

# Data preparation

from Leite et al.

Processing the datasets to create a wide formatting data for analysis, for each assemblage.

## Species X sites occurrence matrix

```
data <- read.csv("data/species_site_occ2.csv", check.names=FALSE)
```

Long format

```
data <- data %>% pivot_longer(cols = 3:182, names_to = "sp", values_to="occor")
str(data)
```

```
## tibble [16,560 x 4] (S3: tbl_df/tbl/data.frame)
## $ landscape: chr [1:16560] "148" "148" "148" "148" ...
## $ site      : chr [1:16560] "148.P10" "148.P10" "148.P10" "148.P10" ...
## $ sp        : chr [1:16560] "Amazilia_lact" "Amazon_aest" "Anabaz_fuscus" "Aphant_cirr" ...
## $ occor      : int [1:16560] 0 0 0 0 0 0 0 0 0 0 ...
```

## Landscape and sites information

```
env <- read.csv("data/sites_information.csv")
str(env)
```

```
## 'data.frame':    92 obs. of  12 variables:
## $ matrix      : chr  "high_quality" "high_quality" "high_quality" "high_quality" ...
## $ landscape    : chr  "P02" "P02" "P02" "P02" ...
## $ site         : chr  "P02.P00" "P02.P01" "P02.P02" "P02.P03" ...
## $ lat_land     : num  -21.9 -21.9 -21.9 -21.9 -21.6 ...
## $ lon_land     : num  -46.7 -46.7 -46.7 -46.7 -46.5 ...
## $ lat_site     : num  -21.9 -21.9 -21.9 -21.9 -21.6 ...
## $ lon_site     : num  -46.7 -46.7 -46.7 -46.7 -46.5 ...
## $ n.visit      : int   4 4 4 4 4 4 4 4 4 4 ...
## $ forest_land  : num   43.5 43.5 43.5 43.5 12 12 12 12 27.6 27.6 ...
## $ forest_site400: num   72.3 45.6 58.3 62.5 24.2 ...
## $ forest_site600: num   61.7 44.7 54.3 56.3 16.6 ...
```

```
## $ forest_site800: num 57.5 43.3 51.4 51.6 16.8 ...
```

## Species information

```
sp <- read.csv("data/species_information.csv")
str(sp)
```

```
## 'data.frame': 180 obs. of 11 variables:
## $ common_name : chr "Sapphire-spangled Emerald" "Turquoise-fronted Amazon" "White-collared" ...
## $ species_name : chr "Amazilia lactea" "Amazona aestiva" "Anabazenops fuscus" "Aphantochroa" ...
## $ sp : chr "Amazilia_lact" "Amazon_aest" "Anabaz_fuscus" "Aphant_cirr" ...
## $ habitat_specif: chr "generalist" "generalist" "specialist" "generalist" ...
## $ body_size : num 4.3 400 39 7.3 540 ...
## $ diet : chr "nectarivorous" "frugivorous" "insectivorous" "nectarivorous" ...
## $ nest : chr "open_semi" "cavities" "cavities" "open_semi" ...
## $ stratum : chr "all" "mid_canopy" "ground_under" "all" ...
## $ frugivory : int 0 50 0 0 10 10 30 50 40 40 ...
## $ insectivory : int 10 0 100 10 70 50 10 0 30 30 ...
## $ lower_stratum : int 30 30 30 80 70 50 40 80 100 100 ...
```

## Mergin information

```
dataset <- data %>% left_join(sp[,3:11], "sp") %>%
  left_join(env[,],by=c("landscape", "site"))
```

