

# Ungulate traits and ecogeography

Rutger Vos (@rvosa)

7-12-2018

## Trait and geospatial data

We have a tab-separated table containing continuous-valued trait data (including latitude and longitude values) from the PanTHERIA data base, filtered on the Ungulates, i.e. the Artiodactyla and the Perissodactyla.

Here we read that table into a data frame:

```
# the table has a header, and the 1st column contains the observation names: species
df <- read.table("ungulates.tsv", sep = "\t", header = T, row.names = 1)
```

We can plot the mid points of the latitudinal and longitudinal ranges of the species, on an `rworldmap`:

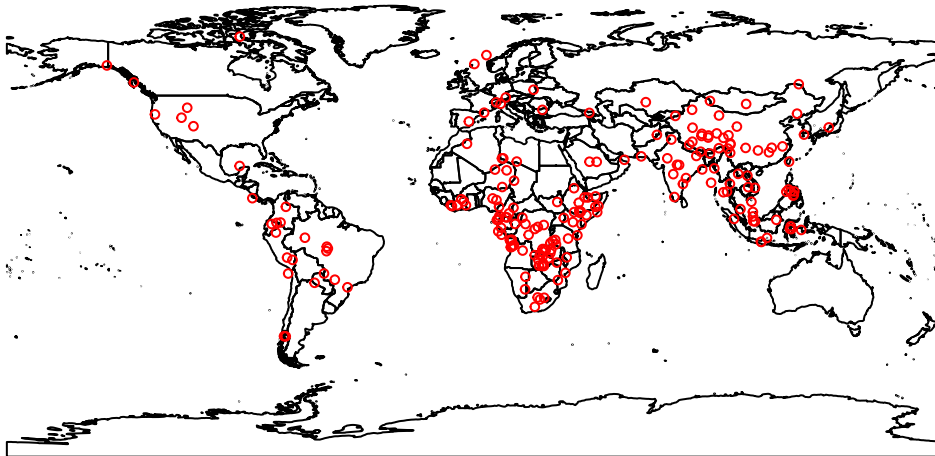
```
library(rworldmap, quietly = T)
```

```
## ### Welcome to rworldmap ###
```

```
## For a short introduction type : vignette('rworldmap')
```

```
newmap <- getMap(resolution = "low")
plot(newmap, asp = 1)
```

```
# notice the two columns in the data frame we read
points(df$X26.7_GR_MidRangeLong_dd, df$X26.4_GR_MidRangeLat_dd, col = "red", cex = .6)
```



