

MalNAF Part 1

Rachael Holmes

13/04/2022

Analysis of composition, native ranges, habitat preference and FI

Step 1. Load Libraries

```
library(tidyverse)
library(readr)
library(magrittr)
library(geojsonio)
library(broom)
library(dplyr)
library(viridis)
library(RColorBrewer)
library(sf)
library(readr)
library(classInt)
library(ggpubr)
library(lcvplants)
library(LCPV)
library(hrbrthemes)
library(knitr)
```

Step 2. Import data

Data are for 10 island groups in Malesia selected as they are distinct phytogeographic units.

Detail available in the methods on creation of the datasets.

```
MalNAF <- read_csv("MalNAF_Naturalized.csv")
Native_range <- read_csv("MalNAF_Native_range.csv")
Malesia_environmental <- read_csv("Malesia_environmental_data.csv")
hab <- read_csv("MalNAF_SynHab.csv")
```

How many taxa/species in each status and what is richness of different islands.

```
# total unique naturalized taxa and species in dataset after harmonization
```

```
MalNAF %$%
  n_distinct(Accepted_Taxon) # taxa
```

```
## [1] 1363
```

```
MalNAF %$%  
  n_distinct(Accepted_binomial) # binomial
```

```
## [1] 1345
```

```
# dataset has multiple entries for some taxa due to multiple references  
# simplify data into single entries to get p/a as binary
```

```
naturalize_taxa <- MalNAF %>%  
  arrange(region_id, Accepted_Taxon)%>%  
  group_by(region_id, Accepted_Taxon) %>%  
  slice_head() # 1 entry per species x region  
  
head(naturalize_taxa)
```

```
## # A tibble: 6 x 16  
## # Groups:   region_id, Accepted_Taxon [6]  
##   taxon_orig_auth    taxon_orig author_name_orig family_orig status_orig country  
##   <chr>             <chr>      <chr>           <chr>      <chr>      <chr>  
## 1 Acacia auriculifo~ Acacia au~ Benth.         Leguminosae invasive   BN  
## 2 Acacia holoserice~ Acacia ho~ G.Don          Leguminosae naturalized BN  
## 3 Acacia mangium Wi~ Acacia ma~ Willd          Fabaceae    invasive   ID  
## 4 Acmeilla uliginosa Acmeilla u~ <NA>           <NA>        naturalized ID  
## 5 Acrocarpus fraxin~ Acrocarpu~ <NA>           <NA>        invasive   ID  
## 6 Aeschynomene indi~ Aeschynom~ <NA>           <NA>        naturalized ID  
## # ... with 10 more variables: region_id <chr>, reference <chr>, source <dbl>,  
## #   Accepted_Family <chr>, Accepted_genus <chr>, Status <chr>,  
## #   Accepted_Taxon <chr>, Accepted_binomial <chr>, Dates <chr>,  
## #   final_status <chr>
```

```
# but to get a dataset for species only need to use accepted_binomial
```

```
naturalize_sp <- MalNAF %>%  
  arrange(region_id, Accepted_binomial)%>%  
  group_by(region_id, Accepted_binomial) %>%  
  slice_head() # 1 entry per species x region  
  
head(naturalize_sp)
```

```
## # A tibble: 6 x 16  
## # Groups:   region_id, Accepted_binomial [6]  
##   taxon_orig_auth    taxon_orig author_name_orig family_orig status_orig country  
##   <chr>             <chr>      <chr>           <chr>      <chr>      <chr>  
## 1 Acacia auriculifo~ Acacia au~ Benth.         Leguminosae invasive   BN  
## 2 Acacia holoserice~ Acacia ho~ G.Don          Leguminosae naturalized BN  
## 3 Acacia mangium Wi~ Acacia ma~ Willd          Fabaceae    invasive   ID  
## 4 Acmeilla uliginosa Acmeilla u~ <NA>           <NA>        naturalized ID  
## 5 Acrocarpus fraxin~ Acrocarpu~ <NA>           <NA>        invasive   ID  
## 6 Aeschynomene indi~ Aeschynom~ <NA>           <NA>        naturalized ID  
## # ... with 10 more variables: region_id <chr>, reference <chr>, source <dbl>,
```

```
## # Accepted_Family <chr>, Accepted_genus <chr>, Status <chr>,
## # Accepted_Taxon <chr>, Accepted_binomial <chr>, Dates <chr>,
## # final_status <chr>
```

```
# Get species richness per island group
```

```
naturalized_rich_sp <- naturalize_sp %>%
  group_by (region_id) %>%
  dplyr::count(region_id)%>%
  arrange(-n) %>%
  rename (richness = n)

naturalized_rich_sp
```

```
## # A tibble: 10 x 2
## # Groups:   region_id [10]
##   region_id      richness
##   <chr>         <int>
## 1 Singapore         569
## 2 Philippines        539
## 3 Java              435
## 4 LSI               400
## 5 New Guinea        338
## 6 Peninsular Malaysia 231
## 7 Borneo            194
## 8 Sumatra           183
## 9 Sulawesi          129
## 10 Maluku            87
```

```
summary(naturalized_rich_sp)
```

```
##   region_id      richness
## Length:10      Min.    : 87.0
## Class :character 1st Qu.:185.8
## Mode  :character Median :284.5
##                      Mean  :310.5
##                      3rd Qu.:426.2
##                      Max.   :569.0
```

```
sd(naturalized_rich_sp$richness)
```

```
## [1] 170.7319
```

```
# add relative and perc naturalized to environmental data
```

```
Malesia_env <- Malesia_environmental %>%
  mutate(Relative_naturalized_richness = Naturalize/Native_ric) %>%
  mutate(Perc_total_flora_naturalized =
    (Naturalize/(Naturalize + Native_ric))*100)
```

```
# get a species list for the naturalized taxa
```

```
mal_sp <- naturalize_sp %>%  
  group_by(Accepted_binomial) %>%  
  slice_head() %>%  
  select(Accepted_Family, Accepted_Taxon, Accepted_binomial)
```

Frequency of naturalized species across Malesia

```
## How frequent are naturalized species across the island groups?
```

```
naturalized_freq <- naturalize_sp %>%  
  group_by (Accepted_binomial) %>%  
  dplyr::count(Accepted_binomial) %>%  
  arrange(-n) %>%  
  rename (frequency = n)
```

```
naturalized_freq
```

```
## # A tibble: 1,345 x 2  
## # Groups:   Accepted_binomial [1,345]  
##   Accepted_binomial      frequency  
##   <chr>                <int>  
## 1 Amaranthus spinosus         10  
## 2 Celosia argentea           10  
## 3 Chromolaena odorata         10  
## 4 Cleome rutidosperma         10  
## 5 Crassocephalum crepidioides 10  
## 6 Eclipta prostrata           10  
## 7 Eichhornia crassipes        10  
## 8 Elephantopus scaber         10  
## 9 Eleusine indica            10  
## 10 Euphorbia hirta             10  
## # ... with 1,335 more rows
```

```
summary(naturalized_freq)
```

```
## Accepted_binomial frequency  
## Length:1345      Min.   : 1.000  
## Class :character 1st Qu.: 1.000  
## Mode  :character Median : 1.000  
##                               Mean  : 2.309  
##                               3rd Qu.: 3.000  
##                               Max.   :10.000
```

```
sd(naturalized_freq$frequency)
```

```
## [1] 2.184386
```

```

# which species are common to every biogeographic unit?

freq_allBU <- naturalized_freq %>%
  filter(frequency == "10")

# add family name to table

most_freq_sp_fam <- freq_allBU %>%
  left_join(naturalize_sp, by = "Accepted_binomial") %>%
  group_by(Accepted_binomial) %>%
  slice_head() %>%
  select (Accepted_binomial, Accepted_Taxon, Accepted_Family)

# how many families represented by the most frequent species?

most_freq_sp_fam %$%
  n_distinct(Accepted_Family)

```

```
## [1] 14
```

```

# which families have the most widespread species?

