Node labels on trees

Azul Carrillo

2022-11-03

The Data set the we are working with comes with thw TDbook r package

load the data packages

```
library("TDbook")
library(ggtree)
```

```
## ggtree v3.4.4 For help: https://yulab-smu.top/treedata-book/
##
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
##
## Guangchuang Yu.  Data Integration, Manipulation and Visualization of
## Phylogenetic Trees (1st edition). Chapman and Hall/CRC. 2022,
## doi:10.1201/9781003279242
##
## LG Wang, TTY Lam, S Xu, Z Dai, L Zhou, T Feng, P Guo, CW Dunn, BR
## Jones, T Bradley, H Zhu, Y Guan, Y Jiang, G Yu. treeio: an R package
## for phylogenetic tree input and output with richly annotated and
## associated data. Molecular Biology and Evolution. 2020, 37(2):599-603.
## doi: 10.1093/molbev/msz240
```

```
library(tidytree)
```

```
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## LG Wang, TTY Lam, S Xu, Z Dai, L Zhou, T Feng, P Guo, CW Dunn, BR
## Jones, T Bradley, H Zhu, Y Guan, Y Jiang, G Yu. treeio: an R package
## for phylogenetic tree input and output with richly annotated and
## associated data. Molecular Biology and Evolution. 2020, 37(2):599-603.
## doi: 10.1093/molbev/msz240
##
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
```

```
##
## Attaching package: 'tidytree'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

load the data from the data package suing the `data()` the data we are loading -tree_boots - df_tip_data -df_ inside_data

Exercise 1. What is the class of this three objects 2. How many elements does `tree_boots` has 3. How many rows does `df_tip_data` and `df_inode_data` have. 4. Compare the row numbers of both objects to the lengths of the `$tip.label` and `$node.label` elements of `tree_boots`.

```
data("tree_boots", "df_inode_data", "df_tip_data")

tree_boots
```

```
##
## Phylogenetic tree with 7 tips and 6 internal nodes.
##
## Tip labels:
##   Rangifer_tarandus, Cervus_elaphus, Bos_taurus, Ovis_orientalis, Suricata_suricatta, Cystophora_cris
## Node labels:
##   Mammalia, Artiodactyla, Cervidae, Bovidae, Carnivora, Caniformia
##
## Rooted; includes branch lengths.
```

```
class(tree_boots) #phylo  7 tips and 6 internal nodes.Rooted; includes branch lengths.
```

```
## [1] "phylo"
```

```
df_inode_data
```

```
##   newick_label      vernacularName                                infoURL
## 1     Mammalia             Mammals    http://eol.org/pages/1642/overview
## 2    Carnivora          Carnivores    http://eol.org/pages/7662/overview
## 3   Caniformia            Dog-like http://eol.org/pages/2849494/overview
## 4      Bovidae              Bovids    http://eol.org/pages/7687/overview
## 5     Cervidae             Cervids    http://eol.org/pages/7685/overview
## 6 Artiodactyla Even-toed ungulates    http://eol.org/pages/7678/overview
##       rank bootstrap posterior
## 1    class        NA        NA
## 2    order        96      0.89
## 3 suborder        98      0.93
## 4   family        99      0.95
## 5   family        98      0.96
## 6    order        92      0.81
```

