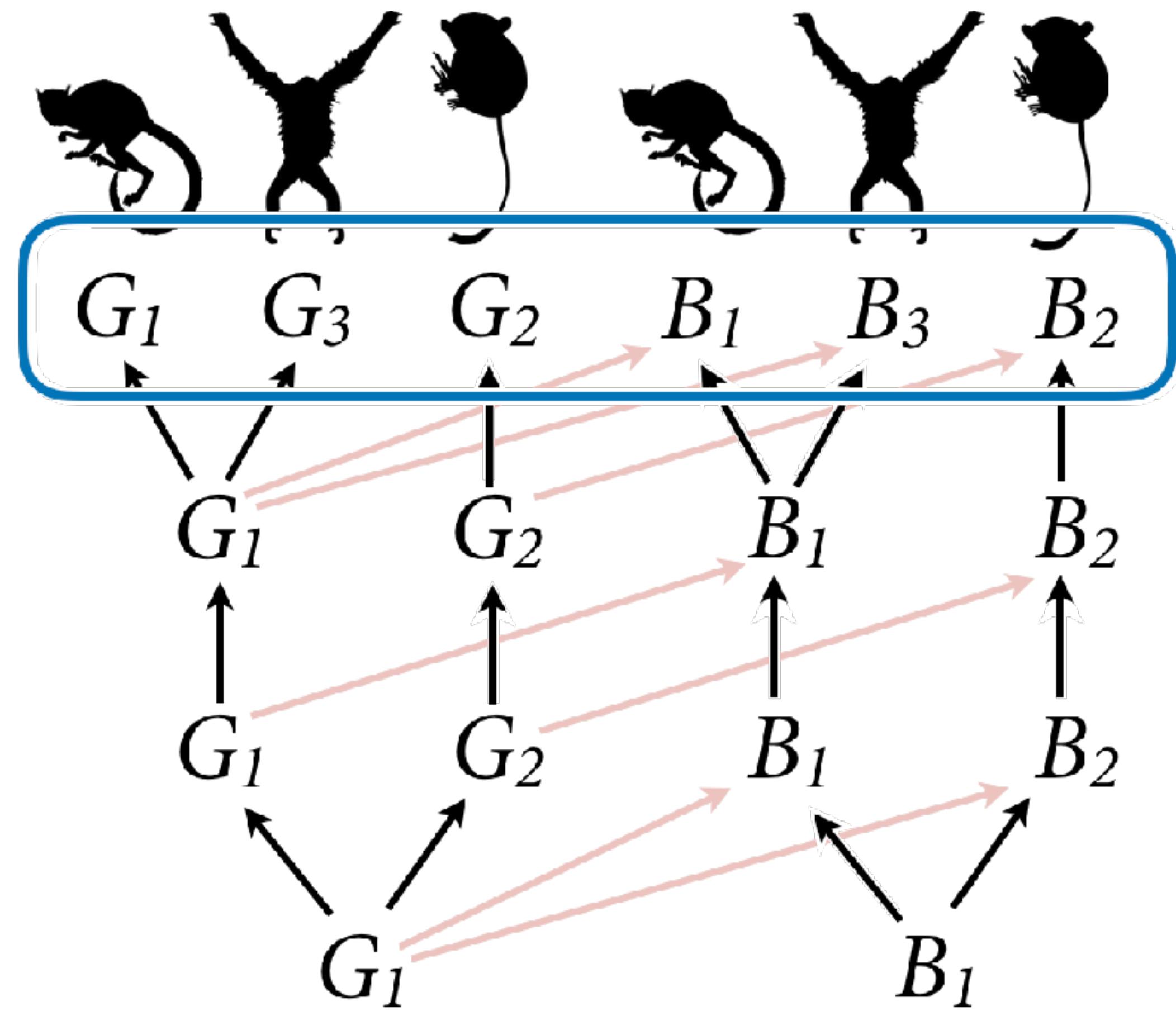
(1) What is the history (phylogeny)?

Gotten much better with genomics

Problems: Huge **uncertainty** in best case,  
process **not stationary**, no one phylogeny  
correct for **all traits**

Cultural/linguistic phylogenies remain  
**incredible**

Basic truth: Phylogenies do not exist



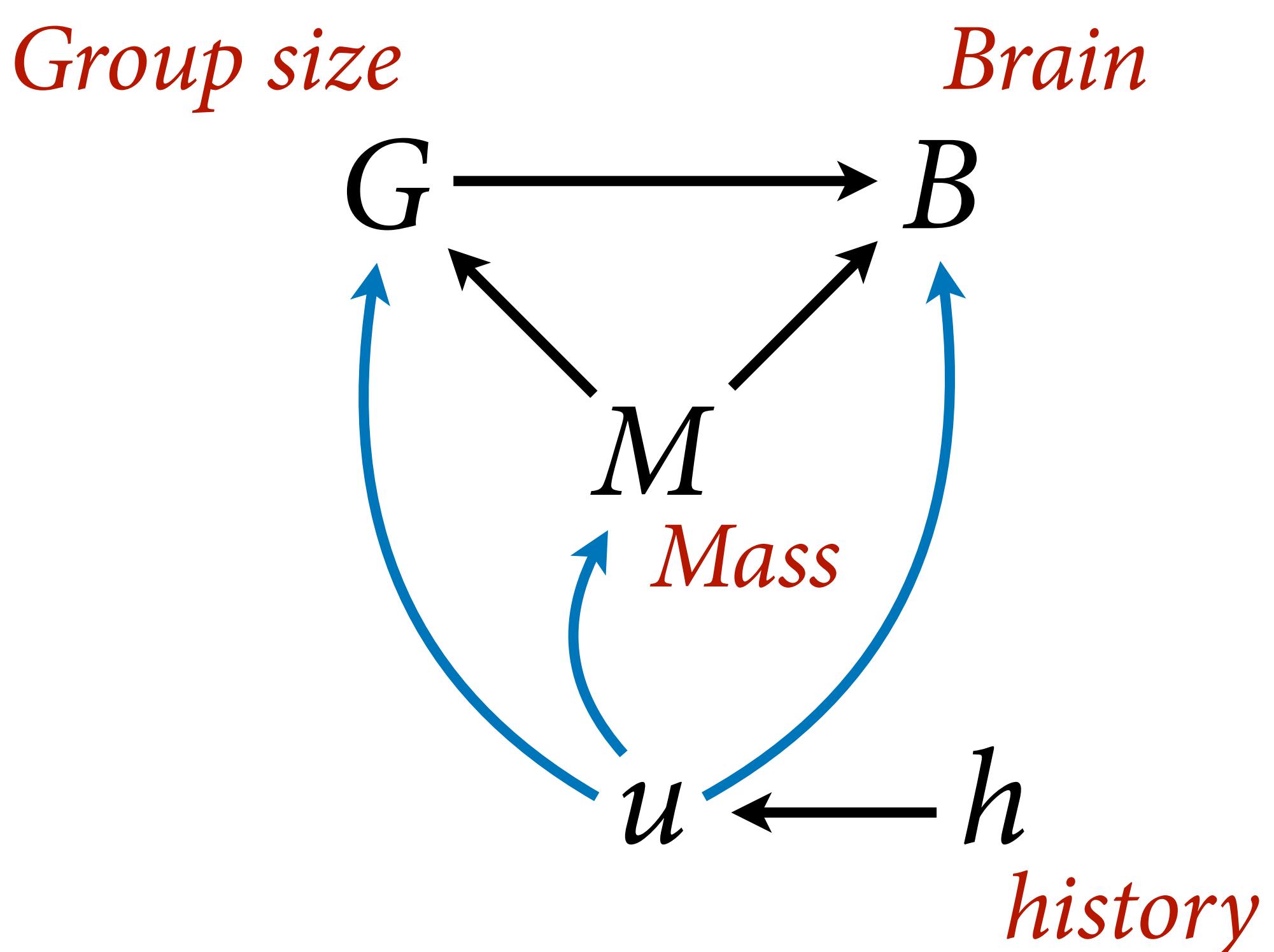
# Phylogenetic regression

(2) How to use it to model causes?

Suppose we have a phylogeny.  
Now what?

No universally correct approach

Default approach is a Gaussian  
process regression



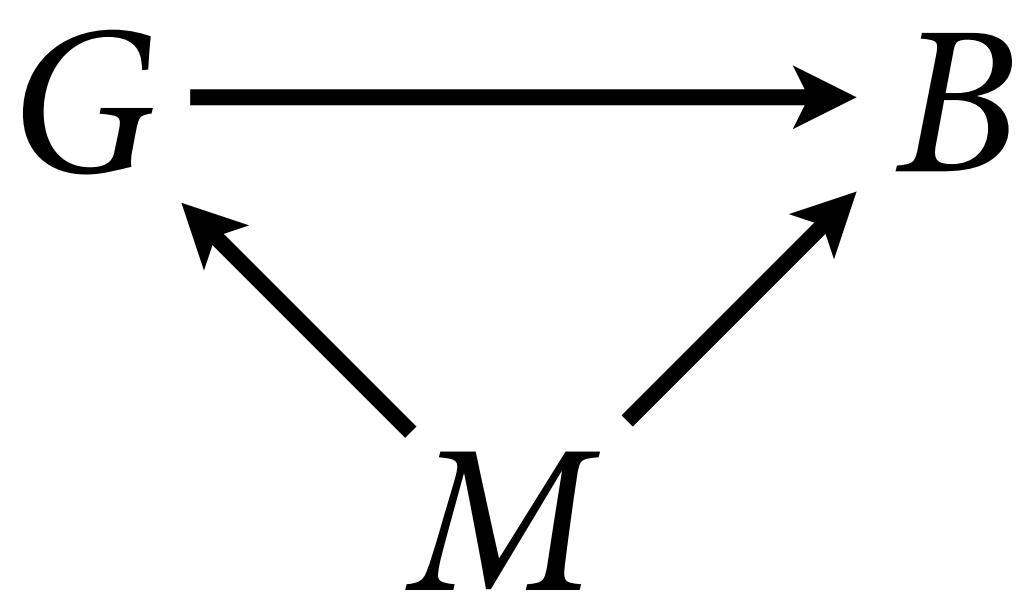
$$B_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

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



$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

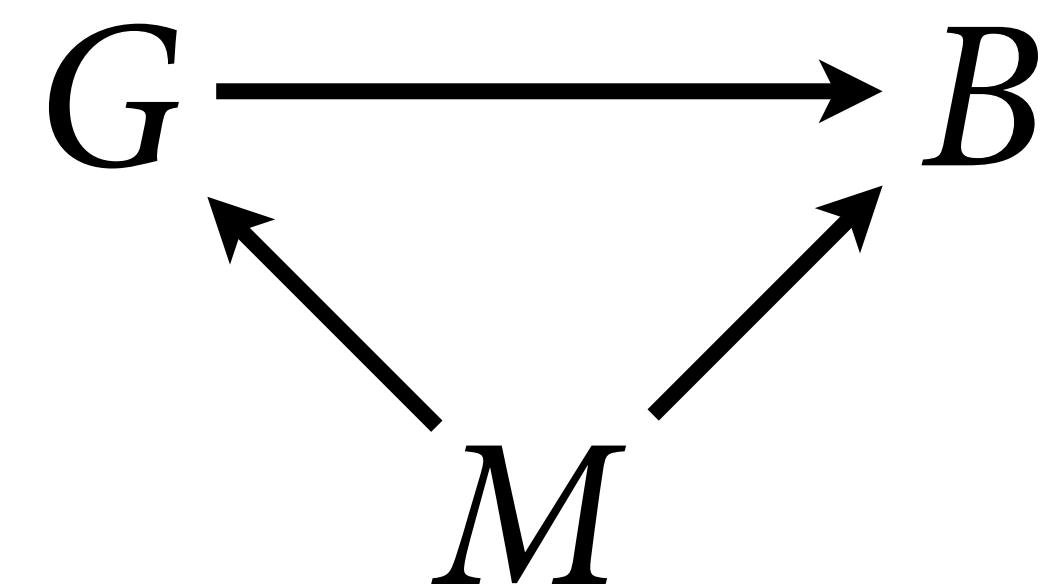
$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