```

```

top_fam_freq <- most_freq_sp_fam %>%
  group_by(Accepted_Family) %>%
  mutate (count = n()) %>%
  select(Accepted_Family, count) %>%
  group_by (Accepted_Family) %>%
  slice_head () %>%
  arrange(-count) %>%
  rename(No_sp = count)

top_fam_freq

```

```

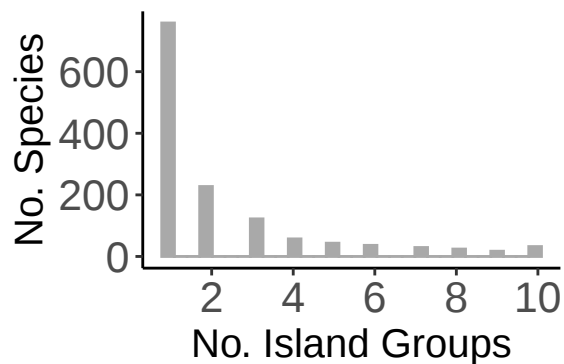
## # A tibble: 14 x 2
## # Groups:   Accepted_Family [14]
##   Accepted_Family No_sp
##   <chr>           <int>
## 1 Asteraceae       6
## 2 Poaceae          6
## 3 Fabaceae         4
## 4 Amaranthaceae    2
## 5 Convolvulaceae  2
## 6 Euphorbiaceae    2
## 7 Lamiaceae        2
## 8 Verbenaceae      2
## 9 Cleomaceae       1
## 10 Muntingiaceae   1
## 11 Piperaceae       1
## 12 Plantaginaceae  1
## 13 Pontederiaceae  1
## 14 Salviniaceae    1

```

```
# plot frequency

p_freq <- naturalized_freq %>%
  ggplot(aes(x = frequency)) +
  geom_histogram(color="darkgray", fill="darkgray") +
  # geom_vline(aes(xintercept=mean(frequency)),
  # color="black", linetype="dashed", size=1) +
  xlab("No. Island Groups") +
  ylab("No. Species") +
  theme (text = element_text(size = 20),
        # Hide panel borders and remove grid lines
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        # Change axis line
        axis.line = element_line(colour = "black"),
        axis.title = element_text(size = 15)) +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 5))

p_freq
```



```
# how many are widespread and occur in 5 island groups or more
```

```
widespread <- naturalized_freq %>%
  filter(frequency >= 5) # 5 or over

widespread %$%
  n_distinct(Accepted_binomial)
```

```
## [1] 181
```

```
rare <- naturalized_freq %>%
  filter(frequency == 1) # only 1

rare %$%
  n_distinct(rare)
```

```
## [1] 758
```

A look at the diverse families and genera

```
## How many families and genera in the dataset?
```

```
# first one will produce a table of all families
```

```
mal_fam <- naturalize_sp %>%  
  group_by(Accepted_Family) %>%  
  slice_head() %>%  
  select(Accepted_Family) %>%  
  arrange(Accepted_Family)  
  
head(mal_fam)
```

```
## # A tibble: 6 x 1  
## # Groups:   Accepted_Family [6]  
## Accepted_Family  
## <chr>  
## 1 Acanthaceae  
## 2 Acoraceae  
## 3 Agavaceae  
## 4 Aizoaceae  
## 5 Alismataceae  
## 6 Alliaceae
```

```
# all genera by families
```

```
mal_gen <- naturalize_sp %>%  
  group_by(Accepted_genus) %>%  
  slice_head() %>%  
  select(Accepted_Family, Accepted_genus) %>%  
  arrange(Accepted_Family)
```

```
# all species in each families
```

```
mal_sp <- naturalize_sp %>%  
  group_by(Accepted_binomial) %>%  
  slice_head() %>%  
  select(Accepted_Family, Accepted_genus, Accepted_Taxon, Accepted_binomial) %>%  
  arrange(Accepted_Family, Accepted_genus, Accepted_Taxon)
```

```
# count no genus in family and gen in family and make new table
```

```
mal_gen_count <- mal_gen %>%  
  group_by (Accepted_Family) %>%  
  mutate (count = n()) %>%  
  select(Accepted_Family, count) %>%  
  group_by (Accepted_Family) %>%  
  slice_head () %>%  
  arrange(-count) %>%  
  rename(No_genus = count)
```

```
# count no. species in family
```

```
mal_sp_count <- mal_sp %>%
  group_by (Accepted_Family) %>%
  mutate (count = n()) %>%
  select(Accepted_Family, count) %>%
  group_by (Accepted_Family) %>%
  slice_head () %>%
  arrange(-count) %>%
  rename(No_species = count)
```

```
## combine the two together
```

```
mal_fam_count <- mal_gen_count %>%
  left_join(mal_sp_count) %>%
  arrange(-No_genus, -No_species)
```

```
head(mal_fam_count)
```

```
## # A tibble: 6 x 3
## # Groups:   Accepted_Family [6]
## Accepted_Family No_genus No_species
## <chr>          <int>    <int>
## 1 Fabaceae      77      171
## 2 Asteraceae    70      103
## 3 Poaceae       66      160
## 4 Acanthaceae   22       46
## 5 Lamiaceae     21       40
## 6 Rubiaceae     20       34
```

```
# average
```

```
mal_fam_count %>%
  summary()
```

```
## Accepted_Family      No_genus      No_species
## Length:161          Min.   : 1.000    Min.   : 1.000
## Class :character    1st Qu.: 1.000    1st Qu.: 1.000
## Mode  :character    Median : 1.000    Median : 2.000
##                      Mean    : 4.391    Mean    : 8.354
##                      3rd Qu.: 4.000    3rd Qu.: 6.000
##                      Max.    :77.000    Max.    :171.000
```

Naturalized plant species richness

```
## Species richness plots
```

```
#import spatial data
```

```
mal <- st_read(
  "Malesia_environmental_data.shp")
```

```
## Reading layer 'Malesia_environmental_data' from data source
```

```
## 'C:\Users\racha\Documents\MalNAF_Analysis_R\Biodiversity_Homogenization\MalNAF_version1_analysis\M
## using driver 'ESRI Shapefile'
## Simple feature collection with 10 features and 35 fields
## Geometry type: MULTIPOLYGON
## Dimension: XY
## Bounding box: xmin: 95.01271 ymin: -11.63633 xmax: 155.9675 ymax: 21.12238
## Geodetic CRS: WGS 84
```

```
# need to combine the .shp file with species richness vars.
```

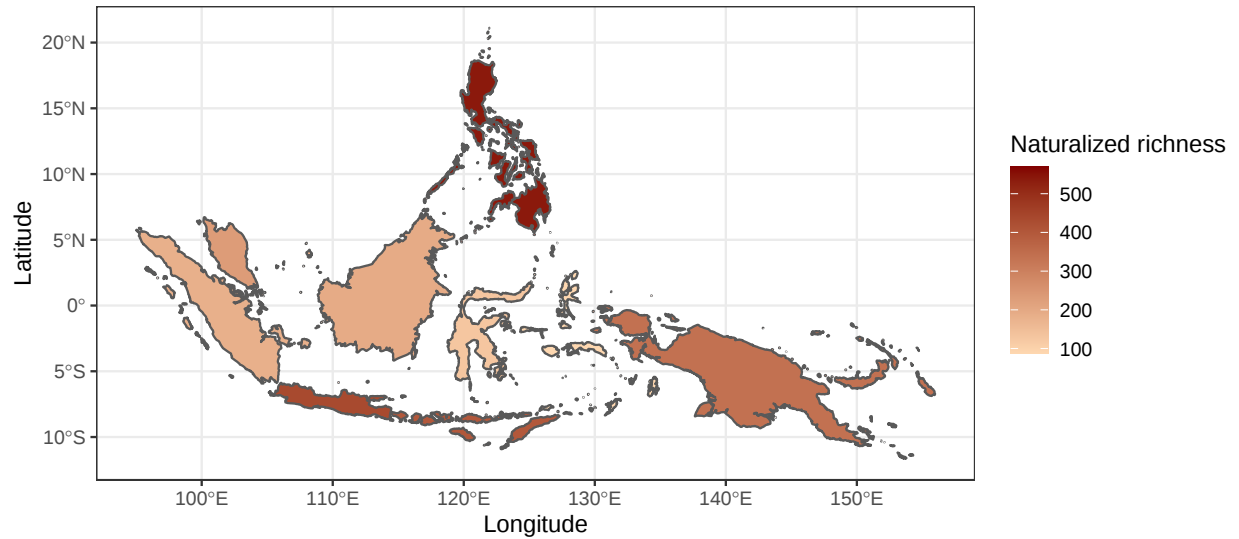
```
mal_env_filt <- Malesia_env %>%
  select("Island_nam", "Relative_naturalized_richness", "Perc_total_flora_naturalized")
```

```
mal_rich <- mal %>%
  left_join (mal_env_filt, by = ("Island_nam"))
```

```
# fig 3
```

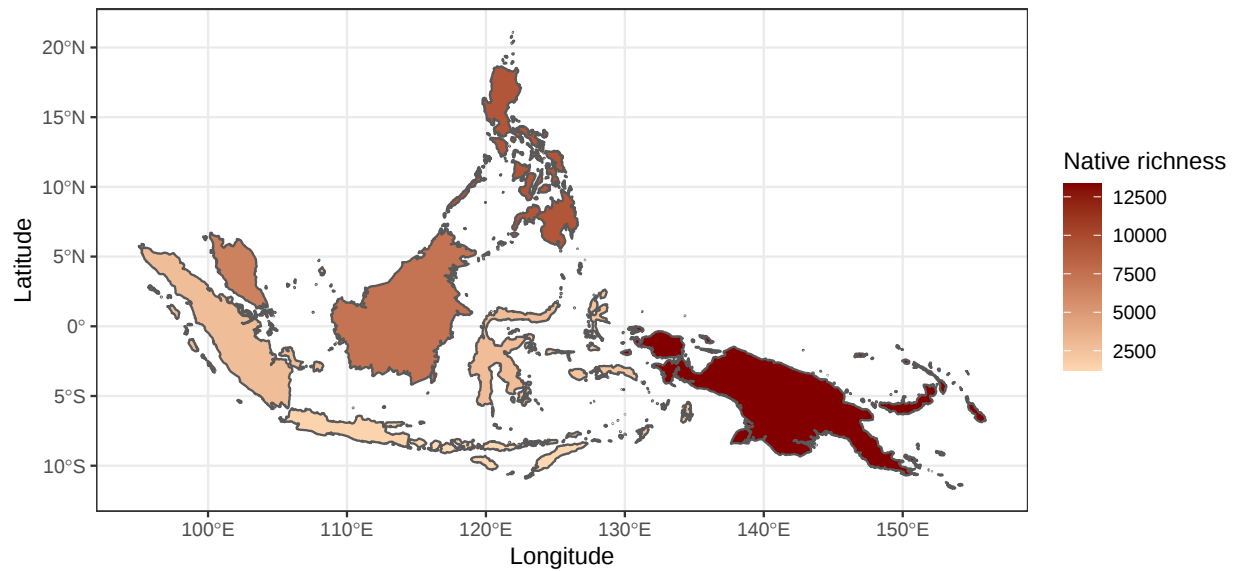
```
p1 <- mal_rich %>%
  ggplot() +
  geom_sf(aes(fill= Naturalize))+
  scale_fill_continuous(
    low = "#fed8b1", high = "#820000")+
  labs(x='Longitude',y='Latitude') +
  theme_bw()+
#theme(legend.position = 'bottom')+
  labs(fill = "Naturalized richness")
```

```
p1 # raw richness
```