```
class(df_inode_data) #data frame 6 rows 6 col
```

```
## [1] "data.frame"
```

```
nrow(df_inode_data)
```

```
## [1] 6
```

```
length(df_inode_data)
```

```
## [1] 6
```

```
df_tip_data
```

```
##          Newick_label  vernacularName
## 1   Rangifer_tarandus        Reindeer
## 2      Cervus_elaphus        Red deer
## 3          Bos_taurus          Cattle
## 4     Ovis_orientalis Asiatic mouflon
## 5  Suricata_suricatta         Meerkat
## 6 Cystophora_cristata     Hooded seal
## 7   Mephitis_mephitis   Striped skunk
##                                                 imageURL imageLicense
## 1  http://media.eol.org/content/2012/06/13/00/48543_orig.jpg     CC-BY-SA
## 2  http://media.eol.org/content/2014/09/16/00/20239_orig.jpg     CC-BY-SA
## 3 https://media.eol.org/content/2014/09/29/06/46535_orig.jpg     CC-BY-SA
## 4  http://media.eol.org/content/2015/05/20/03/80720_orig.jpg     CC-BY-SA
## 5  http://media.eol.org/content/2016/08/16/05/67138_orig.jpg     CC-BY-SA
## 6  http://media.eol.org/content/2013/06/18/07/63362_orig.jpg     CC-BY-SA
## 7  http://media.eol.org/content/2012/06/15/06/75234_orig.jpg        CC-BY
##                   imageAuthor                              infoURL mass_in_kg
## 1 Alexandre Buisse (Nattfodd) http://eol.org/pages/328653/overview     109.09
## 2                Sciadopitys http://eol.org/pages/328649/overview     240.87
## 3           Cynthia Sims Parr http://eol.org/pages/328699/overview     618.64
## 4                Jorg Hempel http://eol.org/pages/311906/overview      39.10
## 5               Sara&Joachim http://eol.org/pages/311580/overview       0.73
## 6      Ecomare, Salko de Wolf http://eol.org/pages/328632/overview     278.90
## 7                Kevin Bowman http://eol.org/pages/328593/overview       2.40
##   trophic_habit ncbi_taxid    rank
## 1     herbivore       9870 species
## 2     herbivore       9860 species
## 3     herbivore       9913 species
## 4     herbivore     469796 species
## 5     carnivore      37032 species
## 6      omnivore      39293 species
## 7      omnivore      30548 species
```

```
class(df_tip_data) #data frame 7 rows 10 col
```

```
## [1] "data.frame"
```

```
nrow(df_tip_data)
```

```
## [1] 7
```

```
length(df_tip_data)
```

```
## [1] 10
```

```
#joining data frameas and the tree
library(tidytree)
colnames(df_bar_data)
```

```
## [1] "id"              "dummy_bar_value"
```

```
df_tip_data$Newick_label
```

```
## [1] "Rangifer_tarandus"   "Cervus_elaphus"      "Bos_taurus"
## [4] "Ovis_orientalis"     "Suricata_suricatta"  "Cystophora_cristata"
## [7] "Mephitis_mephitis"
```

```
tree_boots$tip.label
```

```
## [1] "Rangifer_tarandus"   "Cervus_elaphus"      "Bos_taurus"
## [4] "Ovis_orientalis"     "Suricata_suricatta"  "Cystophora_cristata"
## [7] "Mephitis_mephitis"
```

```
#rename Newick_ label
df_mutate<- mutate(df_tip_data, label = Newick_label)
colnames(df_mutate)
```

```
##  [1] "Newick_label"   "vernacularName" "imageURL"       "imageLicense"
##  [5] "imageAuthor"    "infoURL"        "mass_in_kg"     "trophic_habit"
##  [9] "ncbi_taxid"     "rank"           "label"
```

```
tree_joined <- left_join(tree_boots, df_mutate, by = "label")


df_tip_data
```

```
##          Newick_label  vernacularName
## 1   Rangifer_tarandus        Reindeer
## 2      Cervus_elaphus        Red deer
## 3         Bos_taurus          Cattle
## 4     Ovis_orientalis Asiatic mouflon
## 5  Suricata_suricatta         Meerkat
## 6 Cystophora_cristata     Hooded seal
## 7   Mephitis_mephitis   Striped skunk
##                                                 imageURL imageLicense
## 1  http://media.eol.org/content/2012/06/13/00/48543_orig.jpg     CC-BY-SA
## 2  http://media.eol.org/content/2014/09/16/00/20239_orig.jpg     CC-BY-SA
## 3 https://media.eol.org/content/2014/09/29/06/46535_orig.jpg     CC-BY-SA
## 4  http://media.eol.org/content/2015/05/20/03/80720_orig.jpg     CC-BY-SA
## 5  http://media.eol.org/content/2016/08/16/05/67138_orig.jpg     CC-BY-SA
```