## Phylogeny

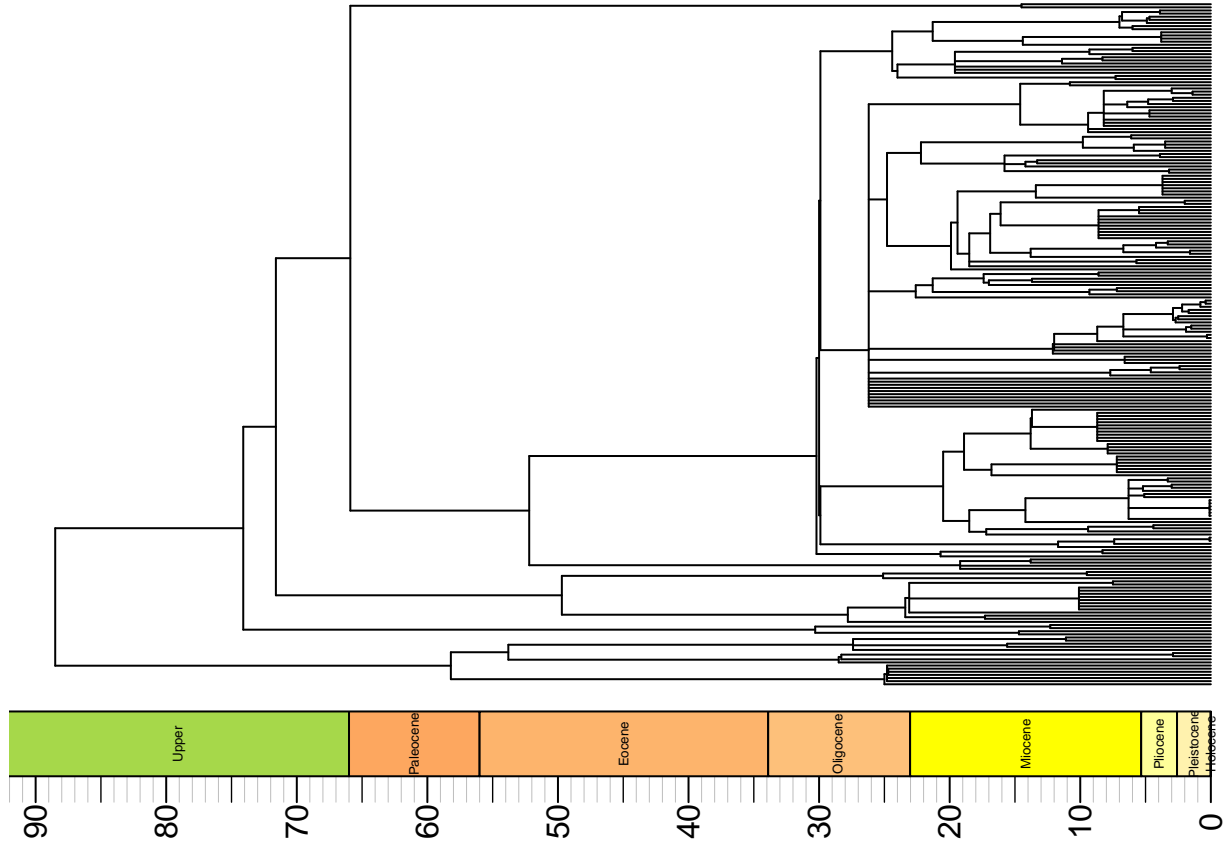
We also have a phylogenetic tree in Newick format. It is the topology of a supertree of the Mammals, but pruned to the samen taxon set as the trait table. To read that tree, we need to load the `ape` package:

```
library(ape, quietly = T)
tree <- read.tree(file = "ungulates.nwk")
```

The tree is a `phylo` object, a pseudo-standard in the R community that a number of packages build on. For example, we can plot the tree against a geological timescale using the `strap` package:

```
library(strap, quietly = T)

# set the age of the root to absolute time
tree$root.time <- 88.5
geoscalePhylo(tree, units = "Epoch", cex.age = 1, cex.ts = 0.5, boxes = T, show.tip.label = F)
```