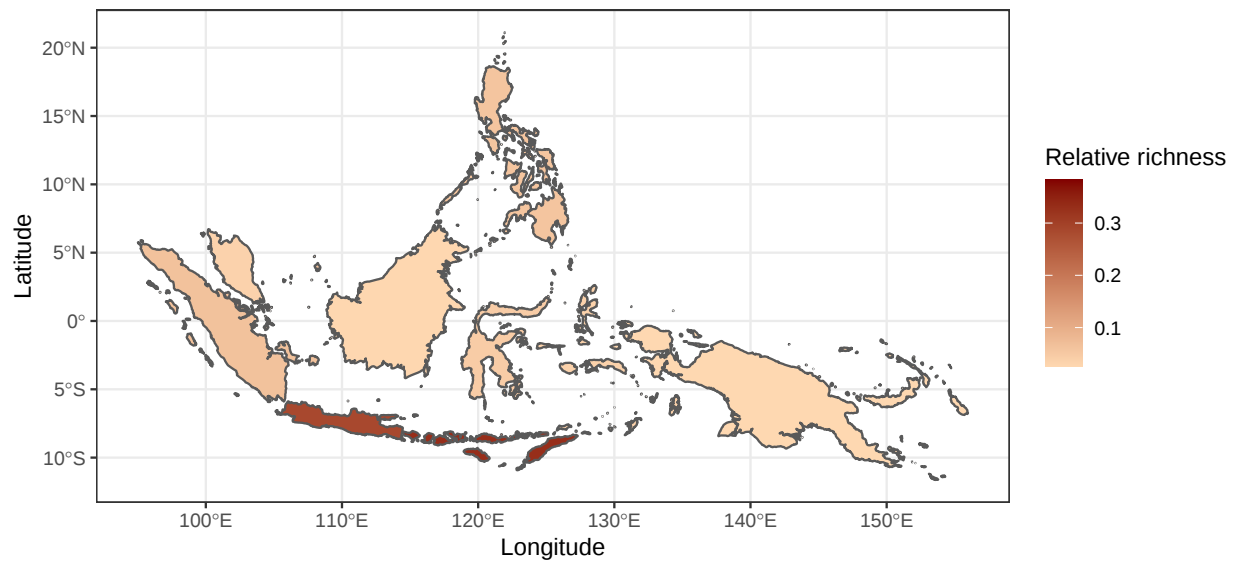
```
p2 <- mal_rich %>%
  ggplot() +
  geom_sf(aes(fill= Native_ric))+
  scale_fill_continuous(
    low = "#fed8b1", high = "#820000")+
  labs(x='Longitude',y='Latitude') +
  theme_bw()+
  #theme(legend.position = 'bottom')+
  labs(fill = "Native richness")
```

p2

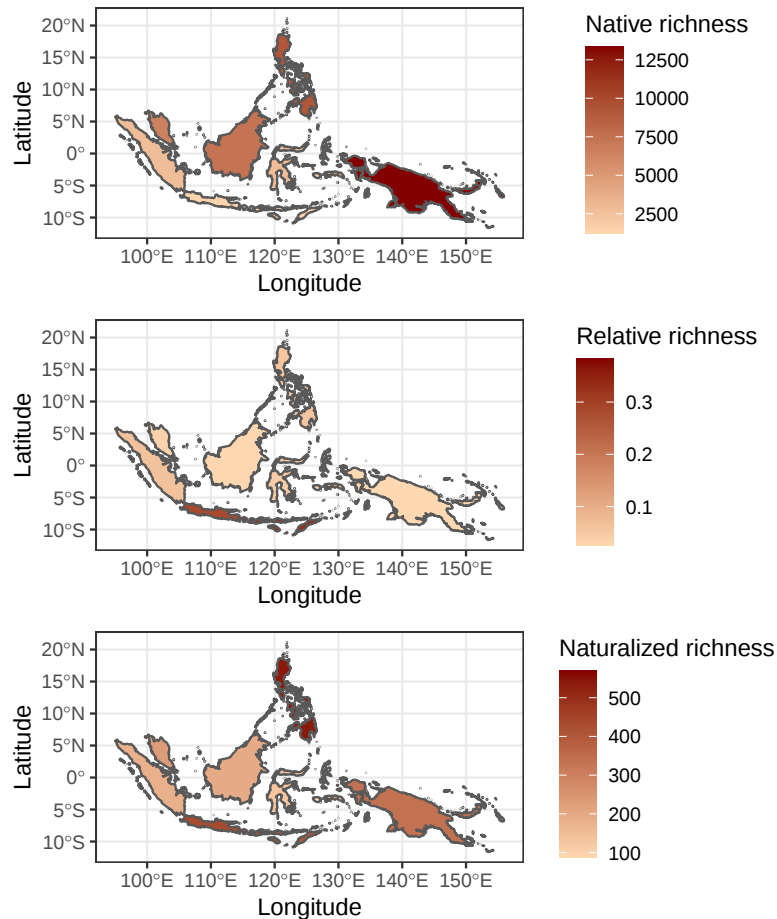


```
p3 <- mal_rich %>%
  ggplot() +
  geom_sf(aes(fill= Relative_naturalized_richness))+
  scale_fill_continuous(
    low = "#fed8b1", high = "#820000")+
  labs(x='Longitude',y='Latitude') +
  theme_bw()+
  #theme(legend.position = 'bottom')+
  labs(fill = "Relative richness") # not sure how to caption
```

p3



```
ggarrange(p2,p3,p1,heights = c(4,4,4),widths = c(6,6,6), nrow=3, align = "v")
```



Family naturalization index

```
## Family Naturalization index

lcvp_fam <- lcvp_group_search(mal_fam$Accepted_Family,
                             search_by = "Family") # search for fam in LCVF

lcvp_fam_count <- lcvp_fam %>%
  filter(Status == "accepted") %>%
  group_by(Family) %>%
  summarise(fam_species_global = n_distinct(Output.Taxon)) # get spp. counts

# count no. species in family

mal_fam_bor <- naturalize_sp %>%
  filter(region_id == "Borneo") %>%
  group_by(Accepted_Family) %>%
  summarise(
    Borneo = n_distinct(Accepted_binomial)
  )

mal_fam_pen <- naturalize_sp %>%
  filter(region_id == "Peninsular Malaysia") %>%
```

```

group_by(Accepted_Family) %>%
summarise(
  Peninsular_Malaysia = n_distinct(Accepted_binomial)
)

mal_fam_sgp <- naturalize_sp %>%
  filter(region_id == "Singapore")%>%
  group_by(Accepted_Family) %>%
  summarise(
    Singapore = n_distinct(Accepted_binomial)
  )

mal_fam_sum <- naturalize_sp %>%
  filter(region_id == "Sumatra")%>%
  group_by(Accepted_Family) %>%
  summarise(
    Sumatra = n_distinct(Accepted_binomial)
  )

mal_fam_jav <- naturalize_sp %>%
  filter(region_id == "Java")%>%
  group_by(Accepted_Family) %>%
  summarise(
    Java = n_distinct(Accepted_binomial)
  )

mal_fam_LSI <- naturalize_sp %>%
  filter(region_id == "LSI")%>%
  group_by(Accepted_Family) %>%
  summarise(
    LSI = n_distinct(Accepted_binomial)
  )

mal_fam_sul <- naturalize_sp %>%
  filter(region_id == "Sulawesi")%>%
  group_by(Accepted_Family) %>%
  summarise(
    Sulawesi = n_distinct(Accepted_binomial)
  )

mal_fam_maluku <- naturalize_sp %>%
  filter(region_id == "Maluku")%>%
  group_by(Accepted_Family) %>%
  summarise(
    Maluku = n_distinct(Accepted_binomial)
  )

mal_fam_phi <- naturalize_sp %>%
  filter(region_id == "Philippines")%>%
  group_by(Accepted_Family) %>%
  summarise(
    Philippines = n_distinct(Accepted_binomial)
  )