## Creating assemblages datasets

Cleaning information columns and standardazing continuous variables for modeling.

```
high.spe <- dataset %>% filter(matrix=="high_quality",
                              habitat_specif=="specialist") %>%

  select(landscape,site,sp,n.visit,occor,body_size,diet,nest,stratum,
         frugivory,insectivory,lower_stratum,forest_land,
         forest_site400, forest_site600, forest_site800) %>%
  mutate(lbody_size = as.numeric(scale(log(body_size))),
         frugivoryOrig = frugivory,
         frugivory = as.numeric(scale(frugivory)),
         insectivoryOrig = insectivory,
         insectivory = as.numeric(scale(insectivory)),
         lower_stratumOrig = lower_stratum,
         lower_stratum = as.numeric(scale(lower_stratum)),
```

```

forest_landOrig = forest_land,
forest_land = as.numeric(scale(forest_land)),
forest_site400Orig = forest_site400,
forest_site400 = as.numeric(scale(forest_site400)),
forest_site600Orig = forest_site600,
forest_site600 = as.numeric(scale(forest_site600)),
forest_site800Orig = forest_site800,
forest_site800 = as.numeric(scale(forest_site800)),
) %>% as.data.frame()

```

```

high.gen <- dataset %>% filter(matrix=="high_quality",
                               habitat_specif=="generalist")%>%
  select(landscape,site,sp,n.visit,occor,body_size,diet,nest,stratum,
         frugivory,insectivory,lower_stratum,forest_land,
         forest_site400, forest_site600, forest_site800)%>%
  mutate(lbody_size = as.numeric(scale(log(body_size))),
         frugivoryOrig = frugivory,
         frugivory = as.numeric(scale(frugivory)),
         insectivoryOrig = insectivory,
         insectivory = as.numeric(scale(insectivory)),
         lower_stratumOrig = lower_stratum,
         lower_stratum = as.numeric(scale(lower_stratum)),
         forest_landOrig = forest_land,
         forest_land = as.numeric(scale(forest_land)),
         forest_site400Orig = forest_site400,
         forest_site400 = as.numeric(scale(forest_site400)),
         forest_site600Orig = forest_site600,
         forest_site600 = as.numeric(scale(forest_site600)),
         forest_site800Orig = forest_site800,
         forest_site800 = as.numeric(scale(forest_site800)),
         ) %>% as.data.frame()

```

```

low.spe <- dataset %>% filter(matrix=="low_quality",
                               habitat_specif=="specialist")%>%
  select(landscape,site,sp,n.visit,occor,body_size,diet,nest,stratum,
         frugivory,insectivory,lower_stratum,forest_land,
         forest_site400, forest_site600, forest_site800)%>%
  mutate(lbody_size = as.numeric(scale(log(body_size))),
         frugivoryOrig = frugivory,

```

```

frugivory = as.numeric(scale(frugivory)),
insectivoryOrig = insectivory,
insectivory = as.numeric(scale(insectivory)),
lower_stratumOrig = lower_stratum,
lower_stratum = as.numeric(scale(lower_stratum)),
forest_landOrig = forest_land,
forest_land = as.numeric(scale(forest_land)),
forest_site400Orig = forest_site400,
forest_site400 = as.numeric(scale(forest_site400)),
forest_site600Orig = forest_site600,
forest_site600 = as.numeric(scale(forest_site600)),
forest_site800Orig = forest_site800,
forest_site800 = as.numeric(scale(forest_site800)),
) %>% as.data.frame()

```

```

low.gen <- dataset %>% filter(matrix=="low_quality",
                             habitat_specif=="generalist")%>%
select(landscape,site,sp,n.visit,occur,body_size,diet,nest,stratum,
       frugivory,insectivory,lower_stratum,forest_land,
       forest_site400, forest_site600, forest_site800)%>%
mutate(lbody_size = as.numeric(scale(log(body_size))),
       frugivoryOrig = frugivory,
       frugivory = as.numeric(scale(frugivory)),
       insectivoryOrig = insectivory,
       insectivory = as.numeric(scale(insectivory)),
       lower_stratumOrig = lower_stratum,
       lower_stratum = as.numeric(scale(lower_stratum)),
       forest_landOrig = forest_land,
       forest_land = as.numeric(scale(forest_land)),
       forest_site400Orig = forest_site400,
       forest_site400 = as.numeric(scale(forest_site400)),
       forest_site600Orig = forest_site600,
       forest_site600 = as.numeric(scale(forest_site600)),
       forest_site800Orig = forest_site800,
       forest_site800 = as.numeric(scale(forest_site800)),
) %>% as.data.frame()