## Combining traits and trees

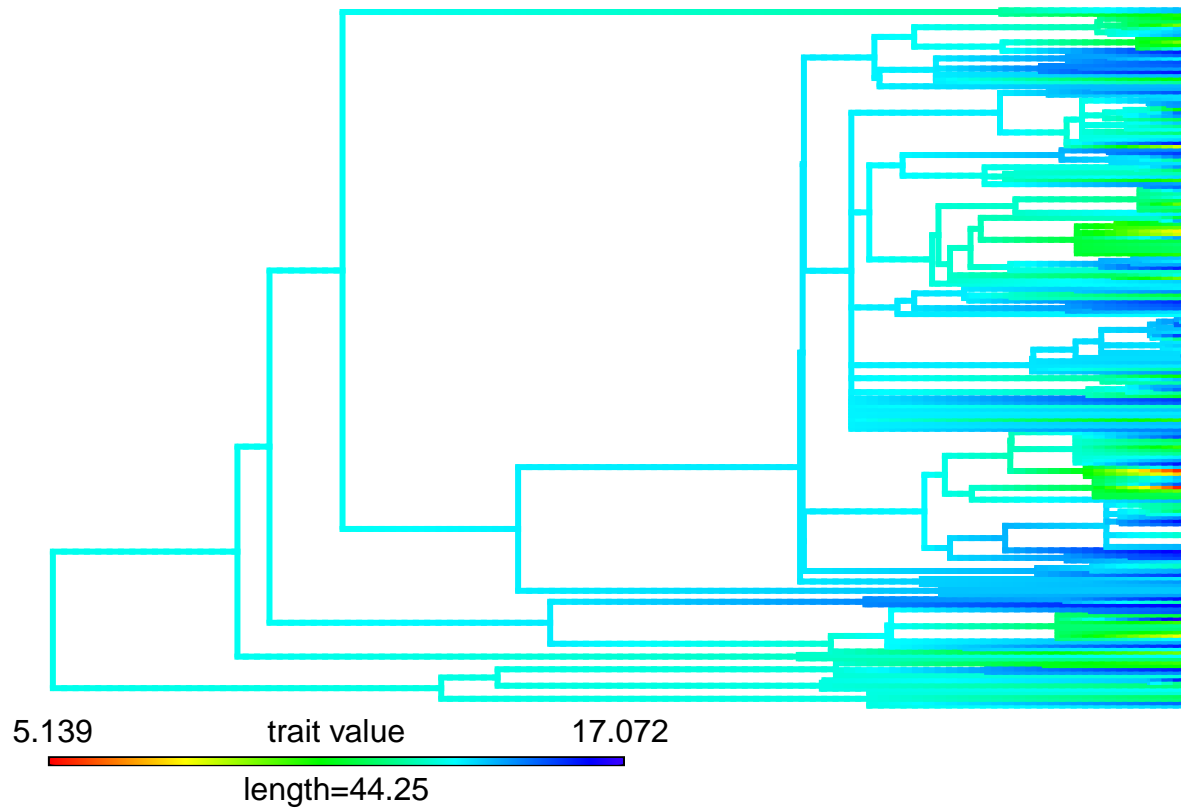
We can combine the trait data and the tree and plot trait values onto the tree. For example, here is the range size of the species, for which we need `phytools`:

```
library(phytools, quietly = T)

# make a vector of the area column, remove records with NA values
log_area <- log(as.vector(df$X26.1_GR_Area_km2))
names(log_area) <- row.names(df)
log_area <- log_area[!is.na(log_area)]

# drop the tips that had NA records for area
tips_to_prune <- setdiff(row.names(df), names(log_area))
pruned <- drop.tip(tree, tips_to_prune)

# do a continuous trait mapping, of log-transformed mass
contMap(pruned, log_area, standardize=TRUE, length=10, outline=F, lwd=3, ftype="off")
```



## Interaction between traits

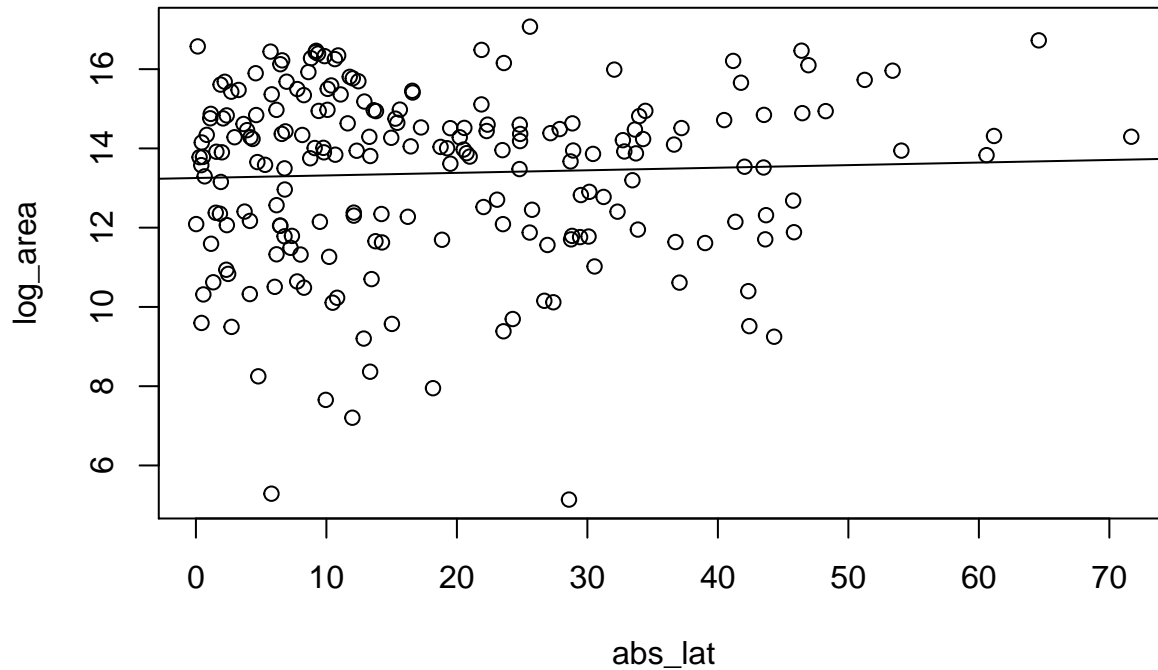
We can plot to see how latitude relates to range size:

```
# make a vector of absolute latitude, give it names, remove NA values
abs_lat <- abs(df$X26.4_GR_MidRangeLat_dd)
names(abs_lat) <- row.names(df)
abs_lat <- abs_lat[!is.na(abs_lat)]

# keep the intersection for each of the two variables
abs_lat <- abs_lat[intersect(names(abs_lat), names(log_area))]
log_area <- log_area[intersect(names(abs_lat), names(log_area))]
```