```

```

mal_fam_nwg <- naturalize_sp %>%
  filter(region_id == "New Guinea") %>%
  group_by(Accepted_Family) %>%
  summarise(
    NWG = n_distinct(Accepted_binomial)
  )

# now combine

fam_island <- mal_fam_bor %>%
  full_join(mal_fam_pen) %>%
  full_join(mal_fam_sgp) %>%
  full_join(mal_fam_sum) %>%
  full_join(mal_fam_jav) %>%
  full_join(mal_fam_LSI) %>%
  full_join(mal_fam_sul) %>%
  full_join(mal_fam_phi) %>%
  full_join(mal_fam_maluku) %>%
  full_join(mal_fam_nwg) %>%
  rename(Family = Accepted_Family) %>%
  replace(is.na(.), 0)

# now to calculate

# need to pivot longer then add species richness per island group

FI_long <- fam_island %>%
  pivot_longer(cols = -Family, names_to = "Island", values_to = "species")

# finally add back in the totals for world and Malesia to allow for calc
# also need total richness per island

family_index <- FI_long %>%
  mutate(Island = replace(Island, Island == "Peninsular_Malaysia",
    "Peninsular Malaysia")) %>%
  mutate(Island = replace(Island, Island == "NWG", "New Guinea")) %>%
  left_join(lcvp_fam_count) %>%
  left_join(naturalized_rich_sp, by = c("Island" = "region_id")) %>%
  mutate(FI = (species/(fam_species_global * richness))*100) %>% # calculate FI
  mutate(Island = replace(Island, Island == "LSI", "Lesser Sunda Islands")) %>%
  mutate(F_ratio = (species/fam_species_global)*100)

# only include families which have at least 10 species naturalized in an
# individual island as these are important in the flora and cannot plot all
# families

family_10 <- family_index %>%
  filter(species > 9) %>% # above 10 in a region
  group_by(Family) %>%
  slice_head() %>%
  select(Family)

```

```
family_index_10 <- family_10 %>%  
  left_join(family_index)
```

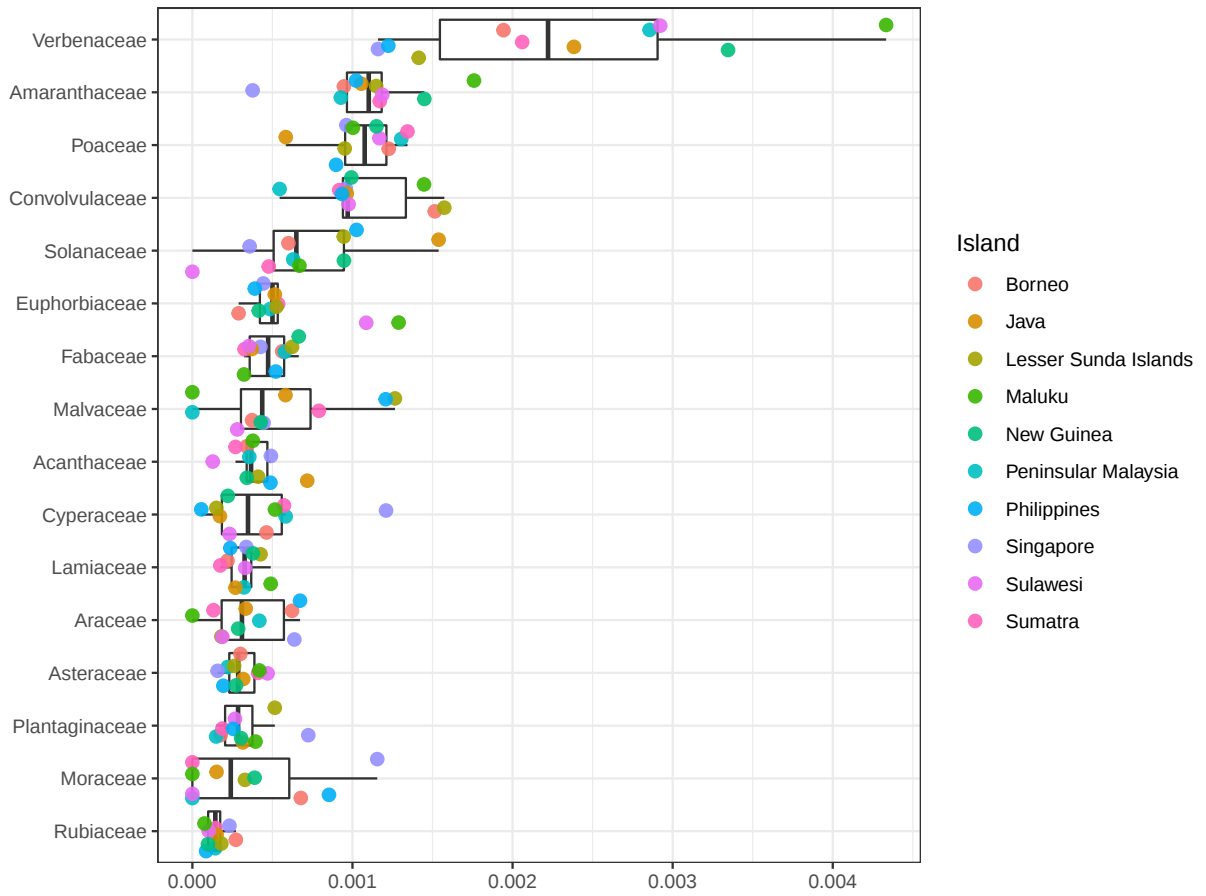
```
family_index_10 %$%  
  n_distinct(Family)
```

```
## [1] 16
```

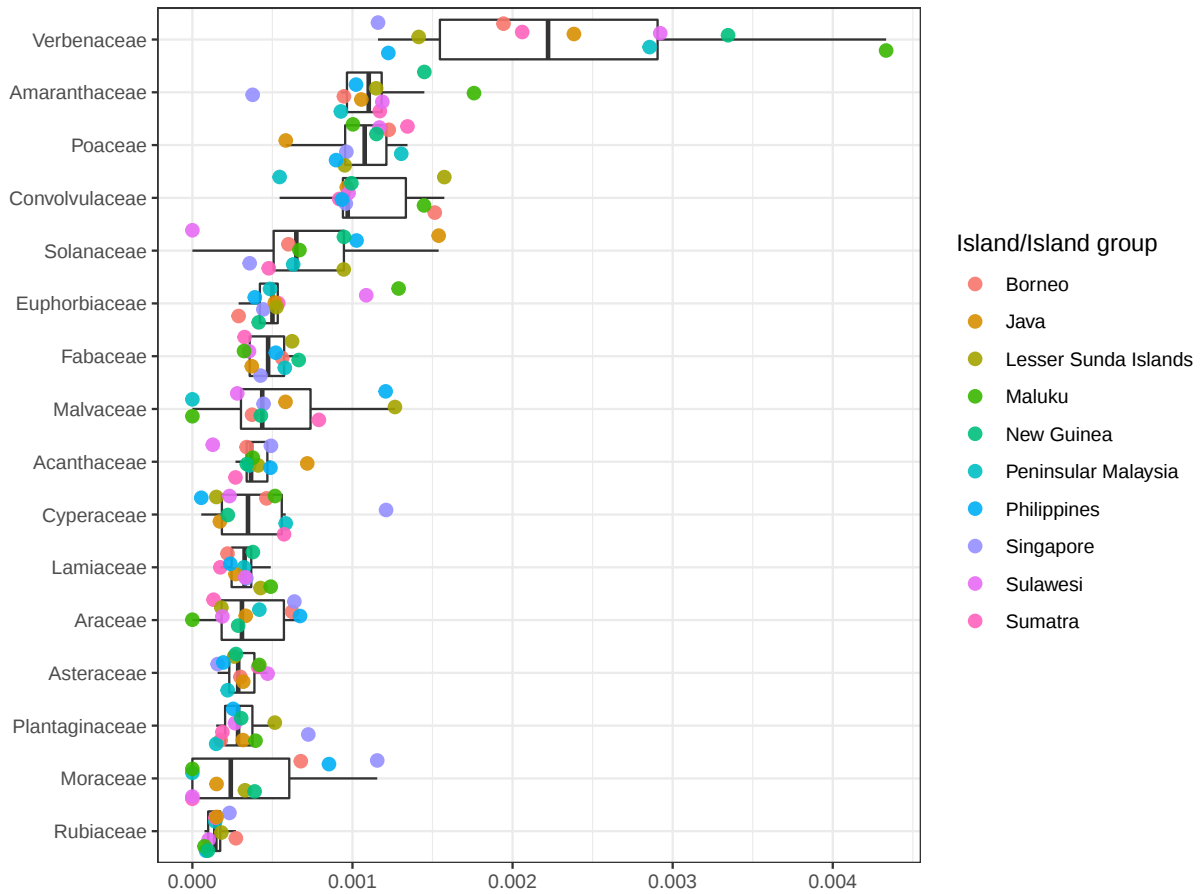
```
FI_1 <- family_index_10 %>%  
  ggplot(aes(y= FI, x= fct_reorder(Family, FI))) +  
  geom_boxplot(outlier.shape = NA) +  
  scale_fill_viridis(discrete = TRUE, alpha=0.6) +  
  geom_jitter(aes(color= Island),  
              size=2.5, alpha=0.9) +  
  theme(  
    legend.title = element_text(size=14, color = "dimgrey"),  
    legend.text = element_text(size=14, color = "dimgrey"),  
    legend.position="right"  
  ) +  
  xlab("") +  
  ylab("") +  
  theme_bw() +  
  coord_flip()
```

```
FI_1 <- FI_1 + labs(fill = "Island/Island group")
```

```
FI_1
```

