

Assignment 6

Michelle

4/20/2020

```
#devtools::install_github("bomeara/geiger")  
library("ape")  
library("corHMM")
```

```
## Loading required package: nloptr
```

```
## Loading required package: GenSA
```

```
library("geiger")
```

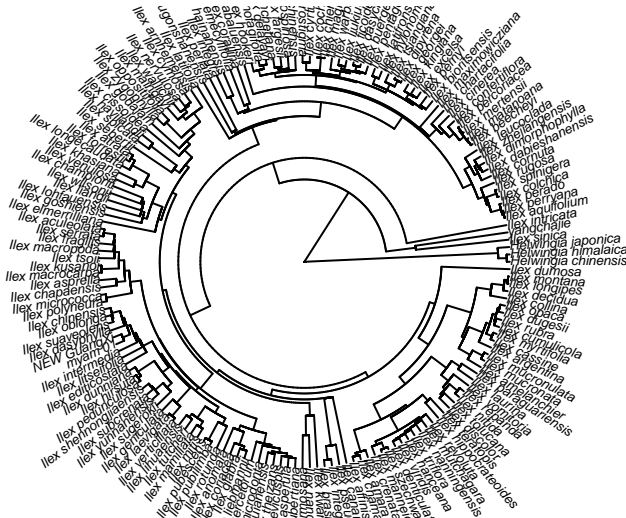
```
## Registered S3 method overwritten by 'geiger':  
##   method      from  
##   unique.multiPhylo ape
```

```
library("phangorn")  
library("openxlsx")  
library("rotl")  
library("phytools")
```

```
## Loading required package: maps
```

```
## Registered S3 method overwritten by 'phytools':  
##   method      from  
##   logLik.gfit geiger
```

```
hollies <- get_study_tree("ot_1984", "tree1")  
hollies.study.metadat<-get_study_meta("ot_1984")  
hollies.tree.ids<-get_tree_ids(hollies.study.metadat)  
plot(hollies, type="fan", cex = 0.4)
```

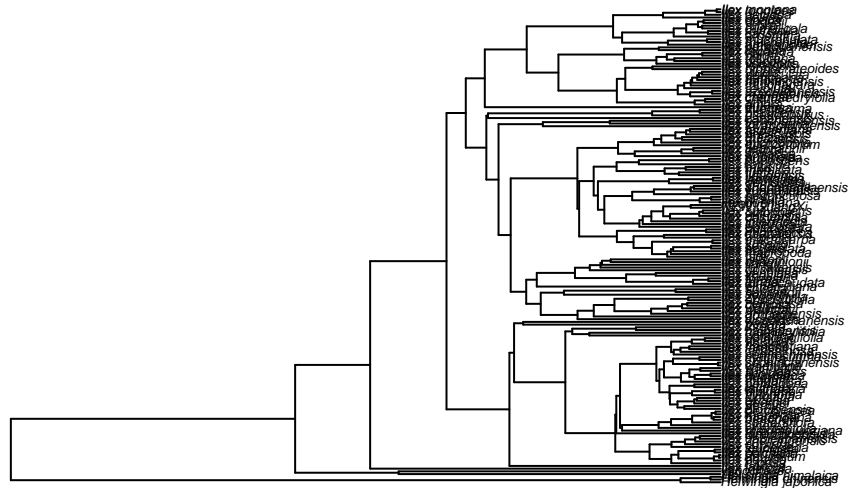


You'll need to get data into R in some way. Look at other phylometh assignments for how to #get trees and data.

```
hollies <- get_study_tree("ot_1984", "tree1")
print(hollies)
```

```
##
## Phylogenetic tree with 175 tips and 174 internal nodes.
##
## Tip labels:
## Helwingia_chinensis, Helwingia_himalaica, Helwingia_japonica, Ilex_sinica, yangchajie, Ilex_intrica
##
## Rooted; includes branch lengths.
```

```
hollies.study.metadata<-get_study_meta("ot_1984")
hollies.tree.ids<-get_tree_ids(hollies.study.metadata)
plot(hollies, cex = 0.4)
```



```
setwd("~/phylometh_exercises-/data\\ set\\ 6\\ and\\ 7")
```

```
hollies.data <- read.xlsx (xlsxFile = "hollies.xlsx", startRow = 2)
head(hollies.data )
```

```
##               species colour fruit.diameter.(mm) margin
## 1      Ilex aculeolata Nakai   black           7.0 serrate
## 2      Ilex affinis Gardner   black           4.0 serrate
## 3      Ilex amara Loes.       red             5.5 serrate
## 4 Ilex amelanchier M.A. Curtis ex Chapm.   red             7.5 entire
## 5      Ilex anomala Hook. & Arn. black           9.0 entire
## 6      Ilex aquifolium L.     red             8.0 spiny
## maximum.height.(m)
## 1      4
## 2      6
## 3      4
## 4      2
## 5      9
## 6     25
```

```
#Take the first 2 names and anytime there is an "_" it will split and put it together.
GetGenusSpeciesOneTaxon <- function(x) {
  x <- gsub("_", " ", x)
  x <- strsplit(x, "\\s+")[1]
  return(paste0(x[1:2], collapse="_"))
}
```

```

}
GetGenusSpeciesManyTaxa <- function(x) {
  return(unnname(apply(x, GetGenusSpeciesOneTaxon)))
}
# To get ride of the space and replace it with an "_"
hollies.data$species <- GetGenusSpeciesManyTaxa(gsub(" ", "_", hollies.data$species))

head(hollies.data )