```
## 6  http://media.eol.org/content/2013/06/18/07/63362_orig.jpg     CC-BY-SA
## 7  http://media.eol.org/content/2012/06/15/06/75234_orig.jpg        CC-BY
##                   imageAuthor                                infoURL mass_in_kg
## 1 Alexandre Buisse (Nattfodd) http://eol.org/pages/328653/overview     109.09
## 2                 Sciadopitys http://eol.org/pages/328649/overview     240.87
## 3           Cynthia Sims Parr http://eol.org/pages/328699/overview     618.64
## 4                 Jorg Hempel http://eol.org/pages/311906/overview      39.10
## 5                Sara&Joachim http://eol.org/pages/311580/overview       0.73
## 6      Ecomare, Salko de Wolf http://eol.org/pages/328632/overview     278.90
## 7                 Kevin Bowman http://eol.org/pages/328593/overview       2.40
##   trophic_habit ncbi_taxid    rank
## 1     herbivore       9870 species
## 2     herbivore       9860 species
## 3     herbivore       9913 species
## 4     herbivore     469796 species
## 5     carnivore      37032 species
## 6      omnivore      39293 species
## 7      omnivore      30548 species
```

```
colnames(df_tip_data)-> my_colnames
my_colnames
```

```
##  [1] "Newick_label"   "vernacularName" "imageURL"       "imageLicense"
##  [5] "imageAuthor"    "infoURL"        "mass_in_kg"     "trophic_habit"
##  [9] "ncbi_taxid"     "rank"
```

```
typeof(my_colnames)#"character"
```

```
## [1] "character"
```

```
is.vector(my_colnames)
```

```
## [1] TRUE
```

```
my_colnames[1]<- "column 1 "
my_colnames[5]
```

```
## [1] "imageAuthor"
```

```
my_colnames
```

```
##  [1] "column 1 "      "vernacularName" "imageURL"       "imageLicense"
##  [5] "imageAuthor"    "infoURL"        "mass_in_kg"     "trophic_habit"
##  [9] "ncbi_taxid"     "rank"
```

```
my_colnames[10] <- "taxonomy"

colnames(df_tip_data)
```

```
##  [1] "Newick_label"   "vernacularName" "imageURL"       "imageLicense"
##  [5] "imageAuthor"    "infoURL"        "mass_in_kg"     "trophic_habit"
##  [9] "ncbi_taxid"     "rank"
```

```
colnames(df_tip_data)[1]<-"label"
colnames(df_tip_data)
```

```
##  [1] "label"          "vernacularName" "imageURL"       "imageLicense"
##  [5] "imageAuthor"    "infoURL"        "mass_in_kg"     "trophic_habit"
##  [9] "ncbi_taxid"     "rank"
```