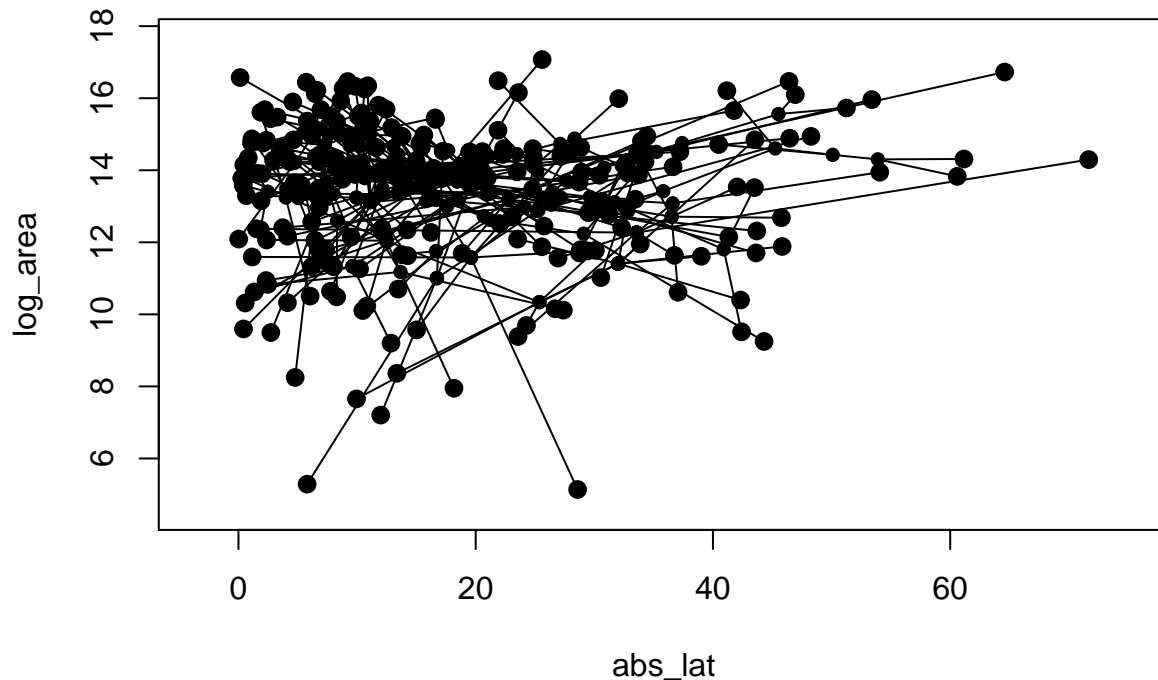
The INCORRECT approach would be to just plot the two variables against each other even though there is autocorrelation. Here is what that would look like:

```
fitwrong <- lm(log_area ~ abs_lat)
plot(abs_lat, log_area)
abline(fitwrong)
```



