

# Lab 3 Question 2 Bioinformatics

*Duc Doung, Martin Smelik, Raymond Sseguya*

*2018 M12 8*

## Assignment 2

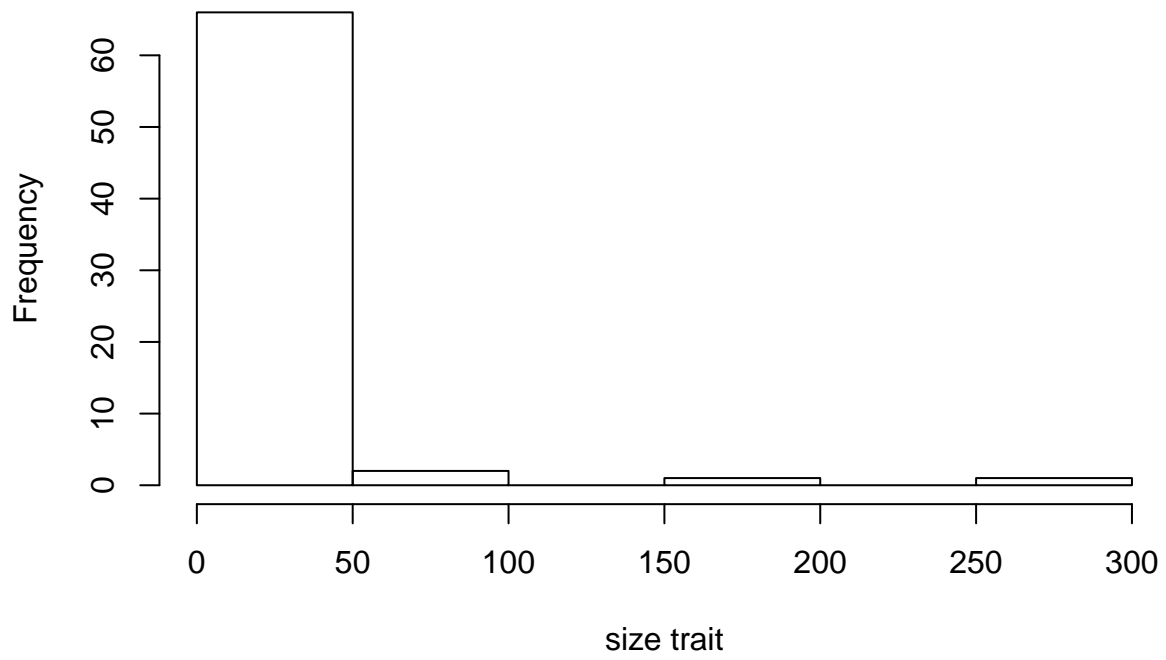
### 2.1

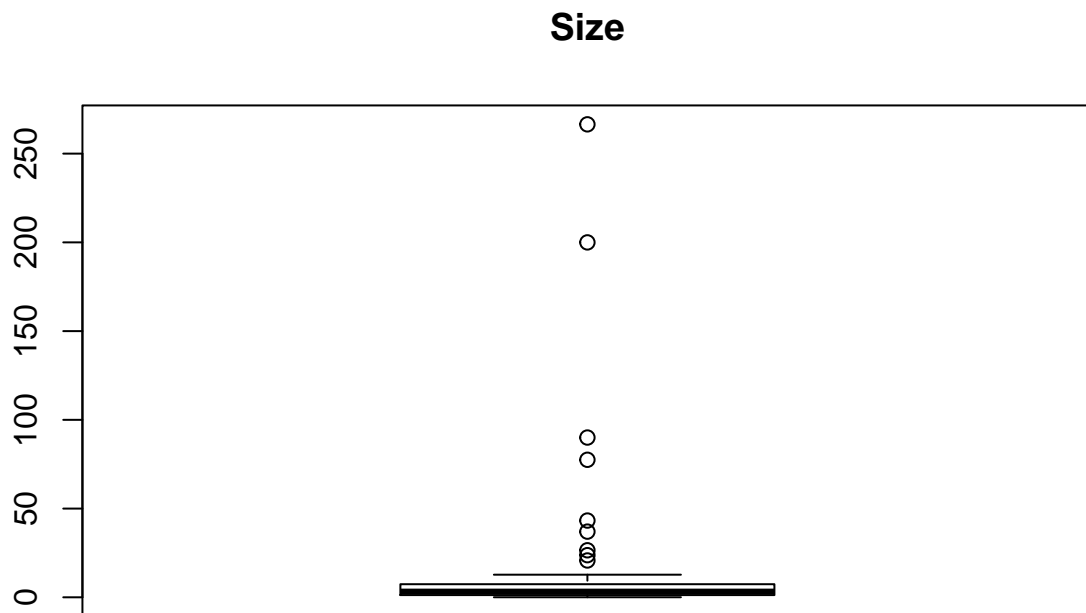
We analysed the carnivores dataset, `data(carni70)`, from the `ade4` R package. The data is about 70 species.

```
head(carni70$tab, 10)
```