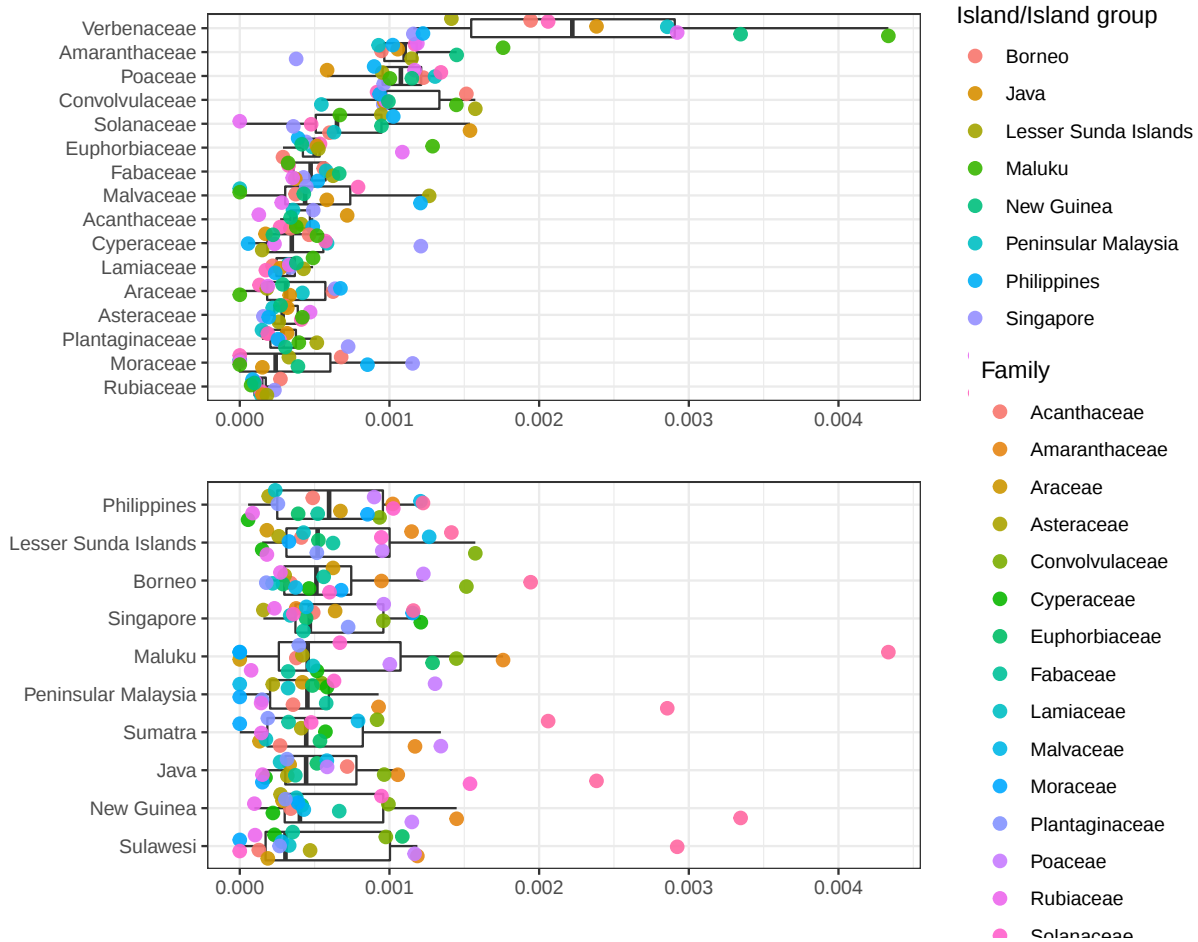


```
FI_1 <- print(FI_1 + labs(colour = "Island/Island group"))
```



```
FI_2 <- family_index_10 %>%
  ggplot(aes(y= FI, x= fct_reorder(Island, FI))) +
  geom_boxplot(outlier.shape = NA) +
  scale_fill_viridis(discrete = TRUE, alpha=0.6) +
  geom_jitter(aes(color= Family),
              size=2.5, alpha=0.9) +
  theme(
    legend.title = element_text(size=14, color = "dimgrey"),
    legend.text = element_text(size=14, color = "dimgrey"),
  ) +
  xlab("")+
  ylab("") +
  theme_bw() +
  coord_flip()

ggarrange(FI_1, FI_2, nrow = 2, align = "hv", legend = "right") # fig.4
```



Native ranges of naturalized species

```
range <- Native_range %>%
  select (binomial, Europe, Africa, Asia.Temperate,
          Asia.Tropical, Australasia, Pacific, Northern.America,
          Southern.America, Antarctica, Hybrid, Unknown)

head(range)
```

```
## # A tibble: 6 x 12
##   binomial      Europe Africa Asia.Temperate Asia.Tropical Australasia Pacific
##   <chr>         <dbl>  <dbl>         <dbl>         <dbl>         <dbl>  <dbl>
## 1 Abelmoschus ma~      0      0             1             1             0      0
## 2 Abelmoschus mo~      0      0             1             1             1      0
## 3 Abrus precator~      0      1             1             1             1      1
## 4 Abutilon hirtum      0      1             1             1             0      0
## 5 Abutilon indic~      0      1             1             1             1      1
## 6 Acacia amplice~      0      0             0             0             1      0
## # ... with 5 more variables: Northern.America <dbl>, Southern.America <dbl>,
## #   Antarctica <dbl>, Hybrid <dbl>, Unknown <dbl>
```

need it in long format for analysis using dplyr

```
range_long <- range %>%
  pivot_longer(cols = -binomial, names_to = "Continent", values_to = "Present") %>%
  filter(Present >= 1) %>%
  select(binomial, Continent)

head(range_long)
```

```
## # A tibble: 6 x 2
##   binomial      Continent
##   <chr>         <chr>
## 1 Abelmoschus manihot Asia.Temperate
## 2 Abelmoschus manihot Asia.Tropical
## 3 Abelmoschus moschatus Asia.Temperate
## 4 Abelmoschus moschatus Asia.Tropical
## 5 Abelmoschus moschatus Australasia
## 6 Abrus precatorius Africa
```

which continent has donated the most naturalized plant species?

```
range_rich <- range_long %>%
  group_by (Continent) %>%
  dplyr::count(Continent)%>%
  arrange(-n) %>%
  rename (richness = n)
# knitr::kable()

range_rich # all good!
```

```
## # A tibble: 11 x 2
## # Groups:   Continent [11]
##   Continent      richness
##   <chr>         <int>
## 1 Asia.Tropical      739
## 2 Southern.America  611
## 3 Asia.Temperate    597
## 4 Northern.America  497
## 5 Africa            446
## 6 Australasia       281
## 7 Pacific           171
## 8 Europe            140
## 9 Antarctica         5
## 10 Unknown           4
## 11 Hybrid            3
```

plot it

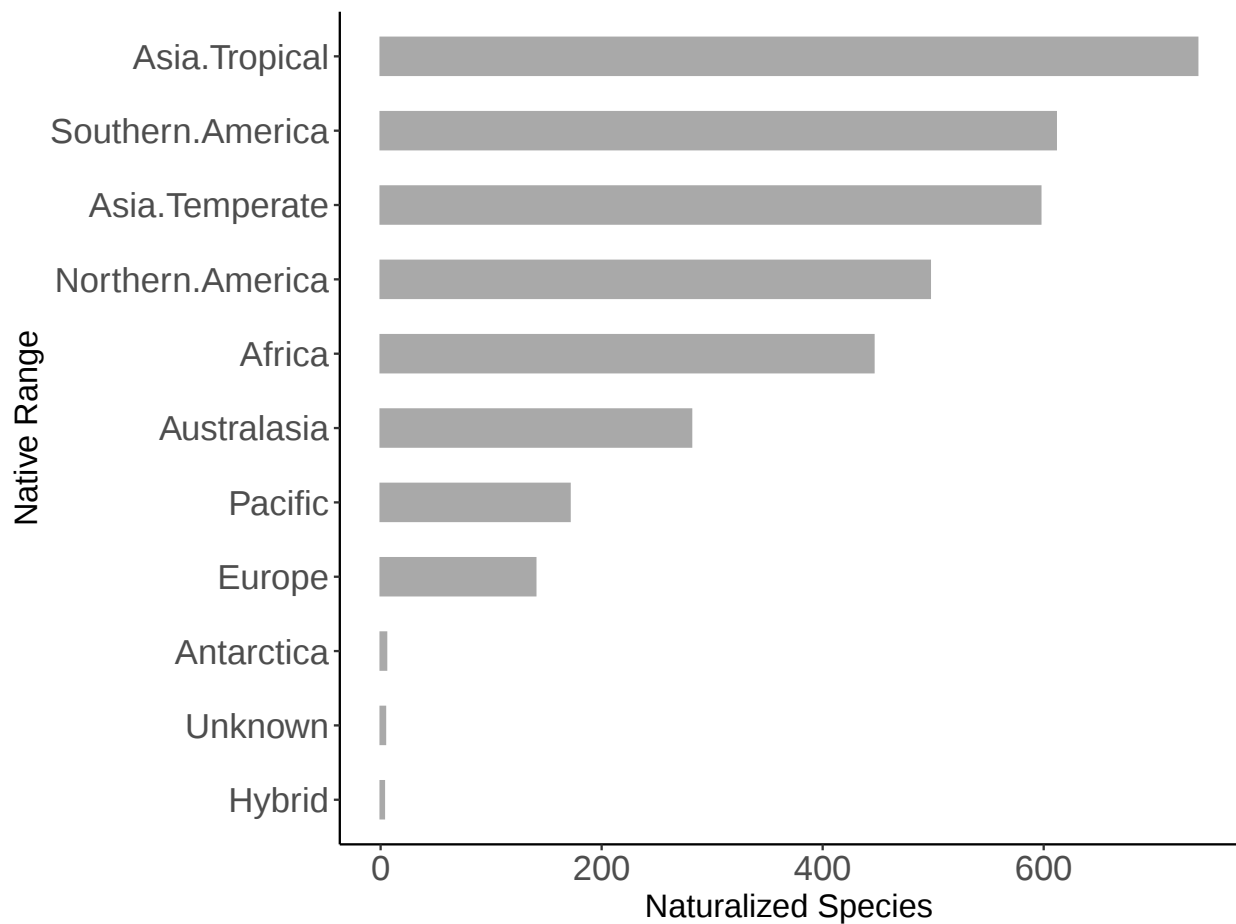
```
p_range <- range_long %>%
  group_by (Continent) %>%
  dplyr::count(Continent)%>%
  ggplot(aes(
    x = n,
    y = reorder(Continent, n))) +
```

```

geom_bar(stat = "identity",
         colour = "darkGray", fill = "darkGray",
         width = 0.5) +
theme (text = element_text(size = 20),
      # Hide panel borders and remove grid lines
      panel.border = element_blank(),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      panel.background = element_blank(),
      # Change axis line
      axis.line = element_line(colour = "black"),
      axis.title = element_text(size = 15)) +
xlab("Naturalized Species") +
ylab("Native Range")