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

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

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



$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

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

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

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

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

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

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

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

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\mathbf{K} = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$$B_i \sim \text{Normal}(\mu_i, \sigma)$$

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

```

dat_list <- list(
  N_spp = nrow(dstan),
  M = standardize(log(dstan$body)),
  B = standardize(log(dstan$brain)),
  G = standardize(log(dstan$group_size)),
  Imat = diag(nrow(dstan)) )

# classical regression form
mBMG0 <- ulam(
  alist(
    B ~ normal( mu , sigma ),
    mu <- a + bM*M + bG*G,
    a ~ normal( 0 , 1 ),
    c(bM,bG) ~ normal( 0 , 0.5 ),
    sigma ~ exponential( 1 )
  ), data=dat_list , chains=4 , cores=4 )

```

```

# multivariate form
mBMG <- ulam(
  alist(
    B ~ multi_normal( mu , K ),
    mu <- a + bM*M + bG*G,
    matrix[N_spp,N_spp]:K <- Imat*(sigma^2),
    a ~ normal( 0 , 1 ),
    c(bM,bG) ~ normal( 0 , 0.5 ),
    sigma ~ exponential( 1 )
  ), data=dat_list , chains=4 , cores=4 )

```

$$B_i \sim \text{Normal}(\mu_i, \sigma)$$

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

```
dat_list <- list(
  N_spp = nrow(dstan),
  M = standardize(log(dstan$body)),
  B = standardize(log(dstan$brain)),
  G = standardize(log(dstan$group_size)),
  Iimat = diag(nrow(dstan)) )
```

```
# classical regression form
mBMG0 <- ulam(
  alist(
    B ~ normal( mu , sigma ),
    mu <- a + bM*M + bG*G
```

> **precis( mBMG0 )**

|       | mean | sd   | 5.5%  | 94.5% | n_eff | Rhat4 |
|-------|------|------|-------|-------|-------|-------|
| a     | 0.00 | 0.02 | -0.03 | 0.03  | 1740  | 1     |
| bG    | 0.12 | 0.02 | 0.09  | 0.16  | 1491  | 1     |
| bM    | 0.89 | 0.02 | 0.86  | 0.93  | 1439  | 1     |
| sigma | 0.22 | 0.01 | 0.20  | 0.24  | 1706  | 1     |

# multivariate form

```
mBMG <- ulam(
  alist(
    B ~ multi_normal( mu , K ),
    mu <- a + bM*M + bG*G,
    matrix[N spp N spp]•K <- Tmat*(sigma^2),
```

> **precis( mBMG )**

|       | mean | sd   | 5.5%  | 94.5% | n_eff | Rhat4 |
|-------|------|------|-------|-------|-------|-------|
| a     | 0.00 | 0.02 | -0.03 | 0.03  | 1880  | 1     |
| bG    | 0.12 | 0.02 | 0.09  | 0.16  | 1384  | 1     |
| bM    | 0.89 | 0.02 | 0.86  | 0.93  | 1395  | 1     |
| sigma | 0.22 | 0.01 | 0.20  | 0.24  | 1433  | 1     |

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

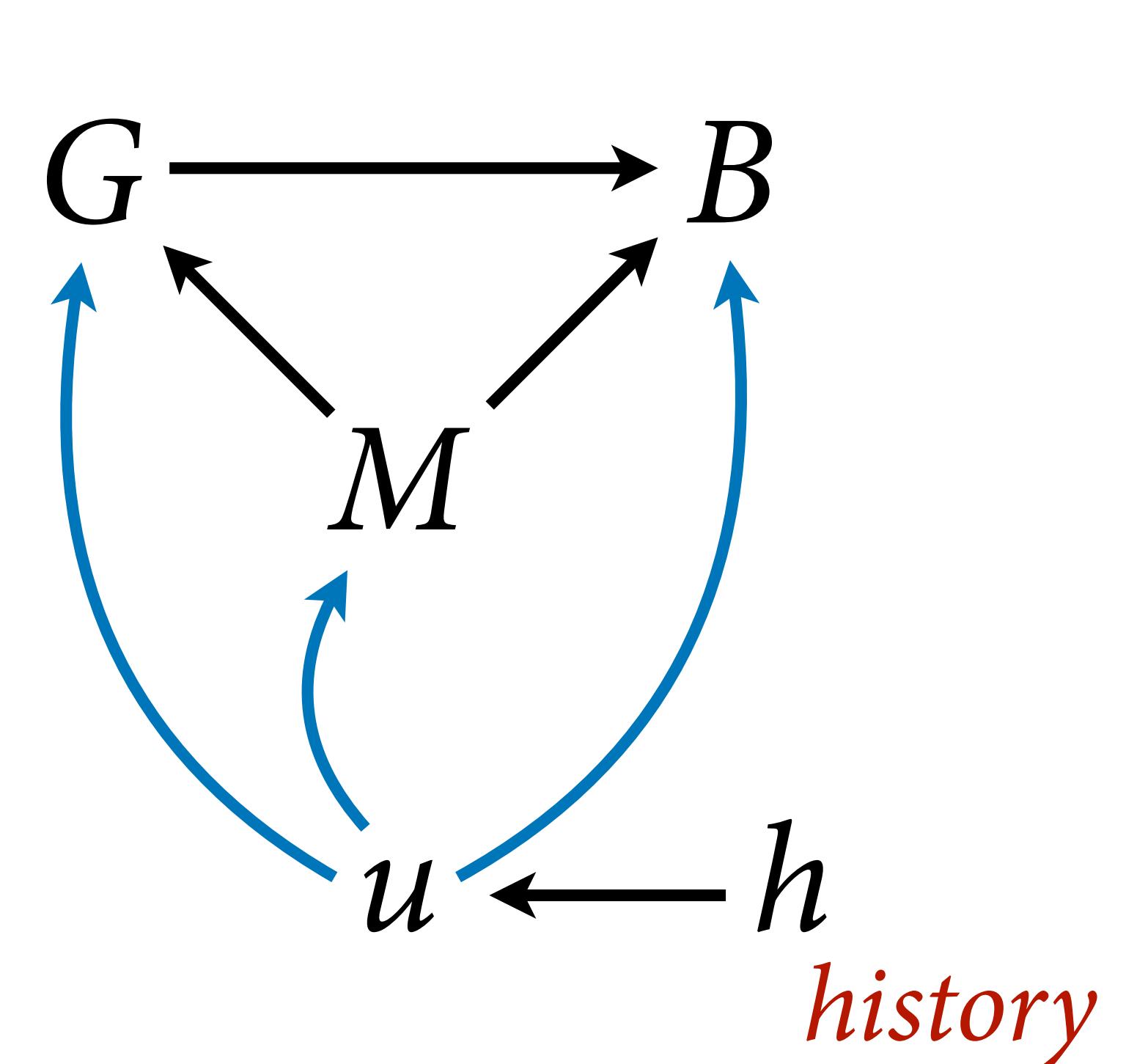
$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i + u_i$$