```

```

##           species colour fruit.diameter.(mm)  margin maximum.height.(m)
## 1  Ilex_aculeolata  black              7.0 serrate                4
## 2    Ilex_affinis  black              4.0 serrate                6
## 3    Ilex_amara    red              5.5 serrate                4
## 4 Ilex_amelanchier  red              7.5  entire                2
## 5    Ilex_anomala  black              9.0  entire                9
## 6  Ilex_aquifolium  red              8.0   spiny               25

```

```

#To read csv files
#discrete.data <- read.csv(file= "hollies.xlsx", stringsAsFactors=FALSE)

```

#To get the color data

```

hollies.colour <- hollies.data$colour
names(hollies.colour) <- hollies.data$species
print(head(hollies.colour))

```

```

## Ilex_aculeolata    Ilex_affinis    Ilex_amara Ilex_amelanchier
##      "black"         "black"         "red"         "red"
##    Ilex_anomala Ilex_aquifolium
##      "black"         "red"

```

```

print(hollies.colour)

```

```

##      Ilex_aculeolata    Ilex_affinis    Ilex_amara
##      "black"         "black"         "red"
##    Ilex_amelanchier    Ilex_anomala    Ilex_aquifolium
##      "red"         "black"         "red"
##      Ilex_argentina    Ilex_arnhemensis    Ilex_asperula
##      "na"         "brown"         "na"
##      Ilex_asprella    Ilex_atrata    Ilex_beecheyi
##      "black"         "red"         "na"
##    Ilex_bioritsensis    Ilex_brasiliensis    Ilex_brevicuspis
##      "red"         "black"         "black"
##      Ilex_buergeri    Ilex_canariensis    Ilex_cassine
##      "red"         "red"         "red"
## Ilex_centrochinensis    Ilex_chamaedryfolia    Ilex_championii
##      "red"         "black"         "red"
##      Ilex_chapaensis    Ilex_chartacifolia    Ilex_chieniana
##      "black"         "red"         "red"
##      Ilex_chinensis    Ilex_ciliospinosa    Ilex_cinerea
##      "red"         "red"         "red"