```

p_range



```

range_rich %>%
summary()

```

```
##    Continent      richness
```

```
## Length:11      Min.   : 3.0
## Class :character 1st Qu.: 72.5
## Mode  :character Median :281.0
##                Mean   :317.6
##                3rd Qu.:547.0
##                Max.   :739.0
```

```
sd(range_rich$richness)
```

```
## [1] 271.9939
```

```
range_rich <- range_rich %>%
  mutate (perc_origin = round(((richness/3494) * 100), 2))

## need a new table combining the taxon x region data with taxon x
# native origin data

mal_range <- naturalize_sp %>%
  left_join(range_long, by = c("Accepted_binomial" = "binomial")) %>%
  select(Accepted_binomial, region_id, Continent)

mal_range
```

```
## # A tibble: 8,155 x 3
## # Groups:   region_id, Accepted_binomial [3,105]
## Accepted_binomial region_id Continent
## <chr> <chr> <chr>
## 1 Acacia auriculiformis Borneo Asia.Tropical
## 2 Acacia auriculiformis Borneo Australasia
## 3 Acacia holosericea Borneo Australasia
## 4 Acacia mangium Borneo Asia.Tropical
## 5 Acacia mangium Borneo Australasia
## 6 Acmella uliginosa Borneo Africa
## 7 Acmella uliginosa Borneo Asia.Tropical
## 8 Acmella uliginosa Borneo Pacific
## 9 Acmella uliginosa Borneo Southern.America
## 10 Acrocarpus fraxinifolius Borneo Asia.Temperate
## # ... with 8,145 more rows
```

```
mal_r_1 <- mal_range %>%
  group_by(region_id) %>%
  count(Continent) %>%
  rename (no_spp = n)

mal_r_2 <- mal_r_1 %>%
  group_by(region_id) %>%
  mutate (total_orig = sum(no_spp))

mal_r_2
```

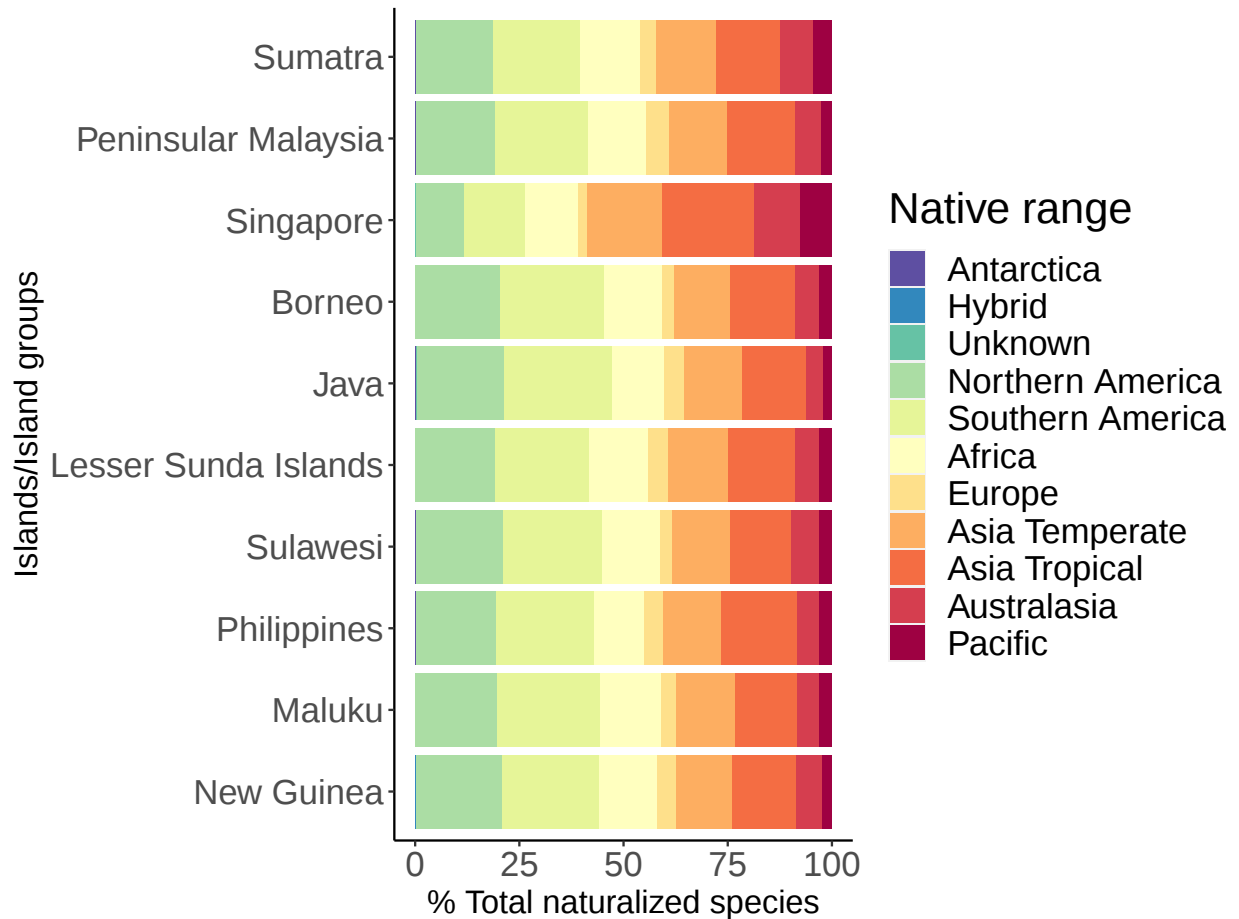
```
## # A tibble: 92 x 4
## # Groups:   region_id [10]
```

```
##   region_id Continent      no_spp total_orig
##   <chr>      <chr>        <int>    <int>
## 1 Borneo     Africa        68      482
## 2 Borneo     Asia.Temperate  65      482
## 3 Borneo     Asia.Tropical   76      482
## 4 Borneo     Australasia    28      482
## 5 Borneo     Europe         13      482
## 6 Borneo     Northern.America 98      482
## 7 Borneo     Pacific         14      482
## 8 Borneo     Southern.America 120     482
## 9 Java       Africa        130     1035
## 10 Java      Antarctica      2      1035
## # ... with 82 more rows
```

```
cont_origin <- mal_r_2 %>%
  mutate(perc_origin = (no_spp / total_orig) * 100)

p_orig_stack <- cont_origin %>%
  mutate(Continent = replace(Continent, Continent == "Asia.Tropical",
                             "Asia Tropical")) %>%
  mutate(Continent = replace(Continent, Continent == "Asia.Temperate",
                             "Asia Temperate")) %>%
  mutate(Continent = replace(Continent, Continent == "Southern.America",
                             "Southern America")) %>%
  mutate(Continent = replace(Continent, Continent == "Northern.America",
                             "Northern America")) %>%
  mutate(region_id = replace(region_id, region_id == "LSI",
                             "Lesser Sunda Islands")) %>%
  mutate(Continent = fct_relevel(Continent, "Unknown", "Hybrid", "Antarctica",
                                "Pacific", "Australasia", "Asia Tropical",
                                "Asia Temperate", "Europe", "Africa",
                                "Southern America", "Northern America")) %>%
  ggplot(aes(fill = Continent, y = perc_origin,
             x = fct_relevel(region_id, "New Guinea", "Maluku", "Philippines",
                             "Sulawesi", "Lesser Sunda Islands", "Java", "Borneo",
                             "Singapore",
                             "Peninsular Malaysia", "Sumatra")))) +
  geom_bar(position="stack", stat="identity") +
  theme(text = element_text(size = 20),
        # Hide panel borders and remove grid lines
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        # Change axis line
        axis.line = element_line(colour = "black"),
        axis.title = element_text(size = 15)) +
  scale_fill_brewer(palette = "Spectral") +
  ylab("% Total naturalized species") +
  xlab("Islands/Island groups") +
  guides(fill=guide_legend(title="Native range", reverse = TRUE)) +
  coord_flip()

p_orig_stack # figure 5
```