##	size	range
## Puma_concolor	37.01	36.00
## Herpailurus_yaguaroundi	2.59	24.00
## Leopardus_wiedii	3.20	21.75
## Leopardus_pardalis	7.90	24.50
## Oreailurus_jacobita	3.99	1.75
## Oncifelis_colocolo	2.99	6.25
## Oncifelis_guigna	2.23	1.25
## Oncifelis_geoffroyi	3.80	4.75
## Leopardus_tigrinus	2.19	20.25
## Lynx_rufus	11.20	13.50

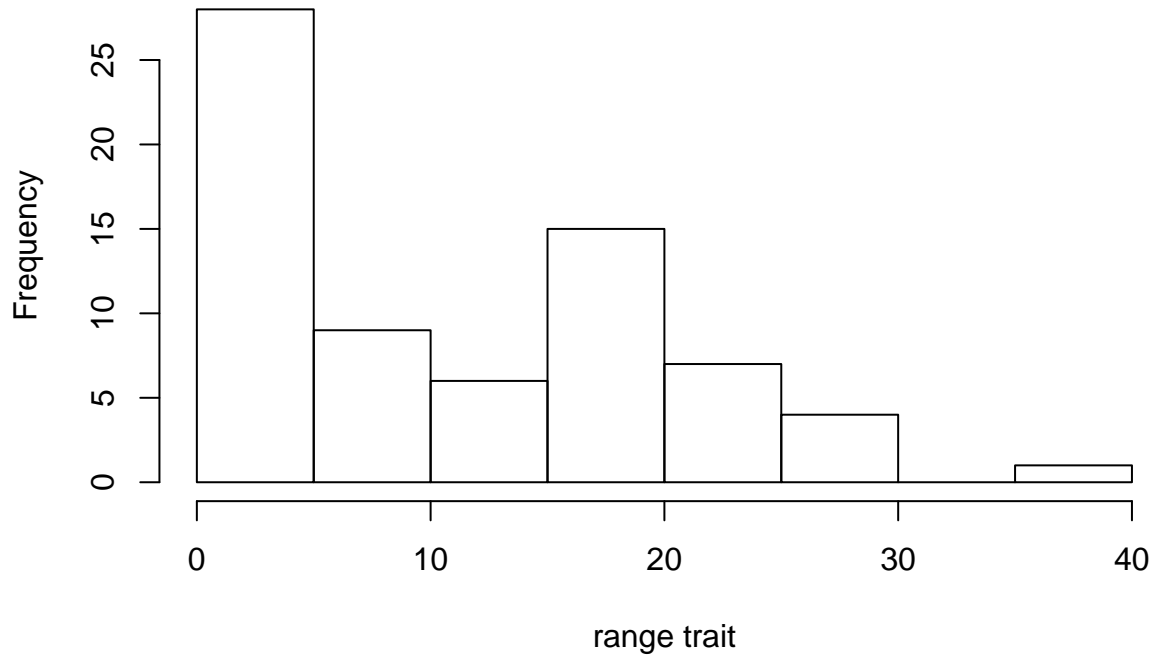
**Histogram of Size**

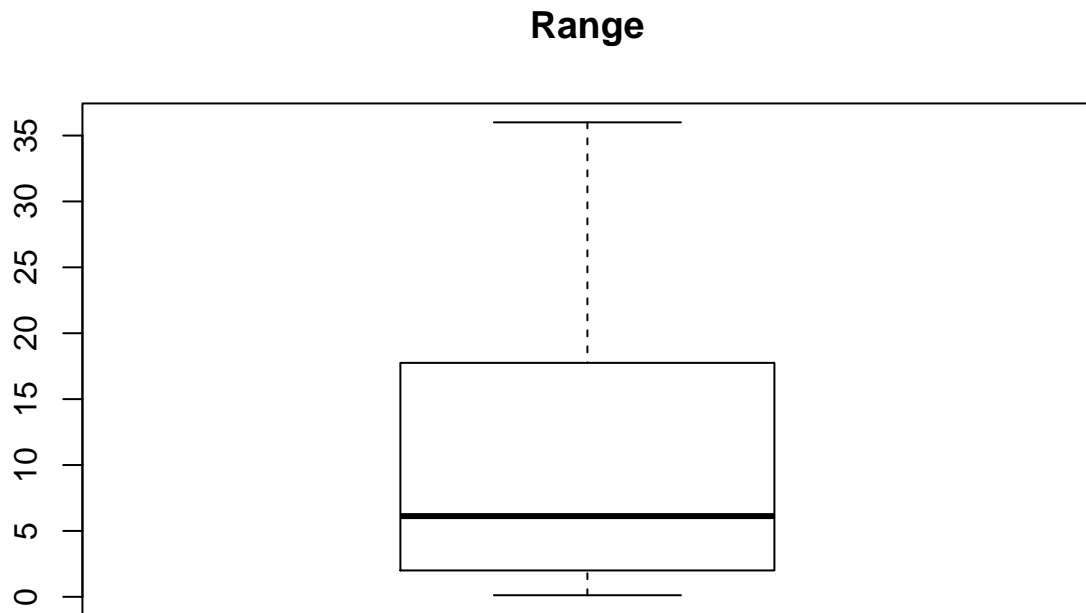




The distribution of the size trait is skewed to the left. In other words, the majority of the 70 species have small sizes. Some of the species have outlier sizes.