```

Selecting only the species that occur in the landscapes

```

sp.high.spe <- high.spe %>% group_by(sp) %>% summarise(n=sum(occor)) %>%
  filter(n>0) %>% select(sp)
high.spe <- high.spe %>% filter(sp %in% sp.high.spe$sp)

sp.high.gen <- high.gen %>% group_by(sp) %>% summarise(n=sum(occor)) %>%
  filter(n>0) %>% select(sp)
high.gen <- high.gen %>% filter(sp %in% sp.high.gen$sp)

sp.low.spe <- low.spe %>% group_by(sp) %>% summarise(n=sum(occor)) %>%
  filter(n>0) %>% select(sp)
low.spe <- low.spe %>% filter(sp %in% sp.low.spe$sp)

sp.low.gen <- low.gen %>% group_by(sp) %>% summarise(n=sum(occor)) %>%
  filter(n>0) %>% select(sp)
low.gen <- low.gen %>% filter(sp %in% sp.low.gen$sp)

```

## Summary species richness

Total richness specialists and generalists in each matrix quality type.

```

riq.high.spe <- length(unique(high.spe$sp))
riq.low.spe <- length(unique(low.spe$sp))
riq.spe <- length(unique(c(unique(low.spe$sp), unique(high.spe$sp))))
shared.spe <- riq.high.spe + riq.low.spe - riq.spe

riq.high.gen <- length(unique(high.gen$sp))
riq.low.gen <- length(unique(low.gen$sp))
riq.gen <- length(unique(c(unique(low.gen$sp), unique(high.gen$sp))))
shared.gen <- riq.high.gen + riq.low.gen - riq.gen

riq.tot <- riq.gen + riq.spe

```

Richness per landscape and site

```

# landscape
lcc.spe <- high.spe %>% filter(occor>0) %>%
  group_by(landscape, sp) %>%
  summarise(riq = n()) %>%
  group_by(landscape) %>%
  summarise(riq=n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),

```

```

        median = median(riq),
        max=max(riq), min=min(riq))
lpp.spe <- low.spe %>% filter(occor>0) %>%
  group_by(landscape, sp) %>%
  summarise(riq = n()) %>%
  group_by(landscape) %>%
  summarise(riq=n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
land.sum <- rbind(lcc.spe, lpp.spe) %>% mutate(matrix = c("High", "low"))

# site
scc.spe <- high.spe %>% filter(occor>0) %>%
  group_by(landscape, site, forest_site400) %>%
  summarise(riq = n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
spp.spe <- low.spe %>% filter(occor>0) %>%
  group_by(landscape, site, forest_site400) %>%
  summarise(riq = n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
site.sum <- rbind(scc.spe, spp.spe) %>% mutate(matrix = c("High", "Low"))

# generalistas

# landscape
lcc.gen <- high.gen %>% filter(occor>0) %>%
  group_by(landscape, sp) %>%
  summarise(riq = n()) %>%
  group_by(landscape) %>%
  summarise(riq=n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
lpp.gen <- low.gen %>% filter(occor>0) %>%