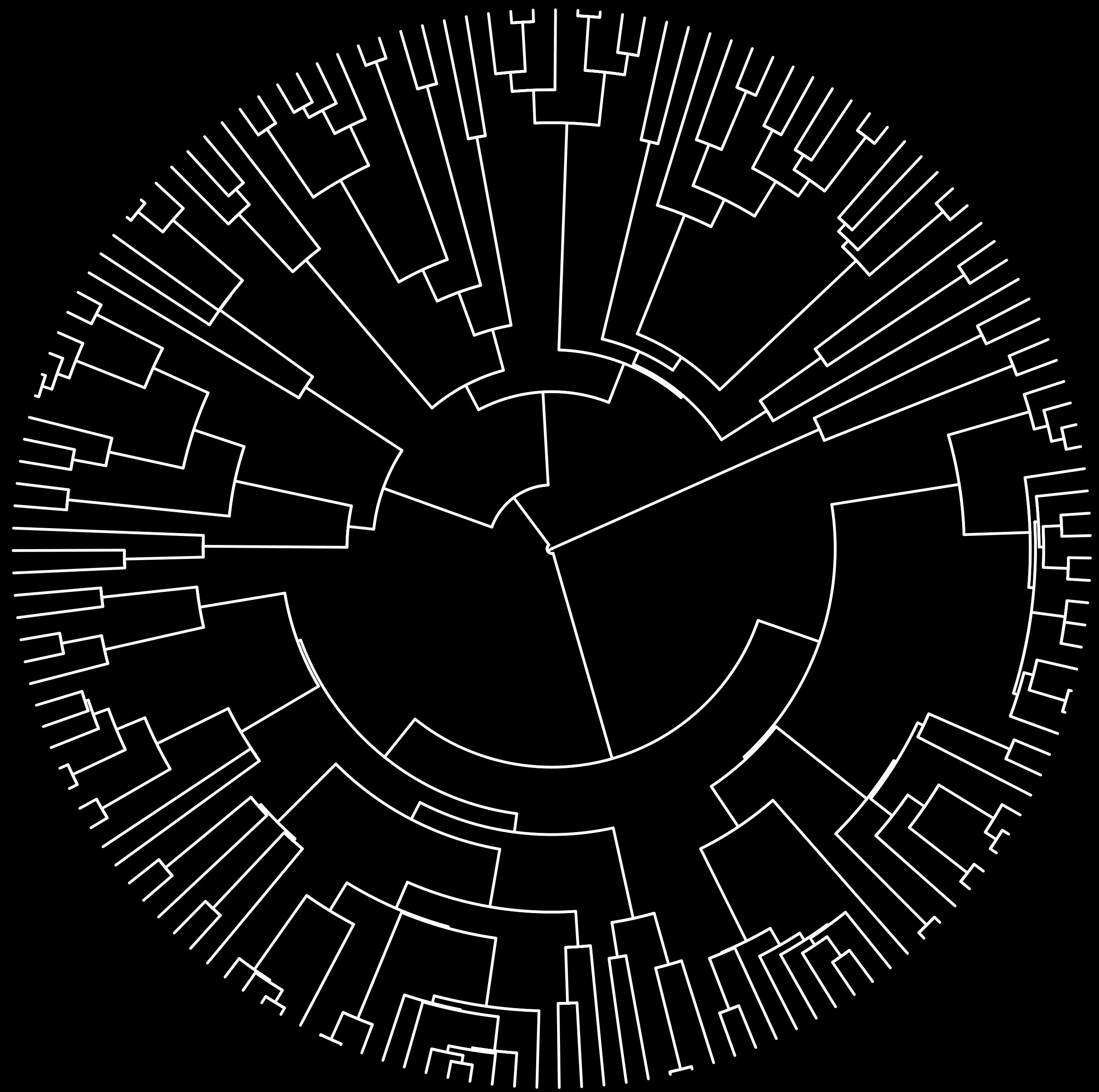
$$\mathbf{K} = \mathbf{I}\sigma^2$$

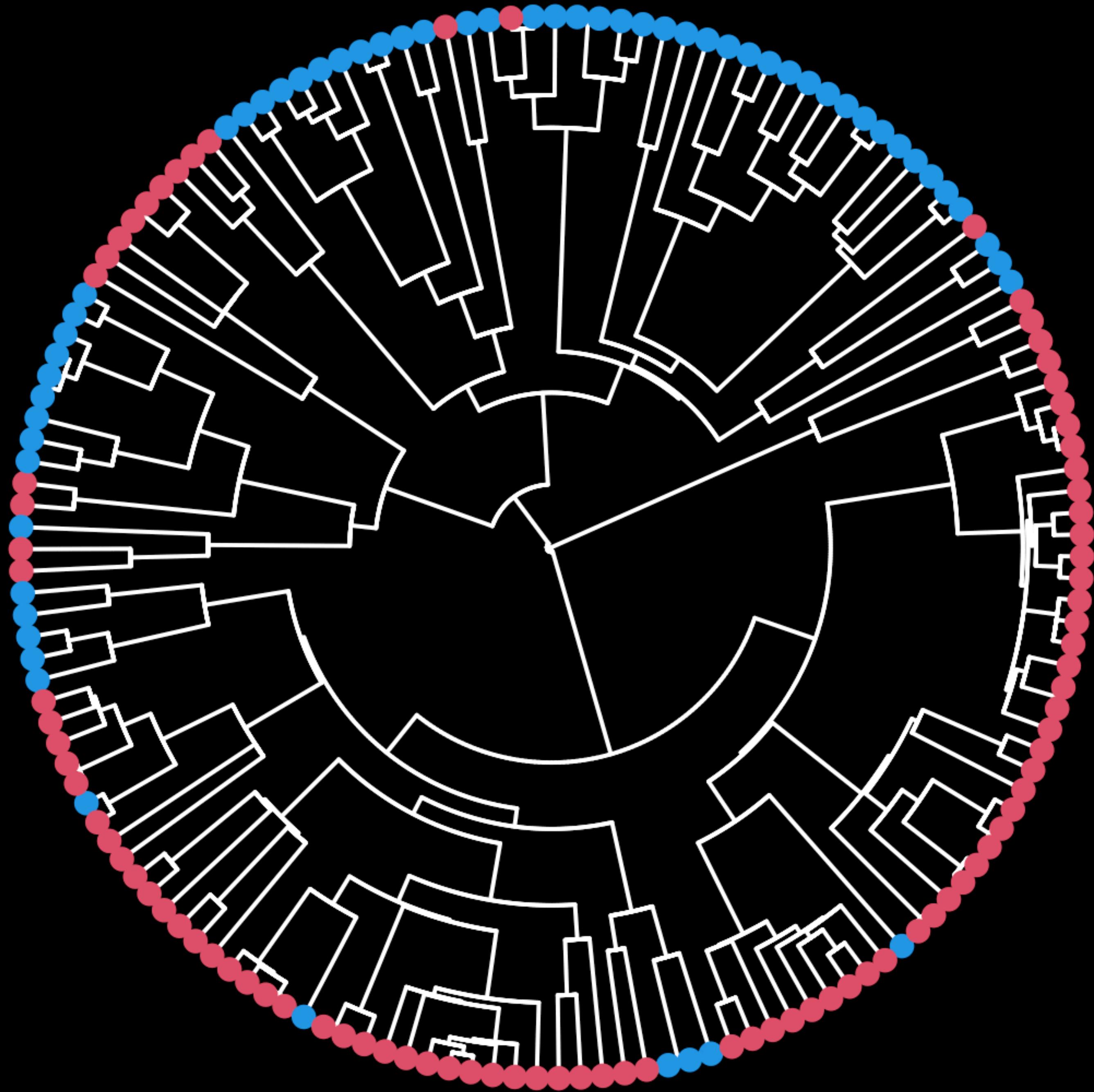
$$\alpha \sim \text{Normal}(0,1)$$

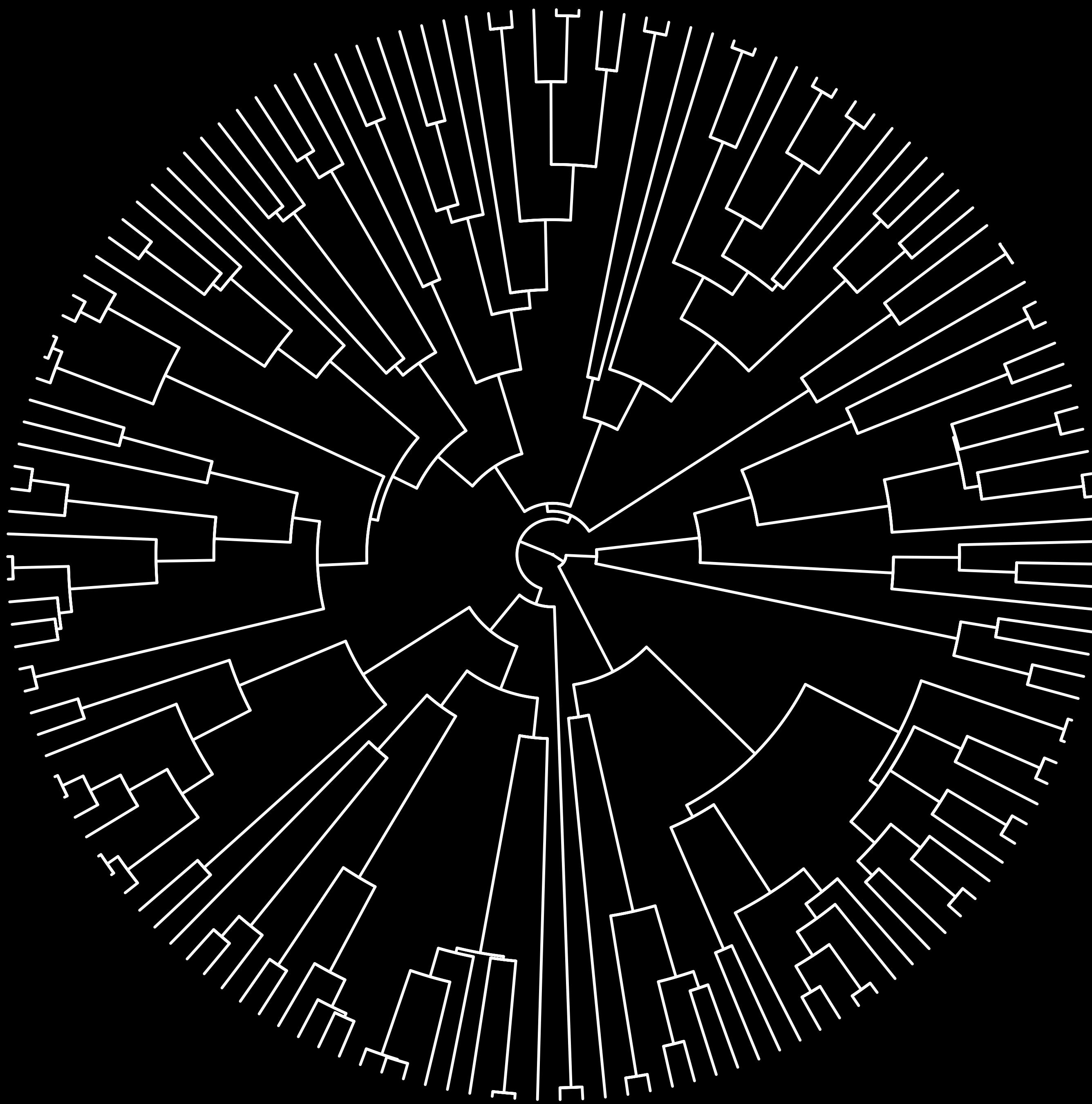
$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

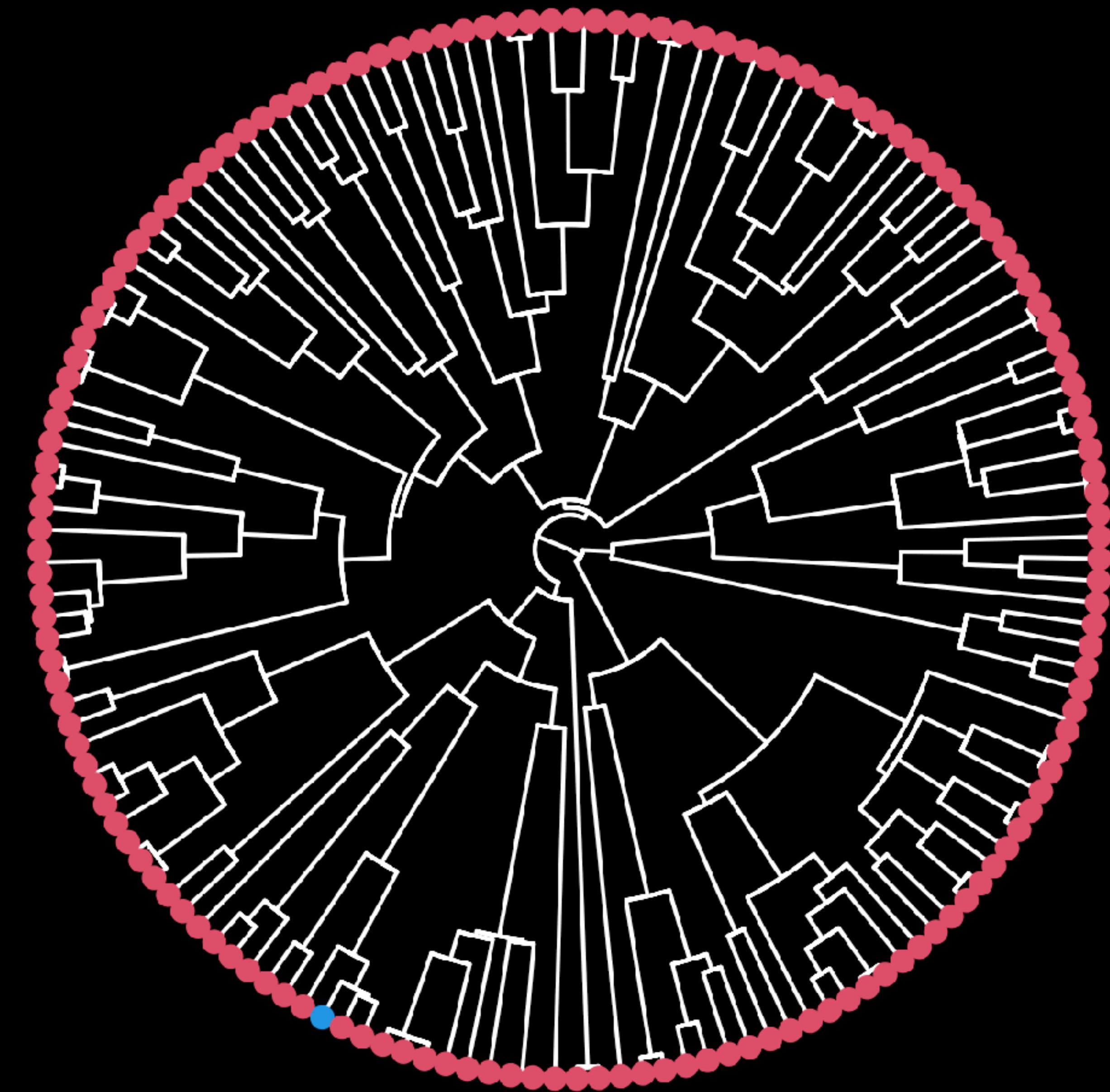
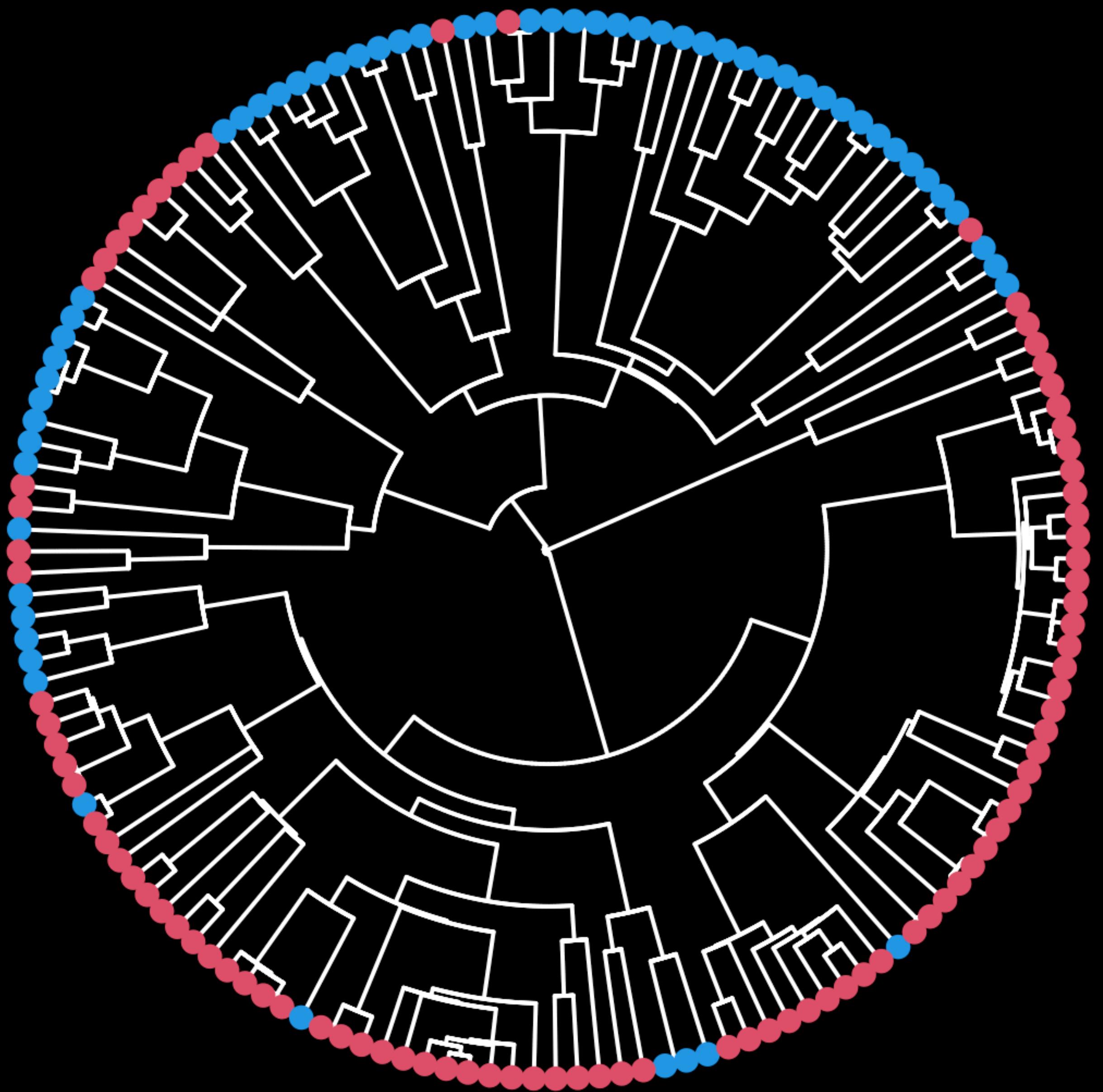
$$\sigma \sim \text{Exponential}(1)$$