```
tree_joined <- left_join(tree_boots, df_tip_data)
```

```
## Joining, by = "label"
```

```
tree_joined
```

```
## 'treedata' S4 object'.
##
## ...@ phylo:
##
## Phylogenetic tree with 7 tips and 6 internal nodes.
##
## Tip labels:
##   Rangifer_tarandus, Cervus_elaphus, Bos_taurus, Ovis_orientalis,
## Suricata_suricatta, Cystophora_cristata, ...
## Node labels:
##   Mammalia, Artiodactyla, Cervidae, Bovidae, Carnivora, Caniformia
##
## Rooted; includes branch lengths.
##
## with the following features available:
##   '', 'vernacularName', 'imageURL', 'imageLicense', 'imageAuthor', 'infoURL',
## 'mass_in_kg', 'trophic_habit', 'ncbi_taxid', 'rank'.
##
## # The associated data tibble abstraction: 13 x 12
## # The 'node', 'label' and 'isTip' are from the phylo tree.
##     node label     isTip verna~1 image~2 image~3 image~4 infoURL mass_~5 troph~6
##    <int> <chr>     <lgl> <chr>   <chr>   <chr>   <chr>   <chr>     <dbl> <chr>
##  1     1 Rangifer~ TRUE  Reinde~ http:/~ CC-BY-~ Alexan~ http:/~  109.   herbiv~
##  2     2 Cervus_e~ TRUE  Red de~ http:/~ CC-BY-~ Sciado~ http:/~  241.   herbiv~
##  3     3 Bos_taur~ TRUE  Cattle  https:~ CC-BY-~ Cynthi~ http:/~  619.   herbiv~
##  4     4 Ovis_ori~ TRUE  Asiati~ http:/~ CC-BY-~ Jorg H~ http:/~   39.1  herbiv~
##  5     5 Suricata~ TRUE  Meerkat http:/~ CC-BY-~ Sara&J~ http:/~    0.73 carniv~
##  6     6 Cystopho~ TRUE  Hooded~ http:/~ CC-BY-~ Ecomar~ http:/~  279.   omnivo~
##  7     7 Mephitis~ TRUE  Stripe~ http:/~ CC-BY   Kevin ~ http:/~    2.4  omnivo~
##  8     8 Mammalia  FALSE <NA>    <NA>    <NA>    <NA>    <NA>      NA    <NA>
##  9     9 Artiodac~ FALSE <NA>    <NA>    <NA>    <NA>    <NA>      NA    <NA>
## 10    10 Cervidae  FALSE <NA>    <NA>    <NA>    <NA>    <NA>      NA    <NA>
## # ... with 3 more rows, 2 more variables: ncbi_taxid <int>, rank <chr>, and
## #   abbreviated variable names 1: vernacularName, 2: imageURL, 3: imageLicense,
## #   4: imageAuthor, 5: mass_in_kg, 6: trophic_habit
```

```
tree_boots
```

```
##
## Phylogenetic tree with 7 tips and 6 internal nodes.
##
## Tip labels:
##   Rangifer_tarandus, Cervus_elaphus, Bos_taurus, Ovis_orientalis, Suricata_suricatta, Cystophora_cris
## Node labels:
##   Mammalia, Artiodactyla, Cervidae, Bovidae, Carnivora, Caniformia
##
## Rooted; includes branch lengths.
```

```
tree_joined@phylo
```

```
##
## Phylogenetic tree with 7 tips and 6 internal nodes.
##
## Tip labels:
##   Rangifer_tarandus, Cervus_elaphus, Bos_taurus, Ovis_orientalis, Suricata_suricatta, Cystophora_cris
## Node labels:
##   Mammalia, Artiodactyla, Cervidae, Bovidae, Carnivora, Caniformia
##
## Rooted; includes branch lengths.
```