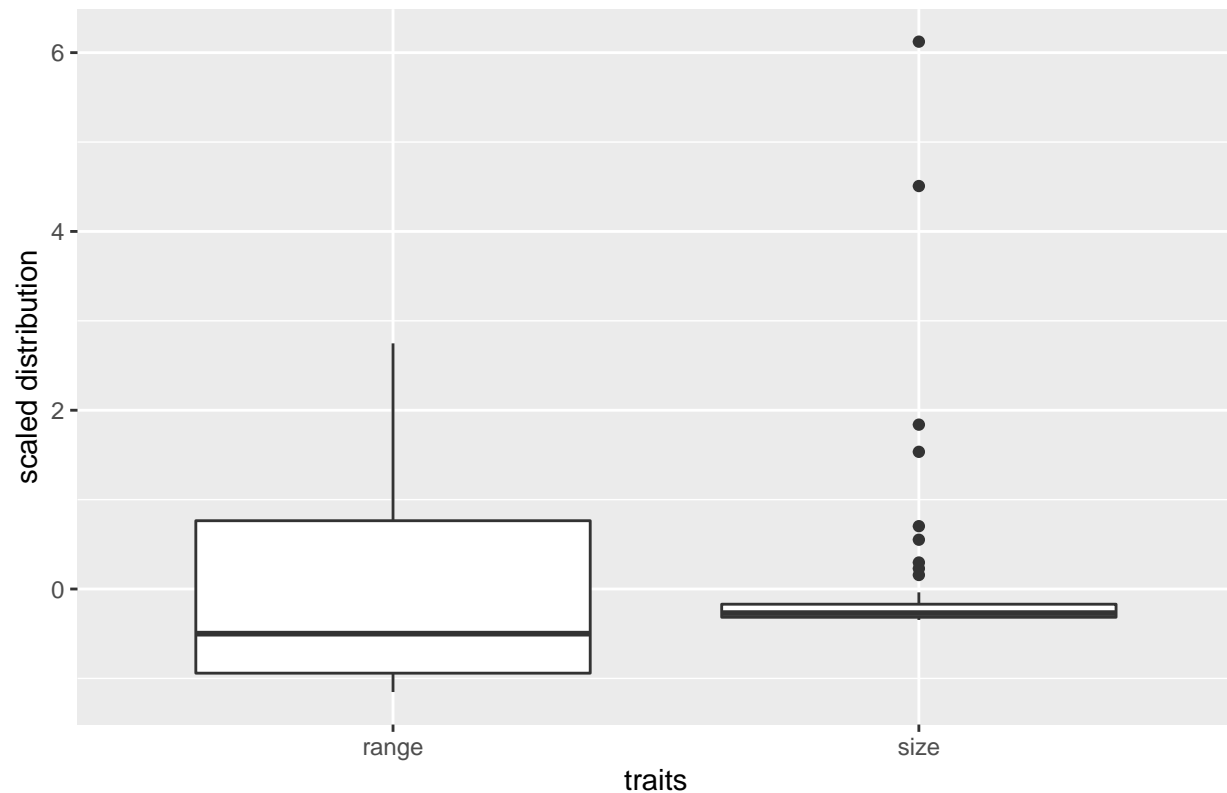
**Histogram of Range**





The distribution of the range trait is also generally more balanced however with evident skewness to the left. There are fewer and fewer species with longer ranges.

## Size and Range compared on the same scale



After scaling the data and comparing the distributions of size and range traits on the same scale with boxplots, it is clear that the size trait is more skewed towards low values than the range trait.

```
rownames(carni70_2[which.max(carni70_2[, "size"]), ])
```

```
## [1] "Ursus_arctos"
```

```
rownames(carni70_2[which.min(carni70_2[, "size"]), ])