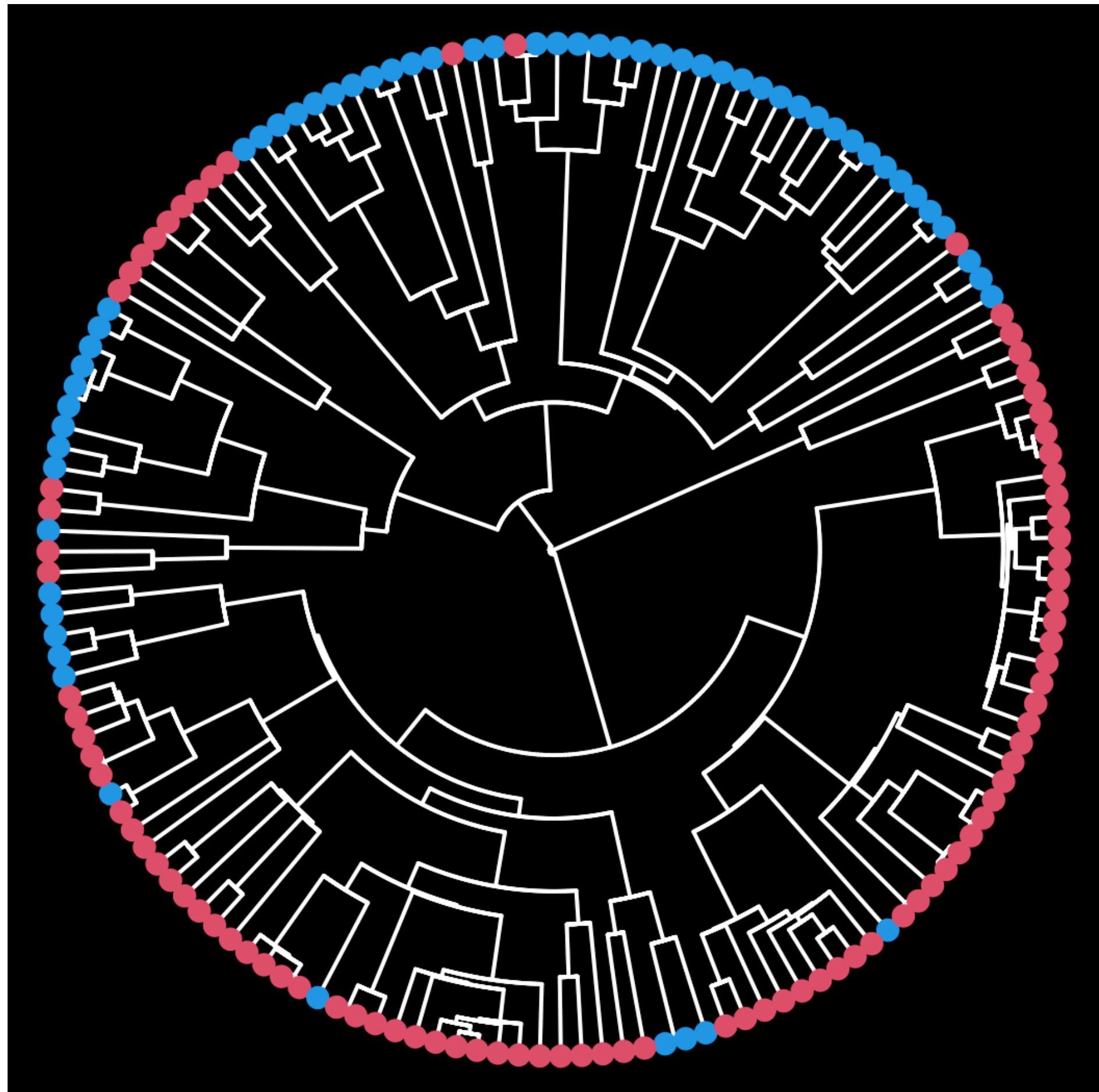


# From Model to Kernel

Evolutionary model + tree structure  
= pattern of covariation at tips

Covariance declines with  
phylogenetic distance

*Phylogenetic distance*: Branch length  
from one species to another



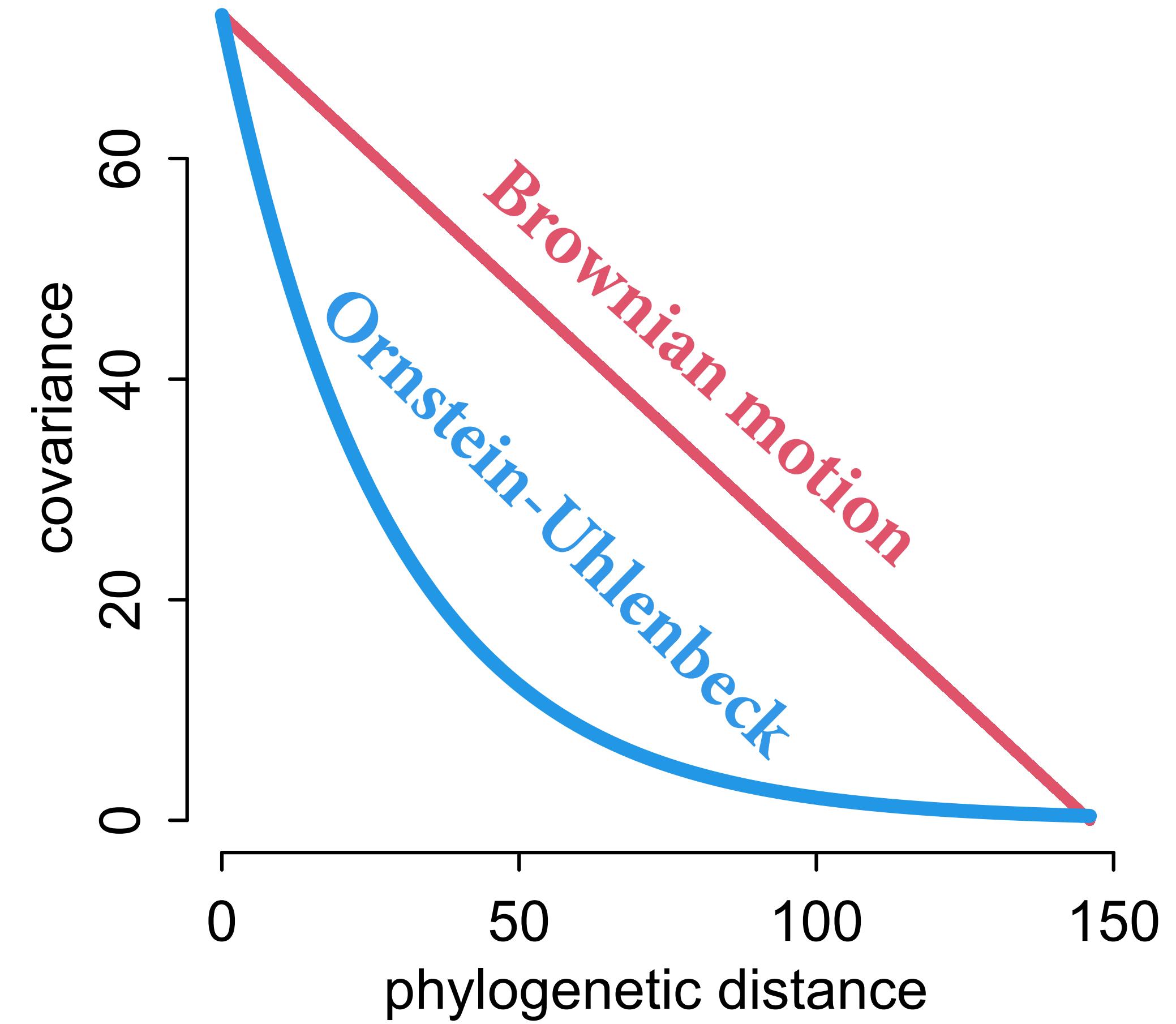
# From Model to Kernel

Evolutionary model + tree structure  
= pattern of covariation at tips

Common simple models:

**Brownian motion**

**Ornstein-Uhlenbeck** (damped Brownian motion)



$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

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

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G,\beta_M \sim \text{Normal}(0,0.5)$$