We can connect the data points by their relationships in the phylogeny:

```
# drop the tips that had NA records for body mass, randomly resolve
tips_to_prune <- setdiff(row.names(df), names(log_area))
pruned <- multi2di(drop.tip(tree, tips_to_prune))
phylomorphospace(pruned, data.frame(abs_lat, log_area), label = "off")
```

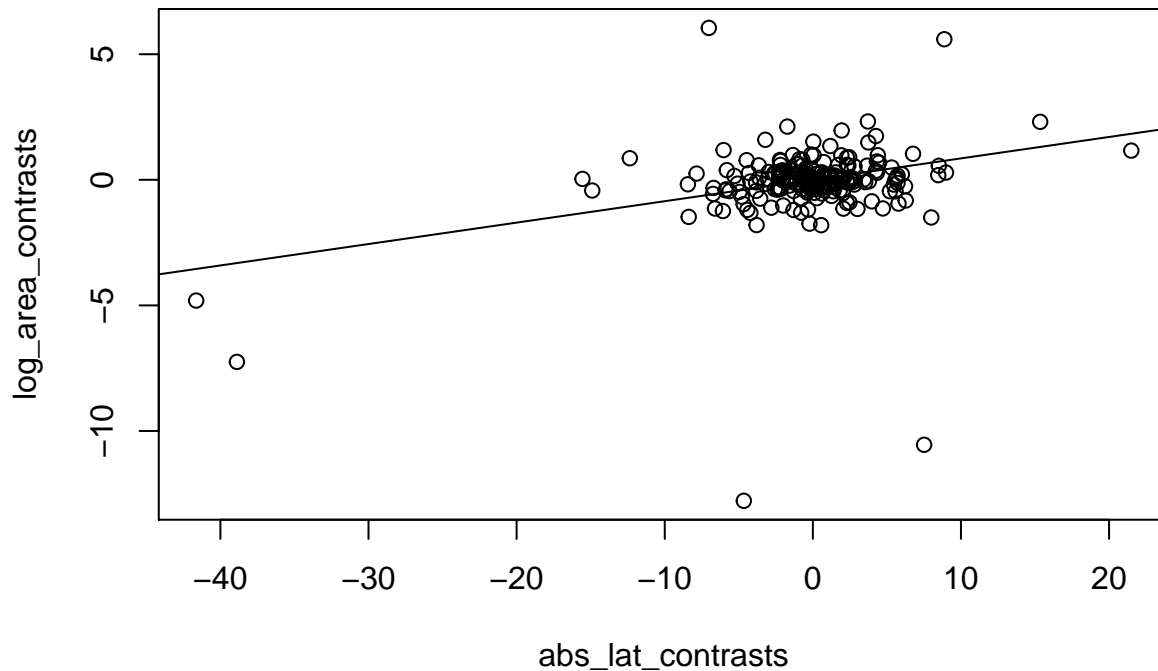


It looks like the placement of the points in morphospace may be shaped by phylogeny. One way to deal with such data is to compute independent contrasts and do subsequent linear modeling on those:

```
# compute independent contrasts
abs_lat_contrasts <- pic(abs_lat, pruned)
```

```
log_area_contrasts <- pic(log_area,pruned)

# fit linear model
fit <- lm(log_area_contrasts ~ abs_lat_contrasts - 1)
plot(abs_lat_contrasts,log_area_contrasts)
abline(fit)
```



There appears to be correlation:

```
summary(fit)

##
## Call:
## lm(formula = log_area_contrasts ~ abs_lat_contrasts - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.3833  -0.4018  -0.0188   0.3849   6.6463
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## abs_lat_contrasts  0.08525     0.01812   4.705 4.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.507 on 205 degrees of freedom
## Multiple R-squared:  0.09746,    Adjusted R-squared:  0.09306
## F-statistic: 22.14 on 1 and 205 DF,  p-value: 4.664e-06
```

## Scaling rules

The result of our analysis shows support for Rapoport's rule: an ecogeographical rule that states that latitudinal ranges of plants and animals are generally smaller at lower latitudes than at higher latitudes. In addition to this correlation, our data set surely shows other correlations, such as trivial scalings:

- Body mass  $\sim$  body length
- Longevity  $\sim$  body mass
- Neonate body mass  $\sim$  gestation length