Habitat preferences of species in Malesia in their naturalized ranges.

```
# how many species in data?
```

```
hab %$%
  n_distinct(binomial)
```

```
## [1] 581
```

```
# how many species in each country?
```

```
sp_no <- hab %>%
  group_by(Region_ID) %>%
  summarise(no_species = n_distinct(binomial))
```

```
sp_no
```

```
## # A tibble: 4 x 2
##   Region_ID no_species
##   <chr>      <int>
## 1 ID        132
## 2 PG         64
## 3 PH        474
## 4 TL         84
```

```
# how many records by country total
```

```
sp_records <- hab %>%
  group_by(Region_ID) %>%
  summarise(hab_records = (count = n())) %>%
  mutate(perc_hab_records = (hab_records/sum(hab_records))*100) %>%
  arrange(-perc_hab_records)

sp_records
```

```
## # A tibble: 4 x 3
##   Region_ID hab_records perc_hab_records
##   <chr>      <int>      <dbl>
## 1 PH        1181        72.4
## 2 ID         214        13.1
## 3 TL         122         7.48
## 4 PG         115         7.05
```

```
# What proportion of naturalized species occur in each habitat type across Malesia?
```

```
all_hab <- hab %>%
  group_by(binomial, Habitat) %>%
  slice_head()%>% # for each species no double habitat entries
  ungroup() %>%
  group_by(Habitat)%>%
  summarise(hab_records = (count = n())) %>%
  mutate(perc_hab_records = (hab_records/sum(hab_records))*100) %>%
  arrange(-perc_hab_records)

all_hab
```

```
## # A tibble: 15 x 3
##   Habitat hab_records perc_hab_records
##   <chr>      <int>      <dbl>
## 1 12a        329        22.7
## 2 12b        291        20.1
## 3 2          207        14.3
## 4 1          113         7.79
## 5 9          103         7.10
## 6 3           96         6.62
## 7 4           93         6.41
## 8 4b          55         3.79
## 9 8           50         3.45
## 10 10         32         2.21
## 11 5           32         2.21
## 12 6           19         1.31
## 13 4a          16         1.10
## 14 11          13         0.896
## 15 7           2         0.138
```

```
# What proportion of records reported from each habitat type across Malesia?
```

```
all_hab2 <- hab %>%
  group_by(binomial, Habitat) %>%
  ungroup() %>%
  group_by(Habitat)%>%
  summarise(hab_records = (count = n())) %>%
  mutate(perc_hab_records = (hab_records/sum(hab_records))*100) %>%
  arrange(-perc_hab_records)
```

```
all_hab2
```

```
## # A tibble: 15 x 3
##   Habitat hab_records perc_hab_records
##   <chr>         <int>         <dbl>
## 1 12a             408             25
## 2 12b             368            22.5
## 3 2              210            12.9
## 4 1              122             7.48
## 5 9              104             6.37
## 6 3              96             5.88
## 7 4              95             5.82
## 8 4b             62             3.80
## 9 8              50             3.06
## 10 10            32             1.96
## 11 5             32             1.96
## 12 6             19             1.16
## 13 11            16             0.980
## 14 4a            16             0.980
## 15 7              2             0.123
```

```
# plot
```

```
Malesia_hab <- hab %>%
  group_by(binomial, Habitat) %>%
  slice_head()%>% # for each species no double habitat entries
  ungroup() %>%
  group_by(Habitat)%>%
  summarise(hab_records = (count = n())) %>%
  mutate(perc_hab_records = (hab_records/sum(hab_records))*100) %>%
  arrange(-perc_hab_records) %>%
  mutate(Habitat = replace(Habitat, Habitat == "1", "Forest")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "2", "Open forest")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "3", "Scrub")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "4", "Grassland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "4a", "Natural grassland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "4b", "Anthropogenic grassland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "5", "Sandy")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "6", "Rocky")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "7", "Dryland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "8", "Saline")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "9", "Riparian")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "10", "Wetland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "11", "Aquatic")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "12", "Anthropogenic")) %>%
```

```

mutate(Habitat = replace(Habitat, Habitat == "12a", "Ruderal/Urban")) %>%
mutate(Habitat = replace(Habitat, Habitat == "12b", "Agricultural")) %>%
mutate(Region = "Malesia")

# set up colour

## You need to expand palette size

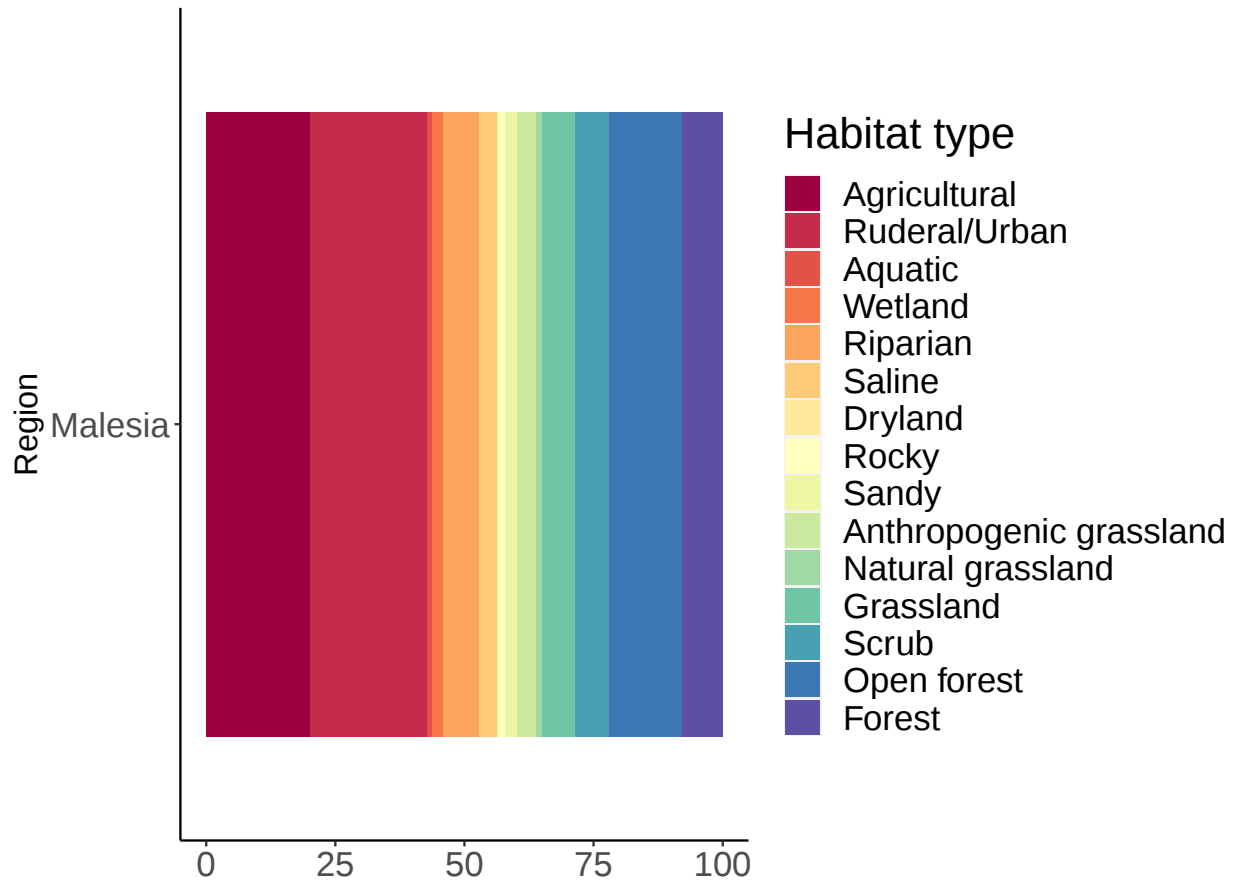
cols <- 15
mycolors <- colorRampPalette(brewer.pal(11, "Spectral"))(cols)
mycolours <- rev(mycolors)

# plot

mal_hab_stack <- Malesia_hab %>%
  ggplot(aes(fill = fct_relevel(Habitat, "Forest", "Open forest", "Scrub", "Grassland",
                                "Natural grassland", "Anthropogenic grassland",
                                "Sandy", "Rocky", "Dryland", "Saline",
                                "Riparian", "Wetland", "Aquatic",
                                "Ruderal/Urban", "Agricultural"),
          y = perc_hab_records, x = Region)) +
  geom_bar(position="stack", stat="identity") +
  theme (text = element_text(size = 20),
        # Hide panel borders and remove grid lines
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        # Change axis line
        axis.line = element_line(colour = "black"),
        axis.title = element_text(size = 15)) +
  scale_fill_manual(values = mycolours) +
  guides(fill=guide_legend(title="Habitat type", reverse = TRUE)) +
  ylab("") +
  xlab("Region") +
  coord_flip()

mal_hab_stack # use as fig. 6a