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

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

*Ornstein-Uhlenbeck kernel*

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

*Maximum covariance prior*

$$\rho \sim \text{HalfNormal}(3,0.25)$$

*Rate prior*

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

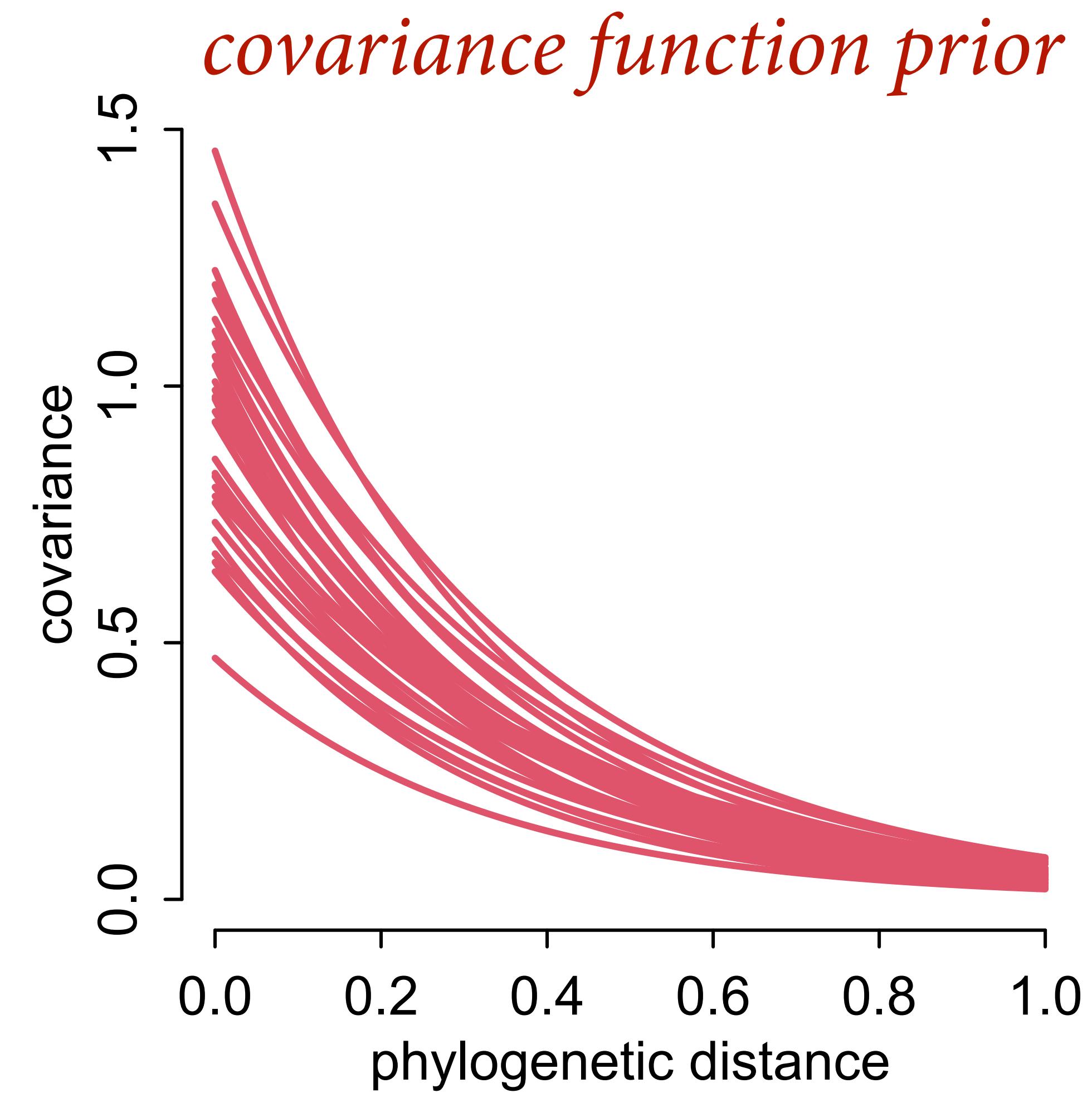
$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

$$\rho \sim \text{HalfNormal}(3,0.25)$$



$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

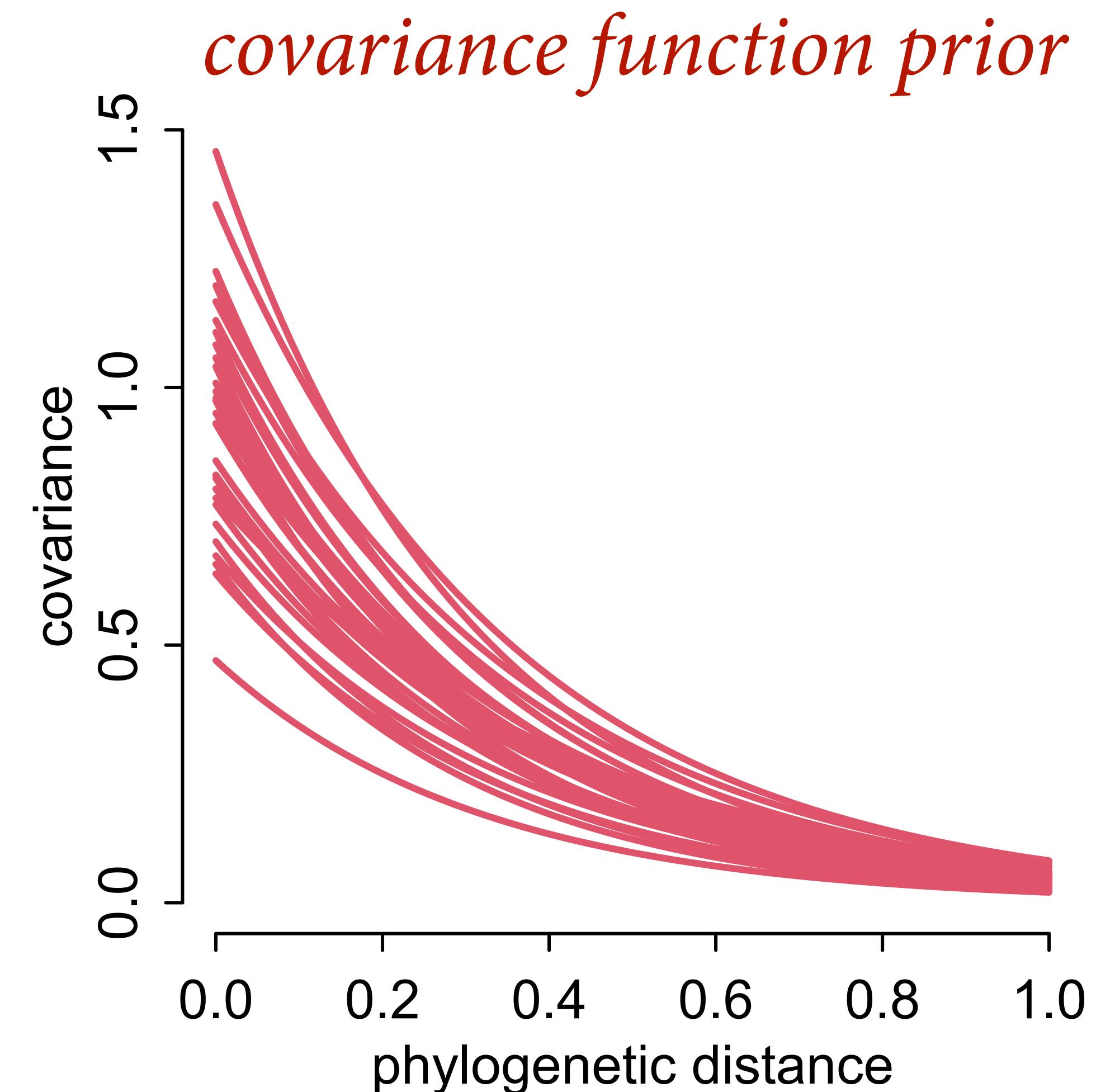
$$\mu_i = \alpha$$

$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

$$\rho \sim \text{HalfNormal}(3,0.25)$$



```

# Ornstein-Uhlenbeck (L1 gaussian process)
# add scaled and reordered distance matrix
dat_list$Dmat <- Dmat[ spp_obs , spp_obs ] / max(Dmat)

mB_OU <- ulam(
  alist(
    B ~ multi_normal( mu , K ),
    mu <- a + 0*M,
    matrix[N_spp,N_spp]:K <- cov_GPL1(Dmat,etasq,rho,0.01),
    a ~ normal(0,1),
    etasq ~ half_normal(1,0.25),
    rho ~ half_normal(3,0.25)
  ), data=dat_list , chains=4 , cores=4 )

```

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha$$

$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

$$\rho \sim \text{HalfNormal}(3,0.25)$$

```

# Ornstein-Uhlenbeck (L1 gaussian process)
# add scaled and reordered distance matrix
dat_list$Dmat <- Dmat[ spp_obs , spp_obs ] / max(Dmat)

mB_OU <- ulam(
  alist(
    B ~ multi_normal( mu , K ),
    mu <- a + U * M,
    matrix[N_spp,N_spp]:K <- cov_GPL1(Dmat,etasq,rho,0.01),
    a ~ normal(0.1),
    etasq ~ half_normal(1,0.25),
    rho ~ half_normal(3,0.25)
  ), data=dat_list , chains=4 , cores=4 )

