

Aggregation and merging in the  
*primate* R-package (Version 0.2.0):  
Using locomotion and height SQL database tables  
to assess arboreal risk.

David Schruth dschruth@anthropoidea.net

Marc Myers mmyers@primate.org

Noel Rowe nrowe@primate.org

May 5, 2022

## 1 Licensing

This package is licensed under the Apache License v2.0: it is therefore free to use and redistribute, however, we, the copyright holders, wish to maintain primary control over any further development. Please be sure to cite us if you use this package in work leading to publication.

## 2 Installation

You currently need to have java installed on your system in order to use RJDBC which depends on rJava and jre and the jdk.

```
# sudo apt-get install default-jre default-jdk  
#R CMD javareconf -e # within the same shell that you run R
```

Building the *primate* package from source requires that you have the proper dependency packages, *caroline*, installed from CRAN. This can typically be accomplished via the following commands from within the R command line environment:

```
install.packages(c('caroline', 'RJDBC'))
```

After a successful installation the *primate* package can be loaded in the normal way: by starting R and invoking the following `library` command:

```
> library(primate)
```

### 3 Introduction

Data from All the World's Primates relational SQL database and other tabular datasets are made available via drivers and connection functions. Additionally we provide several functions and examples to facilitate the merging and aggregation of these tabular inputs.

### 4 Reading in Data

First we need to read the data into R using either locally cached datasets ("pkg.tab") or by using functions to access a remote SQL server mirror (commented out below). The local caches are a fraction of the full dataset and temporary as they will soon exceed the maximum allowable package size.

```
> primates.tab <- AWP.read.pkg.tab(tab.nm='dbo_tblGrovesMonkeys', id.clmn='MonkeyNumberGroves')
> heights.tab <- AWP.read.pkg.tab(tab.nm='HMeasure')
> locomo.tab <- AWP.read.pkg.tab(tab.nm='Locomotion')
> locomot.types <- AWP.read.pkg.tab(tab.nm='dbo_LMType')
> taxa.clmns <- c('Superfamily', 'Family', 'Genus', 'Species')
```

### 5 Aggregate and Merge

First we need to aggregate the data so that we have one record per species. We will use the `groupBy()` and `nerge()` functions (from the *caroline* package) as well as the `regroup.gnsp()` functions from the *primate* package.

```
> ### HEIGHTS ###
> getHMmeans <- function(code){
+   ret <- groupBy(df=heights.tab[heights.tab$HMTypeCd==code,], aggregation='mean', by='Primate')
+   names(ret) <- code; return(ret)
+ }
> heights <- list()
> levels <- c('GRND', 'G5', '510', '1020', '2030', '3040', '40+')
> for(lev in levels)
+   heights[[lev]] <- getHMmeans(lev) #not sure why these numbers are so high... are they correct?
> #names(heights) <- c(paste('h',names(heights)[1:6],sep=''), 'h40up')
>
> AWPheight <- nerge(heights)
> AWPheightsp <- nerge(list(h=tab2df(AWPheight), p=primates.tab[,taxa.clmns]))
> AWPheightgs <- regroup.gnsp(df=AWPheightsp,clmns=colnames(AWPheightsp)[1:7])
> names(AWPheightgs) <- c(sub('X', 'h',x=names(AWPheightgs)[-ncol(AWPheightgs)]), "h40up")
> h.levs <- names(AWPheightgs)
> ### LOCOMOTION ###
> AWPlocLst<-list()
> loc.modes <- c("BRAC", "CLIM", "LEAP", "QUAD", "SUSP")
```

```

> for(loc in loc.modes){
+                               dfsub=locomo.tab[locomo.tab$LTypeCd==loc,]
+   AWPlocLst[[loc]] <- groupBy(df=dfsub,aggregation='mean',by='PrimateID',clmns='PctOfTime
+ }
> AWPlocDF <- nerge(AWPlocLst, all=T)
> AWPlocDFsp <- nerge(list(AWPlocDF, primates.tab[,taxa.clmns]))
> ## add primate names and re-group by them instead (the ids are non-unique at the genus_spe
> loc.pct.modes <- paste("PctOfTime",loc.modes,sep='.')
> AWPlocDFgs <- regroup.gnsp(df=AWPlocDFsp,clmns=loc.pct.modes)
>