```
tree_joined@extraInfo
```

```
## # A tibble: 13 x 10
##     node vernacu~1 image~2 image~3 image~4 infoURL mass_~5 troph~6 ncbi_~7 rank
##    <int> <chr>     <chr>   <chr>   <chr>   <chr>     <dbl> <chr>     <int> <chr>
##  1     1 Reindeer  http:/~ CC-BY-~ Alexan~ http:/~  109.   herbiv~    9870 spec~
##  2     2 Red deer  http:/~ CC-BY-~ Sciado~ http:/~  241.   herbiv~    9860 spec~
##  3     3 Cattle    https:~ CC-BY-~ Cynthi~ http:/~  619.   herbiv~    9913 spec~
##  4     4 Asiatic ~ http:/~ CC-BY-~ Jorg H~ http:/~   39.1  herbiv~  469796 spec~
##  5     5 Meerkat   http:/~ CC-BY-~ Sara&J~ http:/~    0.73 carniv~   37032 spec~
##  6     6 Hooded s~ http:/~ CC-BY-~ Ecomar~ http:/~  279.   omnivo~   39293 spec~
##  7     7 Striped ~ http:/~ CC-BY   Kevin ~ http:/~    2.4  omnivo~   30548 spec~
##  8     8 <NA>      <NA>    <NA>    <NA>    <NA>      NA    <NA>         NA <NA>
##  9     9 <NA>      <NA>    <NA>    <NA>    <NA>      NA    <NA>         NA <NA>
## 10    10 <NA>      <NA>    <NA>    <NA>    <NA>      NA    <NA>         NA <NA>
## 11    11 <NA>      <NA>    <NA>    <NA>    <NA>      NA    <NA>         NA <NA>
## 12    12 <NA>      <NA>    <NA>    <NA>    <NA>      NA    <NA>         NA <NA>
## 13    13 <NA>      <NA>    <NA>    <NA>    <NA>      NA    <NA>         NA <NA>
## # ... with abbreviated variable names 1: vernacularName, 2: imageURL,
## #   3: imageLicense, 4: imageAuthor, 5: mass_in_kg, 6: trophic_habit,
## #   7: ncbi_taxid
```

##plot node labels

```
df_tip_data
```

```
##                label  vernacularName
```

```
## 1   Rangifer_tarandus        Reindeer
## 2      Cervus_elaphus        Red deer
## 3          Bos_taurus          Cattle
## 4     Ovis_orientalis Asiatic mouflon
## 5  Suricata_suricatta         Meerkat
## 6 Cystophora_cristata     Hooded seal
## 7   Mephitis_mephitis   Striped skunk
##                                                 imageURL imageLicense
## 1  http://media.eol.org/content/2012/06/13/00/48543_orig.jpg     CC-BY-SA
## 2  http://media.eol.org/content/2014/09/16/00/20239_orig.jpg     CC-BY-SA
## 3 https://media.eol.org/content/2014/09/29/06/46535_orig.jpg     CC-BY-SA
## 4  http://media.eol.org/content/2015/05/20/03/80720_orig.jpg     CC-BY-SA
## 5  http://media.eol.org/content/2016/08/16/05/67138_orig.jpg     CC-BY-SA
## 6  http://media.eol.org/content/2013/06/18/07/63362_orig.jpg     CC-BY-SA
## 7  http://media.eol.org/content/2012/06/15/06/75234_orig.jpg        CC-BY
##                   imageAuthor                              infoURL mass_in_kg
## 1 Alexandre Buisse (Nattfodd) http://eol.org/pages/328653/overview     109.09
## 2                Sciadopitys http://eol.org/pages/328649/overview     240.87
## 3           Cynthia Sims Parr http://eol.org/pages/328699/overview     618.64
## 4                Jorg Hempel http://eol.org/pages/311906/overview      39.10
## 5               Sara&Joachim http://eol.org/pages/311580/overview       0.73
## 6      Ecomare, Salko de Wolf http://eol.org/pages/328632/overview     278.90
## 7                Kevin Bowman http://eol.org/pages/328593/overview       2.40
##   trophic_habit ncbi_taxid    rank
## 1     herbivore       9870 species
## 2     herbivore       9860 species
## 3     herbivore       9913 species
## 4     herbivore     469796 species
## 5     carnivore      37032 species
## 6      omnivore      39293 species
## 7      omnivore      30548 species
```