```

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

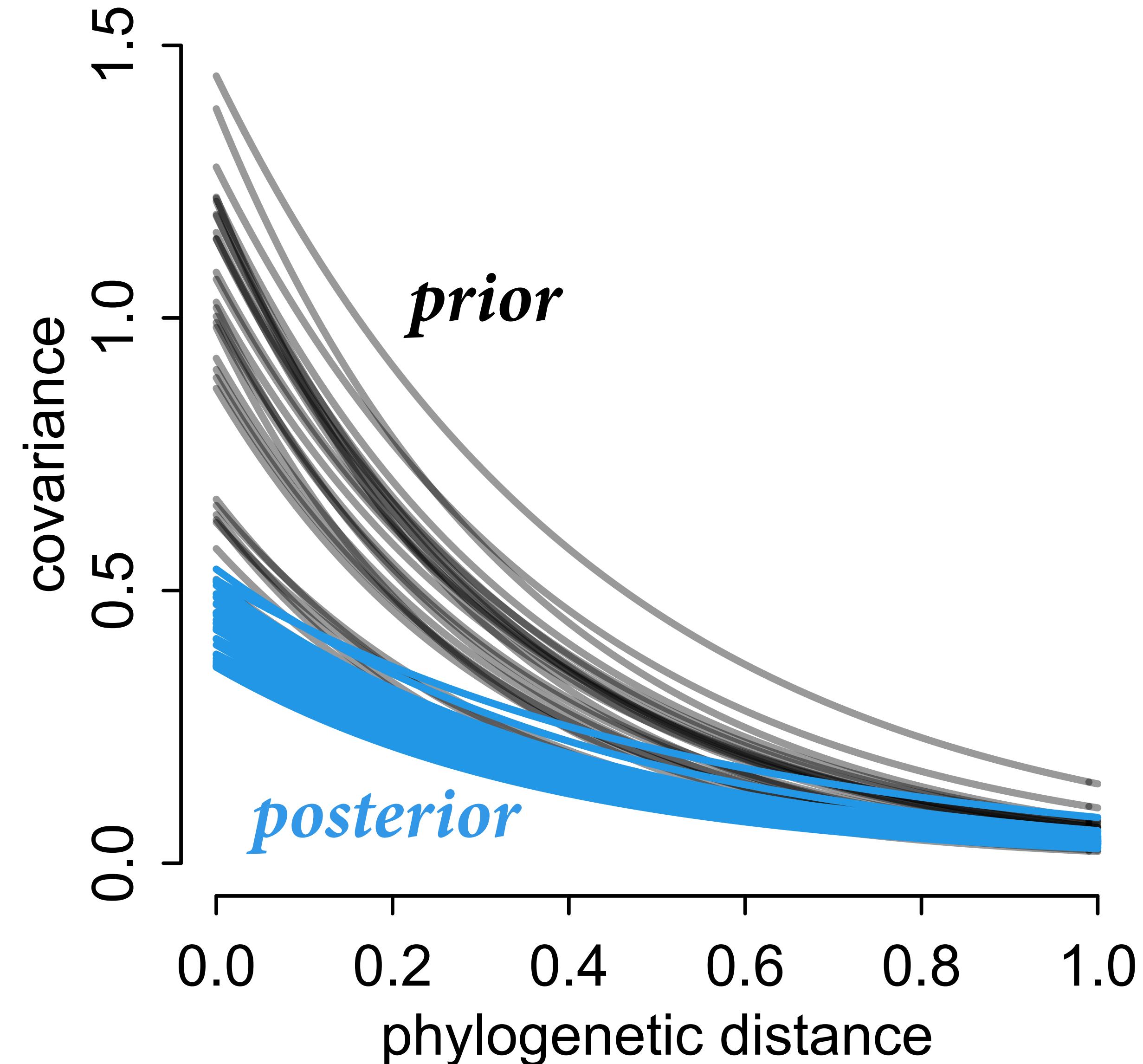
$$\mu_i = \alpha$$

$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

$$\rho \sim \text{HalfNormal}(3,0.25)$$



$B \sim \text{MVNormal}(\mu, \mathbf{K})$   
 $\mu_i = \alpha$   
 $\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$   
 $\alpha \sim \text{Normal}(0, 1)$   
 $\eta^2 \sim \text{HalfNormal}(1, 0.25)$   
 $\rho \sim \text{HalfNormal}(3, 0.25)$

```

mBMG_OU <- ulam(
  alist(
    B ~ multi_normal( mu , K ) ,
    mu <- a + bM*M + bG*G ,
    matrix[N_spp,N_spp]:K <- cov_GPL1(Dmat,etasq,rho,0.01) ,
    a ~ normal(0,1) ,
    c(bM,bG) ~ normal(0,0.5) ,
    etasq ~ half_normal(1,0.25) ,
    rho ~ half_normal(3,0.25)
  ), data=dat_list , chains=4 , cores=4 )

```

$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

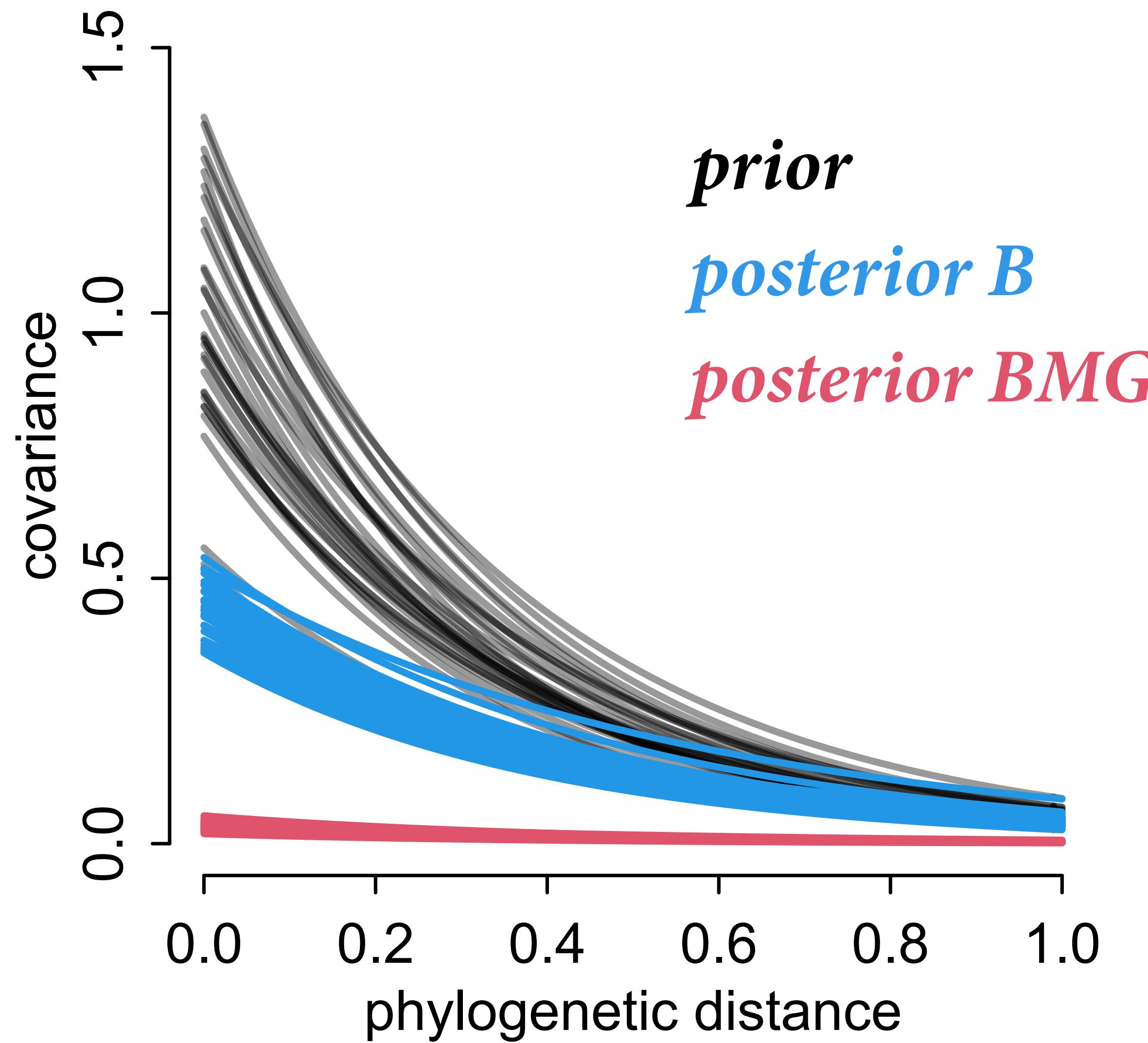
$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

$$\rho \sim \text{HalfNormal}(3,0.25)$$



$$B \sim \text{MVNormal}(\mu, \mathbf{K})$$

$$\mu_i = \alpha + \beta_G G_i + \beta_M M_i$$

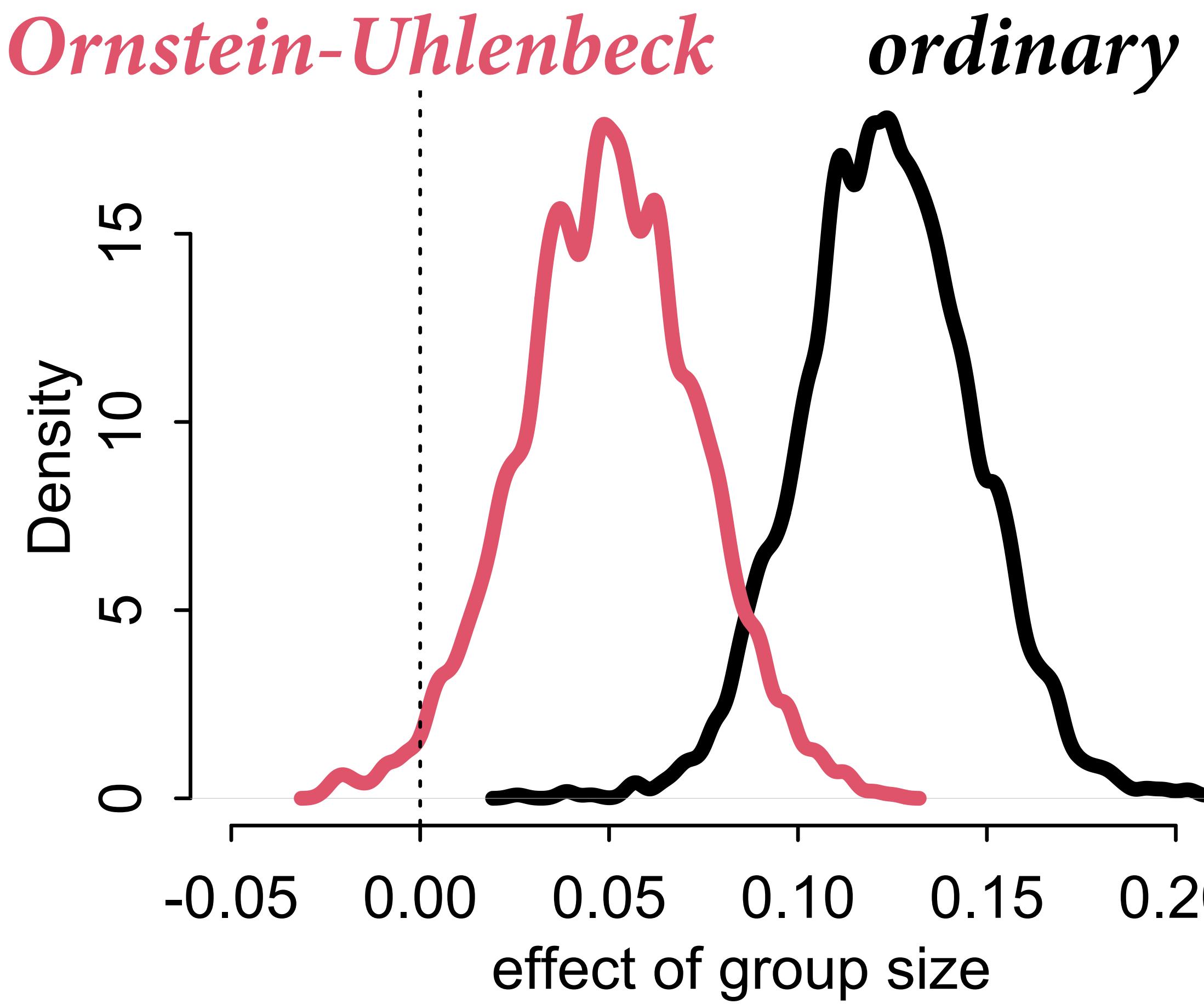
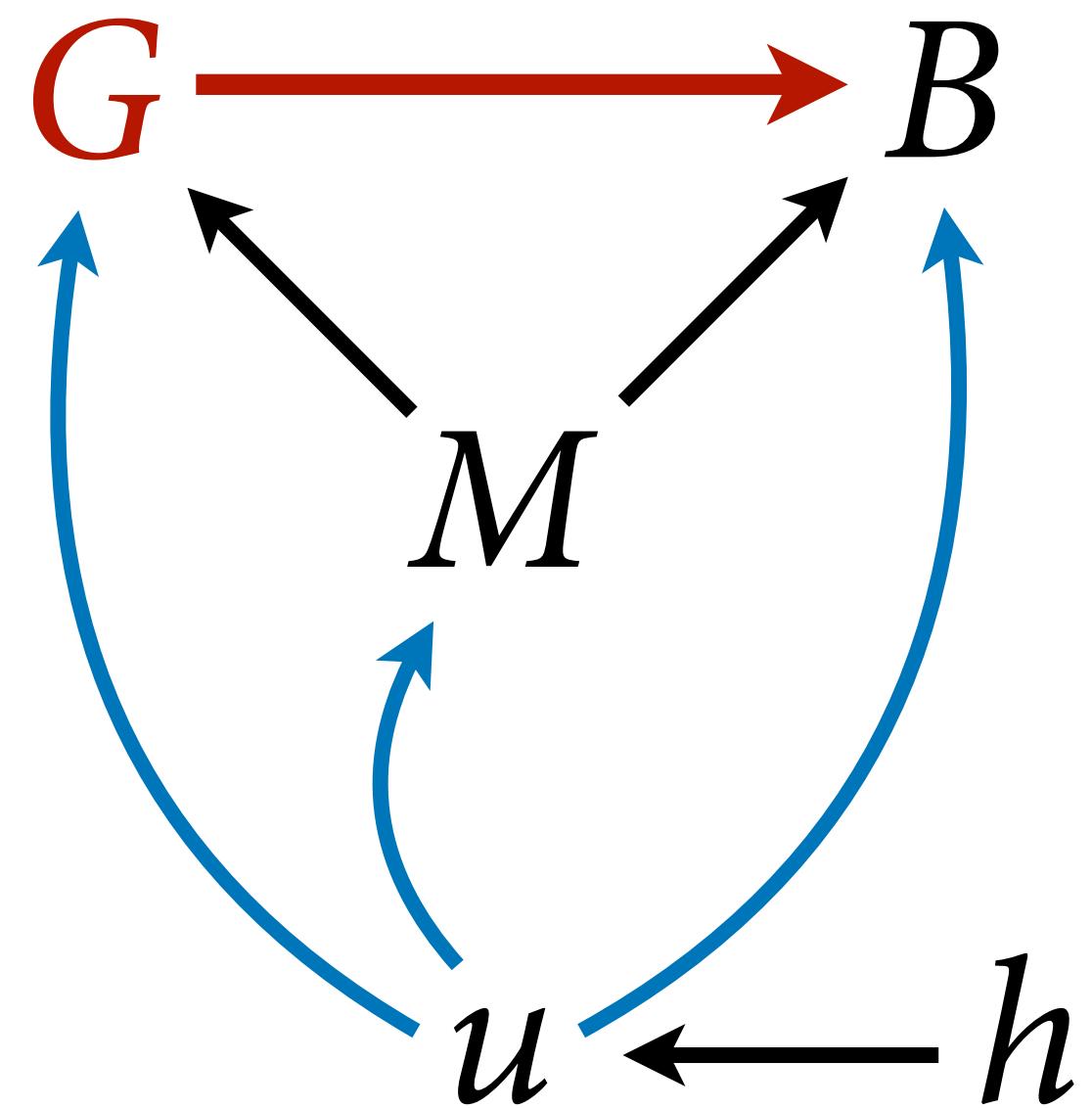
$$\mathbf{K} = \eta^2 \exp(-\rho d_{i,j})$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta_G, \beta_M \sim \text{Normal}(0,0.5)$$