```

```

group_by(landscape, sp) %>%
  summarise(riq = n()) %>%
group_by(landscape) %>%
  summarise(riq=n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
gland.sum <- rbind(lcc.gen, lpp.gen) %>% mutate(matrix = c("High", "Low"))

# site
scc.gen <- high.gen %>% filter(occor>0) %>%
  group_by(landscape, site, forest_site400) %>%
  summarise(riq = n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
spp.gen <- low.gen %>% filter(occor>0) %>%
  group_by(landscape, site, forest_site400) %>%
  summarise(riq = n()) %>% ungroup() %>%
  summarise(mean = mean(riq), desv.pad = sd(riq),
            median = median(riq),
            max=max(riq), min=min(riq))
gsite.sum <- rbind(scc.gen, spp.gen) %>% mutate(matrix = c("High", "Low"))

```

## Summary trait between specialists and generalists

```

bhigh.spe <- high.spe %>% group_by(sp) %>% summarise(body_size = mean(body_size),
                                                    frugi = mean(frugivoryOrig),
                                                    inse = mean(insectivoryOrig),
                                                    grou = mean(lower_stratumOrig),
                                                    nest = first(nest),
                                                    diet = first(diet),
                                                    stratum =first(stratum),
                                                    ) %>%
  mutate(dataset="high.spe") %>%
  as.data.frame()
blow.spe <- low.spe %>% group_by(sp) %>% summarise(body_size = mean(body_size),
                                                    frugi = mean(frugivoryOrig),
                                                    inse = mean(insectivoryOrig),

```

```

        grou = mean(lower_stratumOrig),
        nest = first(nest),
        diet = first(diet),
        stratum =first(stratum),
    ) %>%

mutate(dataset="low.spe") %>%
as.data.frame()

bhigh.gen <- high.gen %>% group_by(sp) %>% summarise(body_size = mean(body_size),
        frugi = mean(frugivoryOrig),
        inse = mean(insectivoryOrig),
        grou = mean(lower_stratumOrig),
        nest = first(nest),
        diet = first(diet),
        stratum =first(stratum),
    ) %>%

mutate(dataset="high.gen") %>%
as.data.frame()

blow.gen <- low.gen %>% group_by(sp) %>% summarise(body_size = mean(body_size),
        frugi = mean(frugivoryOrig),
        inse = mean(insectivoryOrig),
        grou = mean(lower_stratumOrig),
        nest = first(nest),
        diet = first(diet),
        stratum =first(stratum),
    ) %>%

mutate(dataset="low.gen") %>%
as.data.frame()

bares <- rbind(bhigh.spe, blow.spe, bhigh.gen, blow.gen)
colnames(bares)[2:8] <- c("body mass", "% fruits", "% insects", "% lower strata",
        "nest type", "main diet", "foraging strata")

nbares <- bares %>% separate(dataset, c("matrix", "habitat")) %>% select(-matrix) %>%
    distinct() %>%
    mutate(`body mass` = log(`body mass`)) %>%
    mutate_at(c("body mass", "% fruits", "% insects", "% lower strata"), scale)

nbares %>% gather("trait", "value", 2:5) %>%
    ggplot(aes(x=habitat, y=value)) + geom_boxplot() +
    facet_wrap(~trait) +

```



```
scale_x_discrete(name = "", labels=c("Specialists", "Generalists"))
```