```
df_inode_data
```

```
##   newick_label      vernacularName                                infoURL
## 1     Mammalia             Mammals    http://eol.org/pages/1642/overview
## 2    Carnivora          Carnivores    http://eol.org/pages/7662/overview
## 3   Caniformia            Dog-like http://eol.org/pages/2849494/overview
## 4      Bovidae              Bovids    http://eol.org/pages/7687/overview
## 5     Cervidae             Cervids    http://eol.org/pages/7685/overview
## 6 Artiodactyla Even-toed ungulates    http://eol.org/pages/7678/overview
##       rank bootstrap posterior
## 1    class        NA        NA
## 2    order        96      0.89
## 3 suborder        98      0.93
## 4   family        99      0.95
## 5   family        98      0.96
## 6    order        92      0.81
```

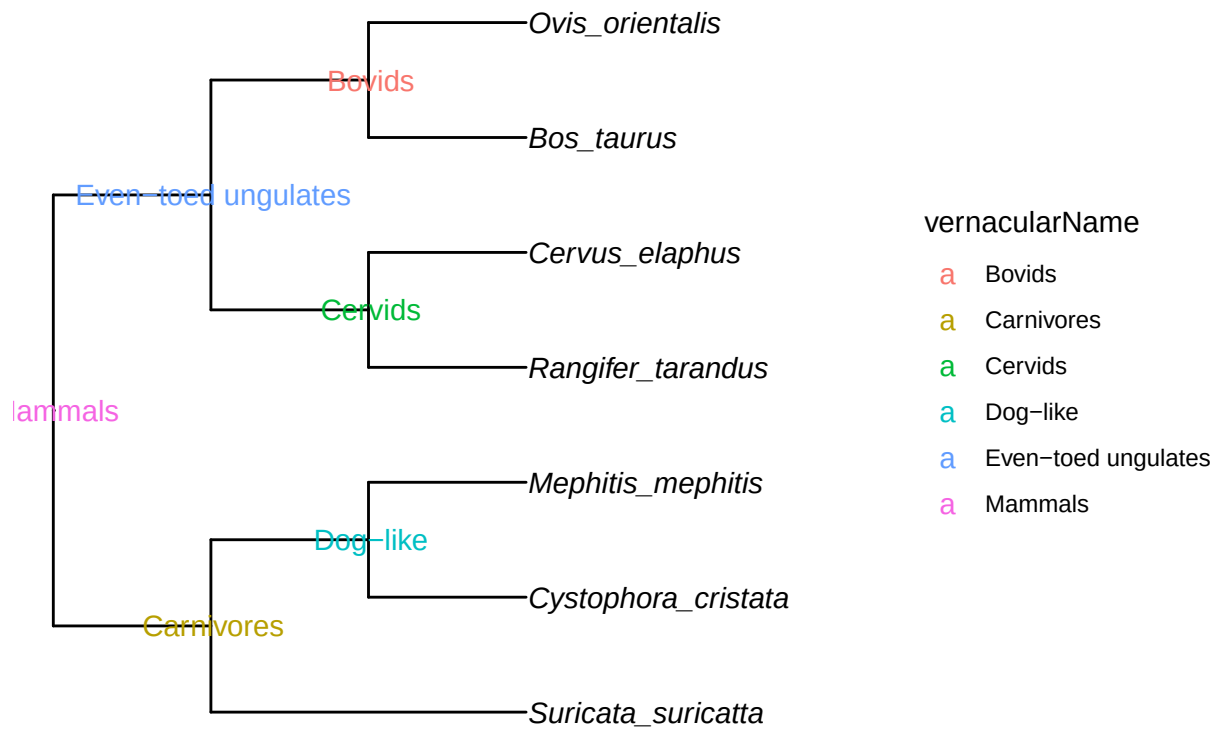
```
colnames(df_inode_data) [1]<-"label"
tree_data2 <- left_join(tree_boots, df_inode_data)
```