```

##	Ilex_cissoidea	Ilex_cochinchinensis	Ilex_colchica
##	"na"	"red"	"red"
##	Ilex_collina	Ilex_confertiflora	Ilex_corallina
##	"red"	"na"	"red"
##	Ilex_coriacea	Ilex_cornuta	Ilex_crenata
##	"black"	"red"	"black"
##	Ilex_cumulicola	Ilex_cymosa	Ilex_cyrtura
##	"red"	"red"	"red"
##	Ilex_dabieshanensis	Ilex_dasyclada	Ilex_dasyphylla
##	"red"	"red"	"red"
##	Ilex_decidua	Ilex_delavayi	Ilex_denticulata
##	"black"	"red"	"black"
##	Ilex_dimorphophylla	Ilex_dipyrena	Ilex_discolor
##	"red"	"red"	"na"
##	Ilex_dugesii	Ilex_dumosa	Ilex_dunniana
##	"red"	"black"	"red"
##	Ilex_editicostata	Ilex_elmerrilliana	Ilex_excelsa
##	"red"	"red"	"red"
##	Ilex_fargesii	Ilex_ficoidea	Ilex_formosana
##	"red"	"red"	"red"
##	Ilex_forrestii	Ilex_fragilis	Ilex_franchetiana
##	"red"	"red"	"red"
##	Ilex_geniculata	Ilex_georgei	Ilex_glabra
##	"na"	"red"	"black"
##	Ilex_godajam	Ilex_goshiensis	Ilex_guianensis
##	"red"	"red"	"black"
##	Ilex_hainanensis	Ilex_hanceana	Ilex_havilandii
##	"na"	"red"	"na"
##	Ilex_hippocrateoides	Ilex_hirsuta	Ilex_hookeri
##	"na"	"na"	"na"
##	Ilex_hylonoma	Ilex_integerrima	Ilex_integra
##	"red"	"na"	"red"
##	Ilex_intermedia	Ilex_intricata	Ilex_khasiana
##	"red"	"red"	"na"
##	Ilex_kinabaluensis	Ilex_kingiana	Ilex_kusanoi
##	"red"	"red"	"na"
##	Ilex_kwangtungensis	Ilex_laevigata	Ilex_latifolia
##	"red"	"red"	"red"
##	Ilex_latifrons	Ilex_laurina	Ilex_leucoclada
##	"na"	"red"	"red"
##	Ilex_liangii	Ilex_liebmannii	Ilex_lihuaensis
##	"red"	"na"	"red"
##	Ilex_litseifolia	Ilex_liukiensis	Ilex_lohfauensis
##	"red"	"na"	"red"
##	Ilex_longicaudata	Ilex_longipes	Ilex_macrocarpa
##	"red"	"red"	"black"
##	Ilex_macropoda	Ilex_macrostigma	Ilex_maingayi
##	"red"	"na"	"na"
##	Ilex_mamillata	Ilex_manneiensis	Ilex_matanoana
##	"red"	"black"	"red"
##	Ilex_maximowicziana	Ilex_melanotricha	Ilex_memecylifolia
##	"na"	"red"	"red"
##	Ilex_mertensii	Ilex_micrococca	Ilex_microdonta
##	"red"	"black"	"red"

##	Ilex_mitis	Ilex_montana	Ilex_mucronulata
##	"red"	"red"	"red"
##	Ilex_mutchagara	Ilex_myrtifolia	Ilex_nanchuanensis
##	"black"	"red"	"red"
##	Ilex_nanningensis	Ilex_nervulosa	Ilex_nitida
##	"red"	"na"	"red"
##	Ilex_nothofagifolia	Ilex_oblonga	Ilex_opaca
##	"red"	"red"	"red"
##	Ilex_oppositifolia	Ilex_paraguariensis	Ilex_pedunculosa
##	"black"	"red"	"red"
##	Ilex_pentagona	Ilex_perado	Ilex_percoriacea
##	"na"	"red"	"red"
##	Ilex_perlata	Ilex_pernyi	Ilex_perryana
##	"na"	"red"	"red"
##	Ilex_polyneura	Ilex_pseudobuxus	Ilex_pubescens
##	"red"	"black"	"red"
##	Ilex_pubilimba	Ilex_quercetorum	Ilex_repanda
##	"red"	"na"	"na"
##	Ilex_revolvata	Ilex_rotunda	Ilex_rubra
##	"na"	"red"	"na"
##	Ilex_rugosa	Ilex_sebertii	Ilex_serrata
##	"red"	"red"	"red"
##	Ilex_shennongjiaensis	Ilex_sinica	Ilex_sp.
##	"red"	"red"	"brown"
##	Ilex_spicata	Ilex_spinigera	Ilex_suaveolens
##	"na"	"red"	"red"
##	Ilex_subcrenata	Ilex_subficoidea	Ilex_szechwanensis
##	"red"	"red"	"black"
##	Ilex_taubertiana	Ilex_teratopis	Ilex_theizans
##	"brown"	"black"	"red"
##	Ilex_tolucana	Ilex_triflora	Ilex_tsiangiana
##	"red"	"black"	"red"
##	Ilex_tsoi	Ilex_tutcheri	Ilex_venulosa
##	"black"	"red"	"red"
##	Ilex_verticillata	Ilex_viridis	Ilex_vomitoria
##	"red"	"black"	"red"
##	Ilex_warburgii	Ilex_wallichii	Ilex_wilsonii
##	"red"	"na"	"red"
##	Ilex_wugonshanensis	Ilex_yunnanensis	Ilex_zhejiangensis
##	"red"	"red"	"red"
##	Ilex_zygophylla		
##	"red"		

```

#table(hollies.colour)
#is.na(hollies.colour)
#na.omit(hollies.colour)
#Omit NA from data set
hollies.colour <- hollies.colour[!grepl("na", hollies.colour)]