$$\eta^2 \sim \text{HalfNormal}(1,0.25)$$

$$\rho \sim \text{HalfNormal}(3,0.25)$$

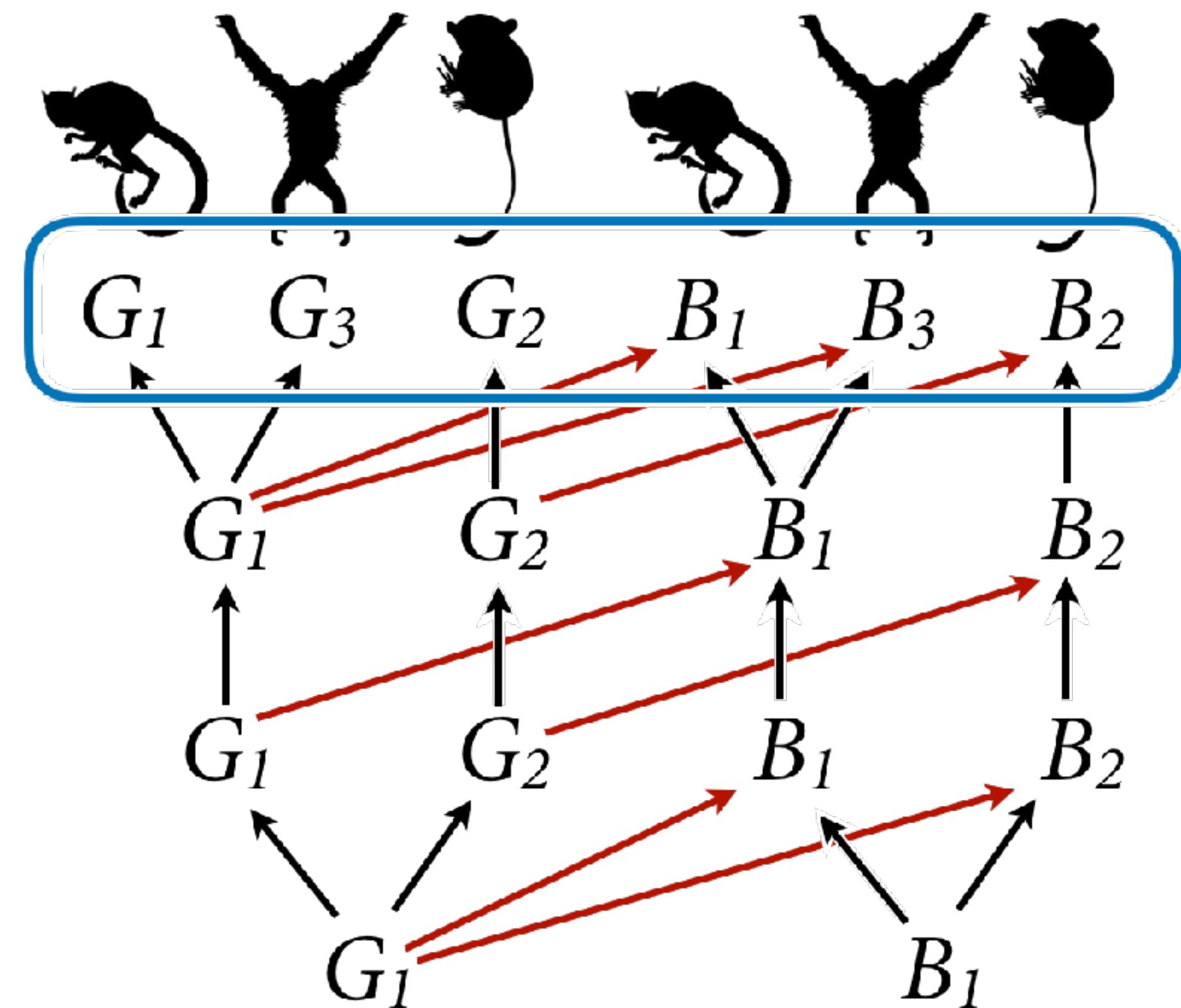


# Phylogenetic regression

Lingering problems:

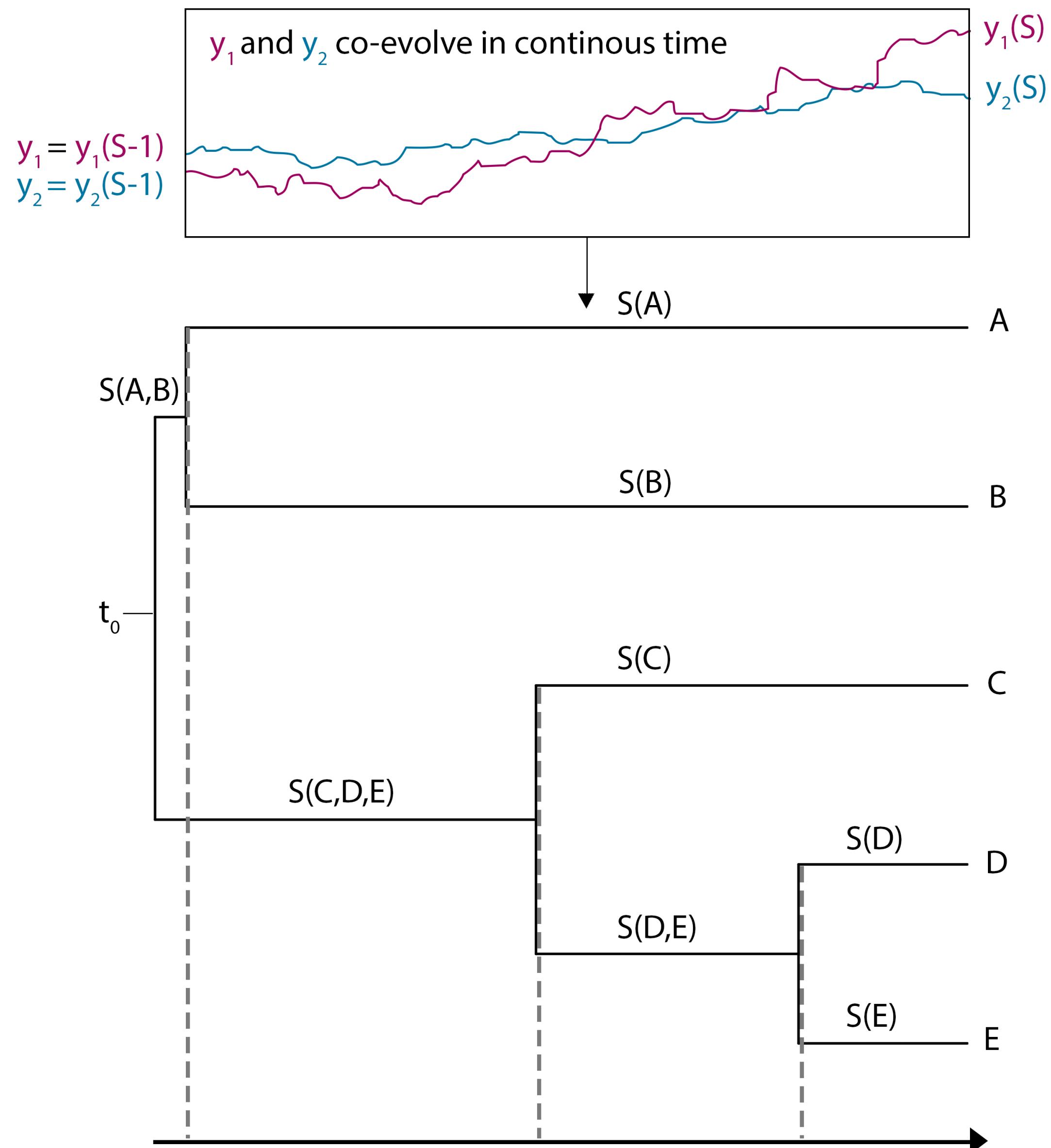
(1) What about phylogenetic uncertainty?

(2) Don't these traits influence one another reciprocally over time?