```
## Joining, by = "label"
```

```
tree_data2
```

```
## 'treedata' S4 object'.
##
## ...@ phylo:
##
## Phylogenetic tree with 7 tips and 6 internal nodes.
##
## Tip labels:
##   Rangifer_tarandus, Cervus_elaphus, Bos_taurus, Ovis_orientalis,
## Suricata_suricatta, Cystophora_cristata, ...
## Node labels:
##   Mammalia, Artiodactyla, Cervidae, Bovidae, Carnivora, Caniformia
##
## Rooted; includes branch lengths.
##
## with the following features available:
##   '', 'vernacularName', 'infoURL', 'rank', 'bootstrap', 'posterior'.
##
## # The associated data tibble abstraction: 13 x 8
## # The 'node', 'label' and 'isTip' are from the phylo tree.
##     node label               isTip vernacularName  infoURL rank  boots~1 poste~2
##    <int> <chr>               <lgl> <chr>           <chr>   <chr>   <int>   <dbl>
##  1     1 Rangifer_tarandus   TRUE  <NA>            <NA>    <NA>       NA   NA
##  2     2 Cervus_elaphus      TRUE  <NA>            <NA>    <NA>       NA   NA
##  3     3 Bos_taurus          TRUE  <NA>            <NA>    <NA>       NA   NA
##  4     4 Ovis_orientalis     TRUE  <NA>            <NA>    <NA>       NA   NA
##  5     5 Suricata_suricatta  TRUE  <NA>            <NA>    <NA>       NA   NA
##  6     6 Cystophora_cristata TRUE  <NA>            <NA>    <NA>       NA   NA
##  7     7 Mephitis_mephitis   TRUE  <NA>            <NA>    <NA>       NA   NA
##  8     8 Mammalia            FALSE Mammals         http:/~ class      NA   NA
##  9     9 Artiodactyla        FALSE Even-toed ungu~ http:/~ order      92    0.81
## 10    10 Cervidae            FALSE Cervids         http:/~ fami~      98    0.96
## # ... with 3 more rows, and abbreviated variable names 1: bootstrap,
## #   2: posterior
```

```
ggtree(tree_data2)+
  geom_tiplab(fontface = "italic")+
  xlim(0,5) +
  geom_nodelab(mapping = aes(label = vernacularName, color = vernacularName))
```




```
portal_tree <-("../raw-data/portal-tree.tre")
```

```
taxonomy <- read.csv("../raw-data/portal-species-taxonomy.csv")
surveys <- read.csv("../raw-data/surveys.csv")
species <- read.csv("../raw-data/species.csv")
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:ggtree':
##
##     inset
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggtree)
library(tidytree)
surveys %>%
  filter(!is.na(weight)) %>%
  group_by(species_id) %>%
  summarise(mean_weight = mean(weight)) -> species_weight

surveys %>%
  filter(!is.na(hindfoot_length))%>%
  group_by(species_id) %>%
  summarise(mean_hindfoot_length =mean(hindfoot_length)) -> species_hindfoot_length

average_data <- full_join(species_weight, species_hindfoot_length)
```

```
## Joining, by = "species_id"
```

```
head(average_data)
```

```
## # A tibble: 6 x 3
##   species_id mean_weight mean_hindfoot_length
##   <chr>            <dbl>                <dbl>
## 1 BA                8.6                 13
## 2 DM               43.2                 36.0
## 3 DO               48.9                 35.6
## 4 DS              120.                  49.9
## 5 NL              159.                  32.3
## 6 OL               31.6                 20.5
```

```
#joined both averages and the taxonomy
taxonomy_average<- full_join(taxonomy, average_data)
```

```
## Joining, by = "species_id"
```

```
colnames(taxonomy_average)
```

```
##  [1] "label"               "species_id"          "genus"
##  [4] "species"             "taxa"                "ott_name"
##  [7] "approximate_match"   "ott_id"              "is_synonym"
## [10] "flags"               "mean_weight"         "mean_hindfoot_length"
```

```
library(dplyr)
portal_tree <-read.tree("../raw-data/portal-tree.tre")
portal_joined <- left_join(portal_tree, taxonomy_average)
```

```
## Joining, by = "label"
```

```
colnames(portal_joined)
```

```
## NULL
```

```
ggtree(portal_joined)+
  geom_tiplab( fontface = "italic",size= 4) +
  xlim(0,20) +
  geom_nodelab()
```