```

## 6 Merge

Finally we merge these two aggregated tables together and look for an interdisciplinary correlation.

```

> lh <- loc.high <- nerge(list(h=AWPheightgs,l=AWPlocDFgs))
> risky.modes <- loc.pct.modes[c(1,3)]
> tmp <- lapply(h.levs, function(x) {ret <- subset(loc.high,get(x)>3 ,risky.modes); names(re
> names(tmp)<- h.levs

> oldpar <- par(mfrow=c(1,length(tmp)), mar=c(2,2,3,0))
> sapply(names(tmp), function(i){boxplot(tmp[[i]], ylim=c(0,100), varwidth=T, main=i)})

      GRND      G5      h510      h1020
stats Numeric,10 Numeric,10 Numeric,10 Numeric,10
n      Numeric,2  Numeric,2  Numeric,2  Numeric,2
conf   Numeric,4  Numeric,4  Numeric,4  Numeric,4
out    Numeric,0  Numeric,0  Numeric,0  Numeric,0
group  Numeric,0  Numeric,0  Numeric,0  Numeric,0
names  Character,2 Character,2 Character,2 Character,2
      h2030      h3040      h40up
stats Numeric,10 Numeric,10 Numeric,10
n      Numeric,2  Numeric,2  Numeric,2
conf   Numeric,4  Numeric,4  Numeric,4
out    10.23333  Numeric,0  Numeric,0
group  2          Numeric,0  Numeric,0
names  Character,2 Character,2 Character,2

> par(oldpar)

```

## References

Noel Rowe and Marc Myers, (2017), "All the World's Primates", Pagonias Press, Charlestown RI

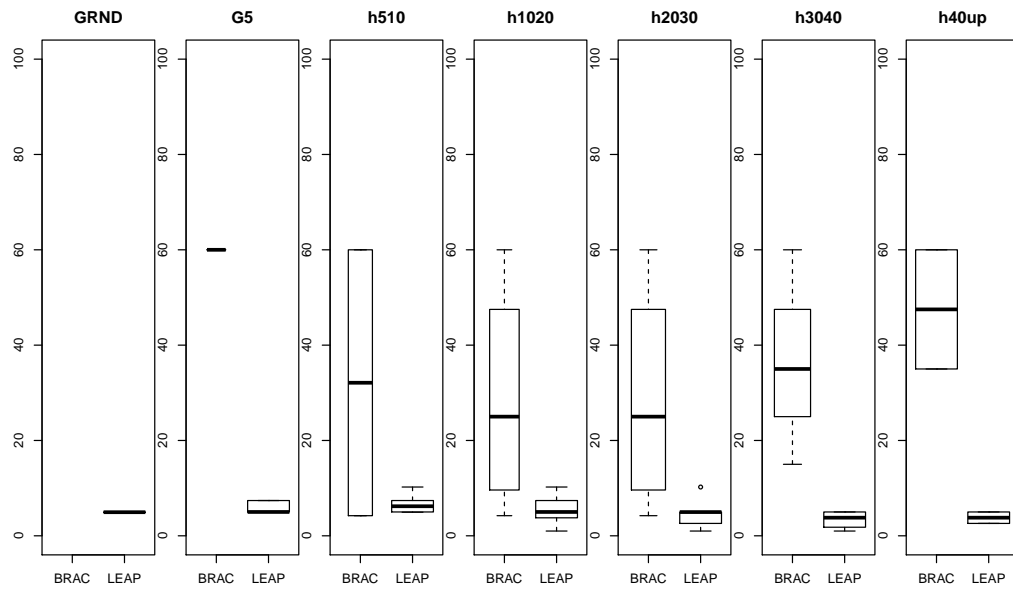


Figure 1: Brachiation and leaping frequencies by canopy height