```

Cleaning your data (Taxa and data match up)

Now write the code to use `CleanData()` to actually clean your data

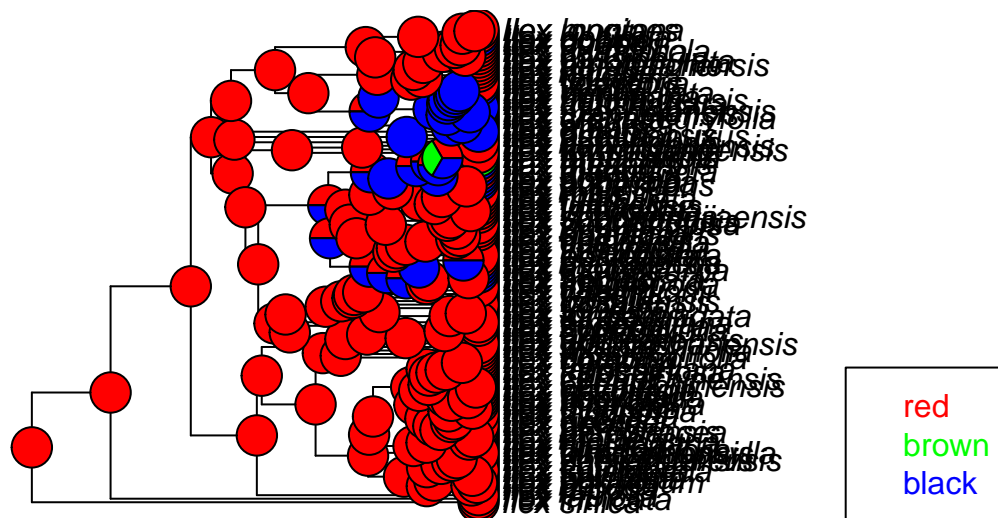
`#CleanData` <- function(phy,data) { `#return(treedata(phy, data))` `#`} `#Data` are often not right in some way. They might not match the taxa in your tree, there may `#be` missing data, etc. `#geiger::treedata` is a great function for getting a tree and data that `#match`, but your data may need other cleaning. `#Do` it as a function so it's repeatable.

```
CleanData <- function(phy, data) {  
  #result=treedata(phy, data,sort = TRUE, warnings = FALSE)  
  #return(result) or the function below  
  return(treedata(phy, data,sort = TRUE, warnings = FALSE))  
}  
hollies.cleaned<-CleanData(hollies, hollies.colour)  
print(str(hollies.cleaned))
```

```
## List of 2  
## $ phy :List of 4  
## ..$ edge      : int [1:268, 1:2] 136 136 137 137 138 139 139 140 141 142 ...  
## ..$ tip.label  : chr [1:135] "Ilex_sinica" "Ilex_intricata" "Ilex_aquifolium" "Ilex_perryana" ...  
## ..$ Nnode      : int 134  
## ..$ edge.length: num [1:268] 50.75 8.94 41.8 9.11 7.5 ...  
## ..- attr(*, "class")= chr "phylo"  
## ..- attr(*, "order")= chr "cladewise"  
## $ data: chr [1:135, 1] "red" "red" "red" "red" ...  
## ..- attr(*, "dimnames")=List of 2  
## .. ..$ : chr [1:135] "Ilex_sinica" "Ilex_intricata" "Ilex_aquifolium" "Ilex_perryana" ...  
## .. ..$ : NULL  
## NULL
```

`#VisualizeData` <- function(phy, data) { `#Important` here is to LOOK at your data before running it. Any weird values? Does it all make sense? What about `#your` `#tree`? Polytomies? `#` Now write the code to use `VisualizeData()` to actually look at your data `#`}

```
VisualizeData <- function(phy, data) {  
  dotTree(phy,data, fsize=0.5,lwd=2)  
}  
VisualizeData(phy=hollies.cleaned$phy,data=hollies.cleaned$data)
```

#Now, plot the likelihood estimates.

```
class(hollies.cleaned$data)
Dat<-as.phyDat(hollies.cleaned$data, type="USER", levels = c("black", "brown","red"))
anc.ml <- ancestral.pml(pml(hollies.cleaned$phy, Dat), type="ml")
plotAnc(hollies.cleaned$phy, anc.ml, 1)
print(Dat)
```

```
Dat<-as.phyDat(hollies.cleaned$data, type="USER", levels = c("black", "brown","red"))
```