```

```
## [1] "Mustela_nivalis"
```

```
rownames(carni70_2[which.max(carni70_2[, "range"]), ])
```

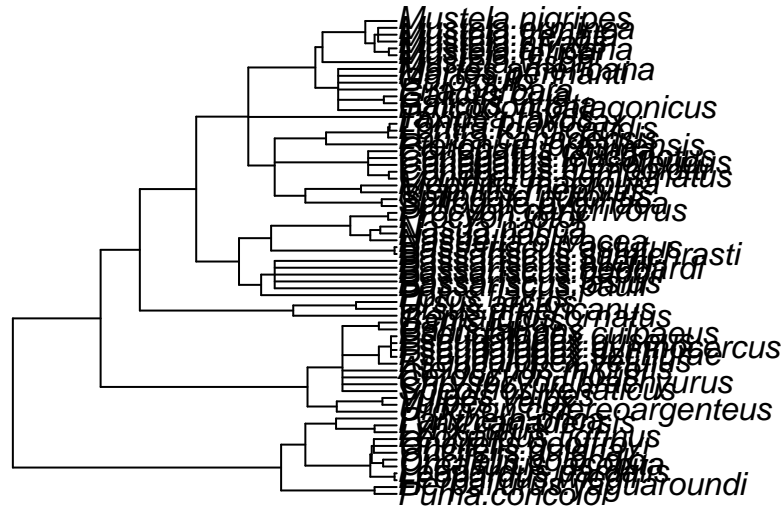
```
## [1] "Puma_concolor"
```

```
rownames(carni70_2[which.min(carni70_2[, "range"]), ])
```

```
## [1] "Bassariscus_pauli"
```

The species “Ursus\_arctos” has the biggest size while “Mustela\_nivalis” the smallest size. The species “Puma\_concolor” has the biggest range while the species “Bassariscus\_pauli” has the smallest range

```
tree_phylo <- ape::read.tree(text=carni70_1)
plot(tree_phylo)
```



Above is a plot of the phylogenetic tree.