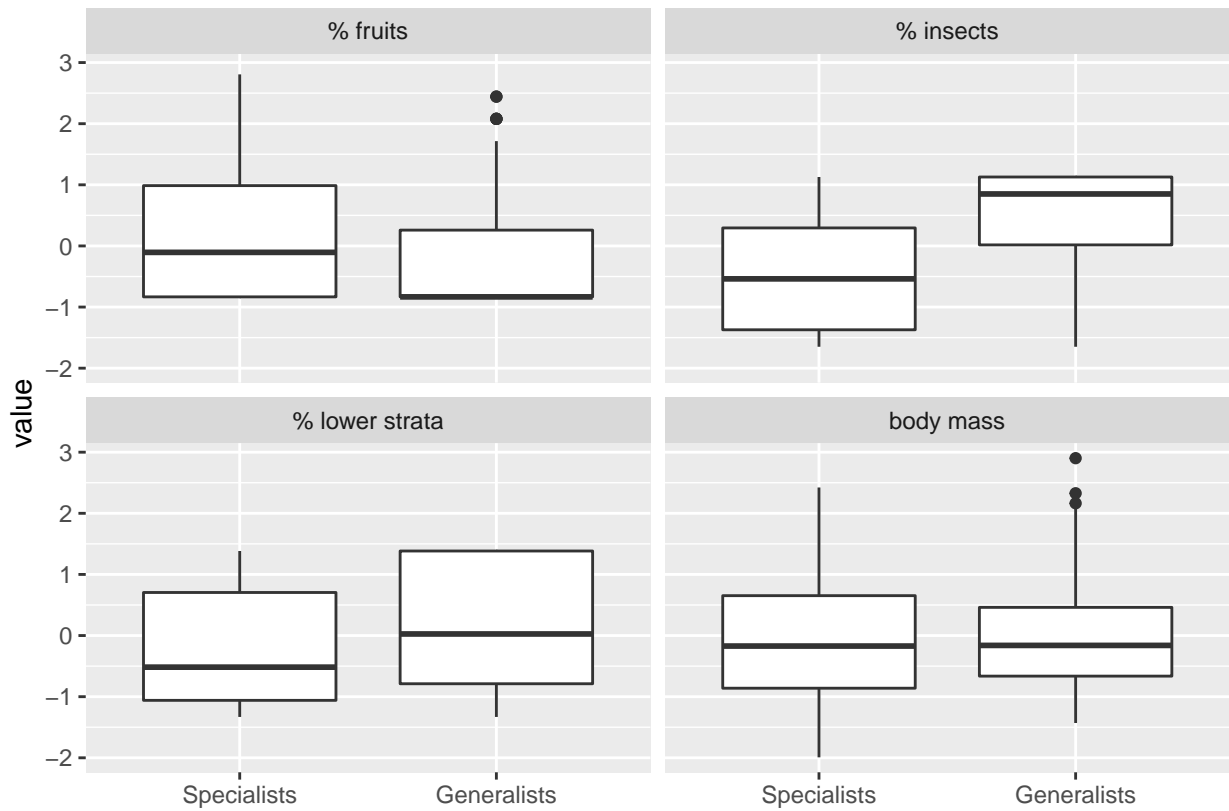


Figure 1: Boxplots of values for the traits measured as continuous variables for the specialists and generalists. Species from the two regions were grouped.

```
tab1 <- nbares %>% tabyl(`nest type`, habitat) %>%
  adorn_percentages("col") %>%
  adorn_pct_formatting(digits = 0) %>% adorn_ns() %>% rename(var=`nest type`)
tab2 <- nbares %>% tabyl(`main diet`, habitat) %>%
  adorn_percentages("col") %>%
  adorn_pct_formatting(digits = 0) %>% adorn_ns() %>% rename(var=`main diet`)
tab3 <- nbares %>% tabyl(`foraging strata`, habitat) %>%
  adorn_percentages("col") %>%
  adorn_pct_formatting(digits = 0) %>% adorn_ns() %>% rename(var=`foraging strata`)

rbind(tab1,tab2,tab3) %>% rename(Traits = var, Specialists = spe, Generalists = gen) %>%
  kable()
```

Traits	Generalists	Specialists
cavities	22% (19)	25% (23)
closed	21% (18)	27% (25)
open_semi	57% (50)	48% (45)
frugivorous	23% (20)	17% (16)
granivorous	6% (5)	2% (2)

Traits	Generalists	Specialists
insectivorous	40% (35)	77% (72)
nectarivorous	15% (13)	0% (0)
onivorous	16% (14)	3% (3)
all	20% (17)	5% (5)
ground_under	32% (28)	58% (54)
mid_canopy	48% (42)	37% (34)

## Saving datasets

```
save(env, high.spe, high.gen, low.spe, low.gen, file="datasets.Rdata")
```