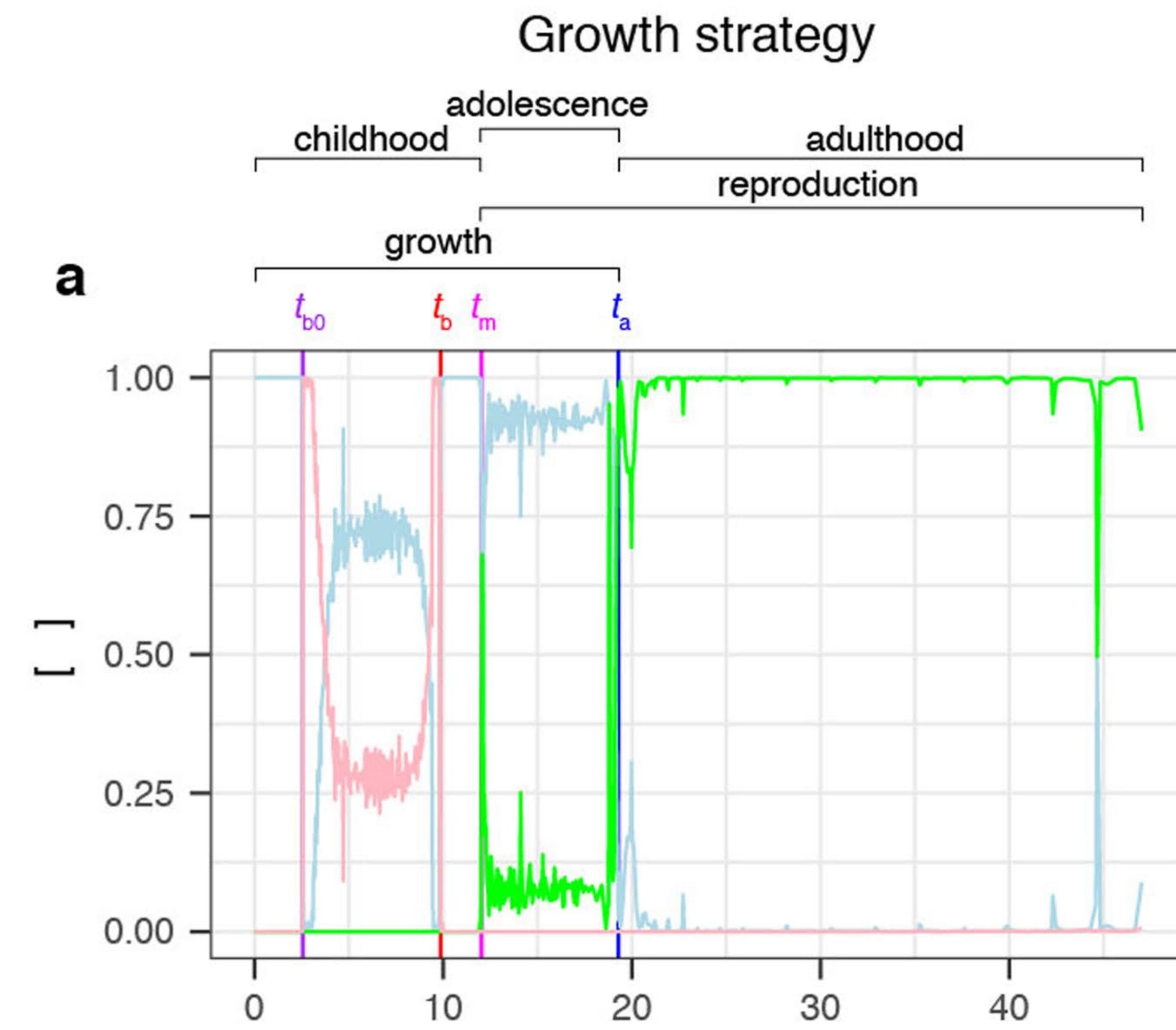


Ringen, Martin, Jaeggi 2021

*Drift-coevolution dynamics*



González-Forero & Gardner 2018  
*Optimal life history approach*



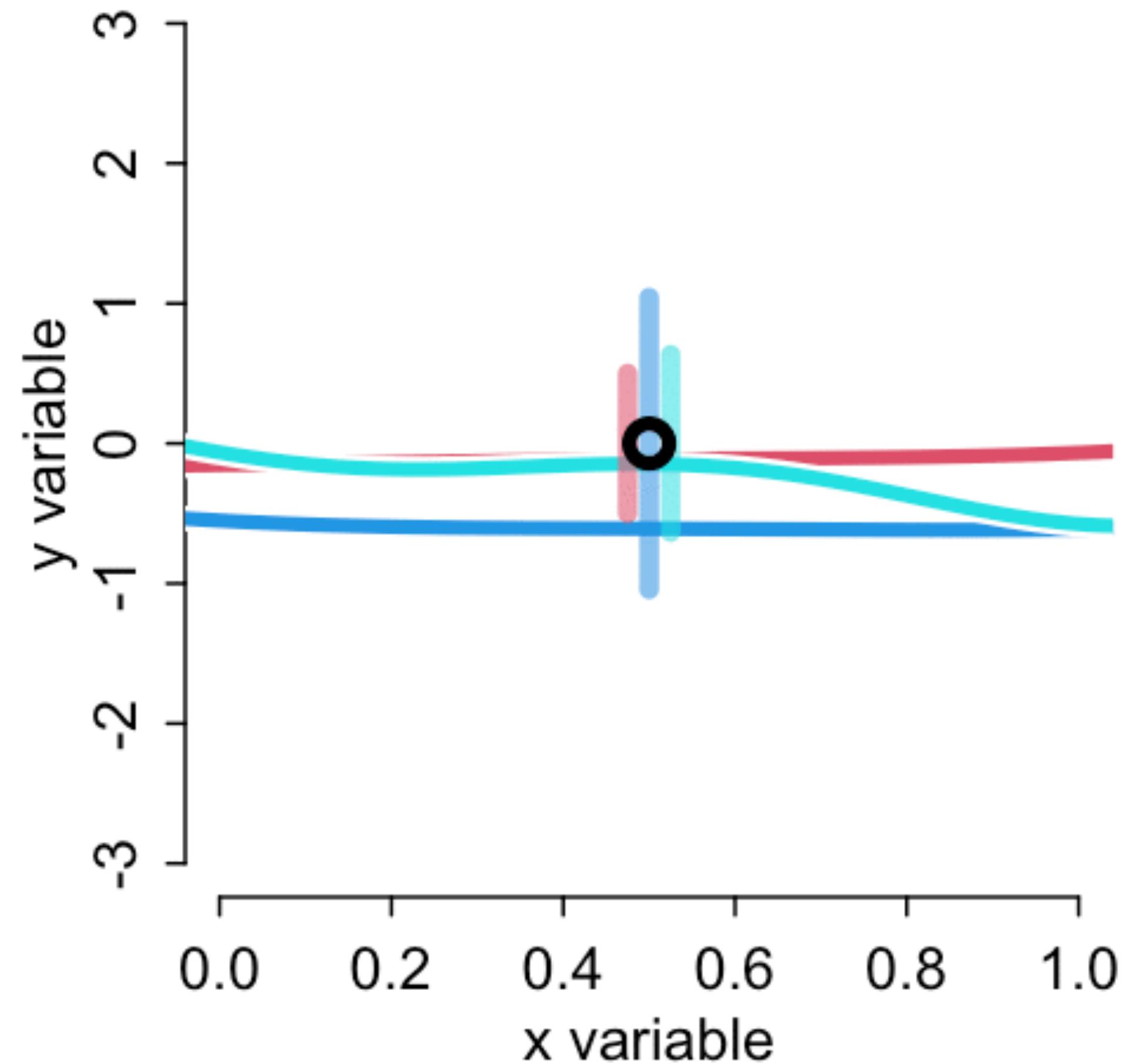
# Gaussian Processes

Partial pooling for continuous categories

Very general approximation engine

Causal theory => covariance kernel

Sensitivity to kernel priors — choose wisely

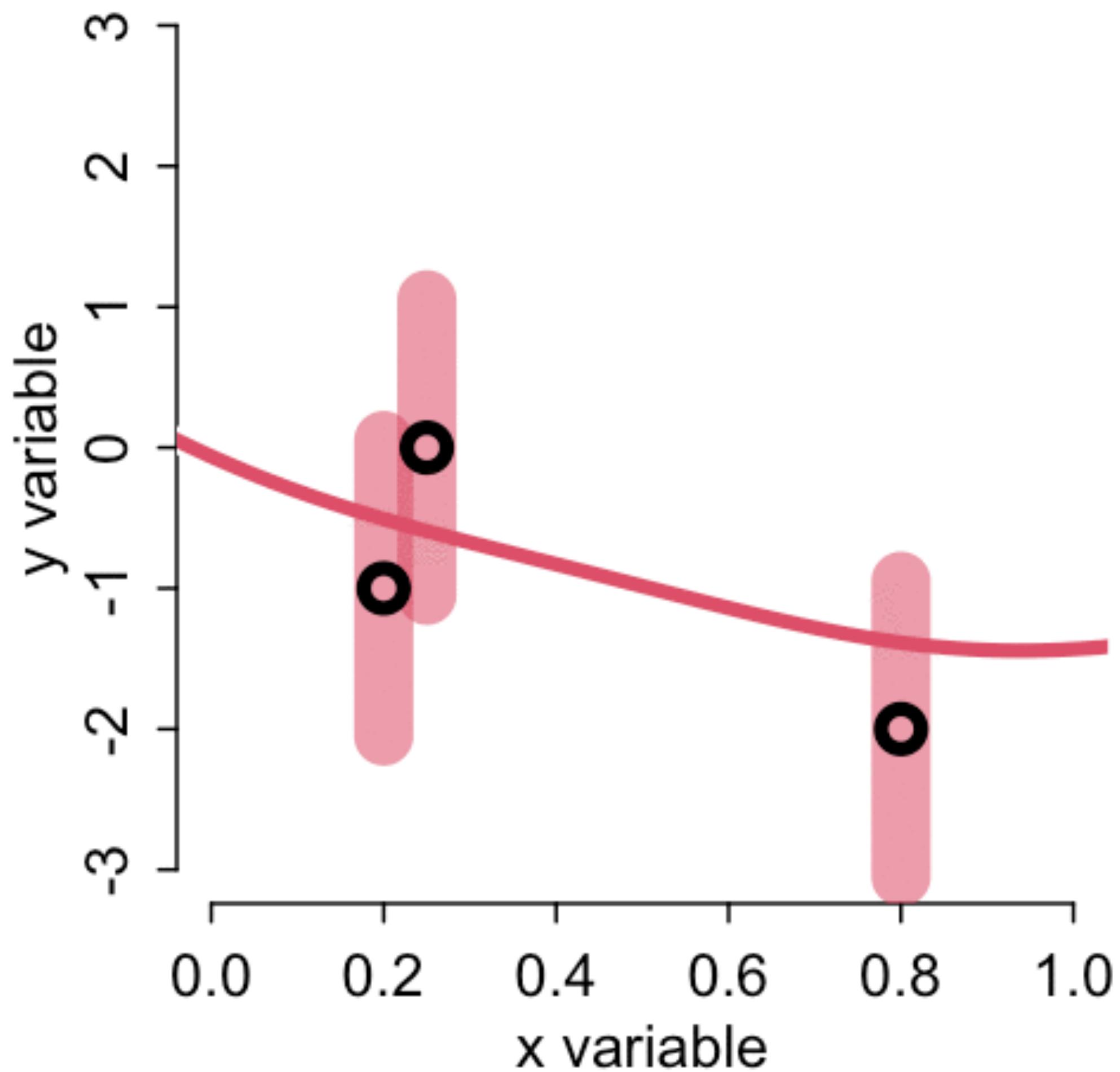


# Gaussian Possibilities

Automatic relevance determination  
(ARD): Multiple distance  
dimensions inside the kernel

Multi-output Gaussian processes:  
Draw vectors from kernel

Telemetry, navigation: Real-time  
tracking and error correction  
(*Kálmán filter*)



# Course Schedule

|         |  |                  |
|---------|--|------------------|
| Week 1  | Bayesian inference                     | Chapters 1, 2, 3 |
| Week 2  | Linear models & Causal Inference       | Chapter 4        |
| Week 3  | Causes, Confounds & Colliders          | Chapters 5 & 6   |
| Week 4  | Overfitting / MCMC                     | Chapters 7, 8, 9 |
| Week 5  | Generalized Linear Models              | Chapters 10, 11  |
| Week 6  | Ordered categories & Multilevel models | Chapters 12 & 13 |
| Week 7  | More Multilevel models                 | Chapters 13 & 14 |
| Week 8  | Social Networks & Gaussian Processes   | Chapter 14       |
| Week 9  | Measurement & Missingness              | Chapter 15       |
| Week 10 | Generalized Linear Madness             | Chapter 16       |

[https://github.com/rmcelreath/stat\\_rethinking\\_2022](https://github.com/rmcelreath/stat_rethinking_2022)