## 2.2

### 2.2.1 Both traits evolve as independent Brownian motions

```
## row names of the data matrix must match tip names of your phylogeny!
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results for constrained rate BM1 model --
## LogLikelihood:      -589.2373
## AIC:                1186.475
## AICc:               1186.771
## 4 parameters
##
## Estimated rate matrix
## -----
##           size      range
## size  73.70335  0.00000
## range  0.00000 13.54524
##
```

```
## Estimated root state
## -----
##           size    range
## theta: 38.43947 13.78439
```

Because of the assumption that the two traits follow independent processes, the off diagonal elements of the estimated rate matrix are zero. From the rate matrix, we see that the changes in the size trait along the phylogenetic tree are more frequent (at 73) than the changes of the range trait on the same tree. That explains why the histogram of the size trait is more left skewed than that of the range trait (13). This model also suggests that the ancestor(root) had a size of 38.43947 and a range of 13.78439.

## 2.2.2 The traits evolve as a correlated Brownian motion

```
## $ParamsInModel
## $ParamsInModel$Sxx
##           size    range
## size 171.415584 0.00000
## range  1.327048 26.01045
##
## $ParamsInModel$vX0
##           [,1]
## size    8.434762
## range 11.084473
##
##
## $ParamSummary
## $ParamSummary$StS
##           size    range
## size 29383.3025 227.4768
## range  227.4768 678.3047
##
## $ParamSummary$LogLik
## [1] -656.7365
##
## $ParamSummary$dof
## [1] 5
##
## $ParamSummary$m2loglik
## [1] 1313.473
##
## $ParamSummary$aic
## [1] 1323.473
##
## $ParamSummary$aic.c
## [1] 1323.921
##
## $ParamSummary$sic
## [1] 1338.181
##
## $ParamSummary$bic
## [1] 1338.181
##
```



```
## $ParamSummary$RSS
## [1] 140
```

Under the assumption that the two traits follow correlated Brownian motion processes, the ancestor (root) had a size of 8.434762 and a range of 11.084473. Even under this model, the size trait changes more frequently than the range trait, with 29383.3025 for size versus 678.3047 for range. The difference between size and range is clearly indicated in the skewnesses of histograms.

### 2.2.3 independent Ornstein Uhlenbeck processes

```
mvOU(tree_phylo, data=carni70$tab$size, model = c("OU1"),
      diagnostic = FALSE, echo = TRUE)
```

```
## species in the matrix are assumed to be in the same order as in the phylogeny, otherwise specify row
##
## -- Summary results --
## LogLikelihood:   -324.2639
## AIC:             654.5278
## AICc:            654.8914
## 3 parameters
##
## Estimated theta values
## -----
## OU1 38.43947
##
## ML alpha values
## -----
## 8.737191e-10
##
## ML sigma values
## -----
## 73.70246
```

```
mvOU(tree_phylo, data=carni70$tab$range, model = c("OU1"),
      diagnostic = FALSE, echo = TRUE)
```

```
## species in the matrix are assumed to be in the same order as in the phylogeny, otherwise specify row
##
## -- Summary results --
## LogLikelihood:   -248.9895
## AIC:             503.979
## AICc:            504.3427
## 3 parameters
##
## Estimated theta values
## -----
##
```

```
## OU1 11.08609
##
## ML alpha values
## -----
## 0.2295954
##
## ML sigma values
## -----
## 38.05329
```

## 2.2.4 traits evolve as a bivariate Ornstein-Uhlenbeck process

```
mvOU(tree_phylo, data=carni70$tab, model = c("OU1"),
      diagnostic = TRUE, echo = TRUE)
```

```
## row names of the data matrix must match tip names of your phylogeny!
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results --
## LogLikelihood: -572.6063
## AIC: 1161.213
## AICc: 1162.312
## 8 parameters
##
## Estimated theta values
## -----
##          size      range
## OU1 37.55833 11.42602
##
## ML alpha values
## -----
##          size      range
## size 8.345549e-05 -0.003859286
## range -3.859286e-03 0.233577603
##
## ML sigma values
## -----
##          size      range
## size 73.766441 5.859783
## range 5.859783 38.247492
```