```



What is the proportion by country?

```
country_hab <- hab %>%
  group_by(Region_ID, Habitat)%>%
  summarise(hab_records = (count = n())) %>%
  mutate(perc_hab_records = (hab_records/sum(hab_records))*100) %>%
  arrange(-perc_hab_records) %>%
  mutate(Region_ID = replace(Region_ID, Region_ID == "PH", "Philippines")) %>%
  mutate(Region_ID = replace(Region_ID, Region_ID == "ID", "Indonesia")) %>%
  mutate(Region_ID = replace(Region_ID, Region_ID == "TL", "Timor Leste")) %>%
  mutate(Region_ID = replace(Region_ID, Region_ID == "PG", "Papua New Guinea")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "1", "Forest")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "2", "Open forest")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "3", "Scrub")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "4", "Grassland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "4a", "Natural grassland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "4b", "Anthropogenic grassland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "5", "Sandy")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "6", "Rocky")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "7", "Dryland")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "8", "Saline")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "9", "Riparian")) %>%
  mutate(Habitat = replace(Habitat, Habitat == "10", "Wetland")) %>%
```

```

mutate(Habitat = replace(Habitat, Habitat == "11", "Aquatic")) %>%
mutate(Habitat = replace(Habitat, Habitat == "12", "Anthropogenic")) %>%
mutate(Habitat = replace(Habitat, Habitat == "12a", "Ruderal/Urban")) %>%
mutate(Habitat = replace(Habitat, Habitat == "12b", "Agricultural"))

# set up colour

## You need to expand palette size

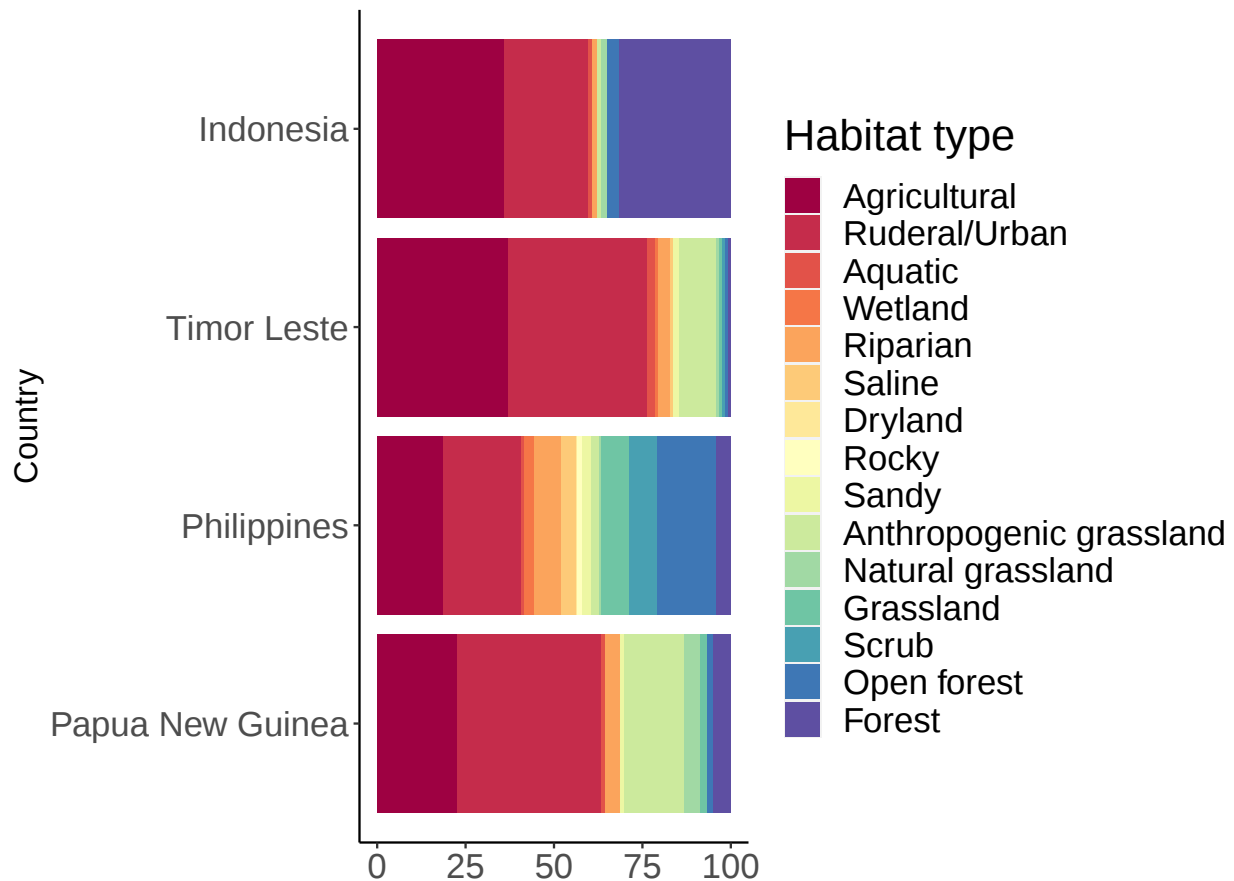
cols <- 15
mycolors <- colorRampPalette(brewer.pal(11, "Spectral"))(cols)
mycolours <- rev(mycolors)

# plot

p_hab_stack <- country_hab %>%
  ggplot(aes(fill = fct_relevel(Habitat, "Forest", "Open forest", "Scrub", "Grassland",
                                "Natural grassland", "Anthropogenic grassland",
                                "Sandy", "Rocky", "Dryland", "Saline",
                                "Riparian", "Wetland", "Aquatic",
                                "Ruderal/Urban", "Agricultural"),
          y = perc_hab_records,
          x = fct_relevel(Region_ID, "Papua New Guinea", "Philippines",
                          "Timor Leste", "Indonesia"))) +
  geom_bar(position="stack", stat="identity") +
  theme (text = element_text(size = 20),
        # Hide panel borders and remove grid lines
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        # Change axis line
        axis.line = element_line(colour = "black"),
        axis.title = element_text(size = 15)) +
  scale_fill_manual(values = mycolours) +
  guides(fill=guide_legend(title="Habitat type", reverse = TRUE)) +
  ylab("") +
  xlab("Country") +
  coord_flip()

p_hab_stack # use as fig. 6b

```



```
# what proportion do widespread species occur in vs. non-widespread

widespread_sp <- widespread %>%
  filter (frequency > 4) %>%
  select (Accepted_binomial)

habitat_filt <- hab %>%
  group_by(binomial, Habitat) %>%
  slice_head()%>% # for each species no double habitat entries
  ungroup()

hab_wide <- habitat_filt %>%
  semi_join(widespread_sp, by = c("binomial" = "Accepted_binomial")) %>%
  mutate(Distribution = "Widespread")

hab_rare <- habitat_filt %>%
  anti_join(widespread_sp, by = c("binomial" = "Accepted_binomial")) %>%
  mutate(Distribution = "Rare")

hab_frequency <- hab_wide %>%
  full_join(hab_rare)

freq_hab <- hab_frequency %>%
```

```

ungroup() %>%
group_by(Distribution, Habitat)%>%
summarise(hab_records = (count = n())) %>%
mutate(perc_hab_records = (hab_records/sum(hab_records))*100) %>%
arrange(-perc_hab_records) %>%
mutate(Habitat = replace(Habitat, Habitat == "1", "Forest")) %>%
mutate(Habitat = replace(Habitat, Habitat == "2", "Open forest")) %>%
mutate(Habitat = replace(Habitat, Habitat == "3", "Scrub")) %>%
mutate(Habitat = replace(Habitat, Habitat == "4", "Grassland")) %>%
mutate(Habitat = replace(Habitat, Habitat == "4a", "Natural grassland")) %>%
mutate(Habitat = replace(Habitat, Habitat == "4b", "Anthropogenic grassland")) %>%
mutate(Habitat = replace(Habitat, Habitat == "5", "Sandy")) %>%
mutate(Habitat = replace(Habitat, Habitat == "6", "Rocky")) %>%
mutate(Habitat = replace(Habitat, Habitat == "7", "Dryland")) %>%
mutate(Habitat = replace(Habitat, Habitat == "8", "Saline")) %>%
mutate(Habitat = replace(Habitat, Habitat == "9", "Riparian")) %>%
mutate(Habitat = replace(Habitat, Habitat == "10", "Wetland")) %>%
mutate(Habitat = replace(Habitat, Habitat == "11", "Aquatic")) %>%
mutate(Habitat = replace(Habitat, Habitat == "12a", "Ruderal/Urban")) %>%
mutate(Habitat = replace(Habitat, Habitat == "12b", "Agricultural"))

## You need to expand palette size

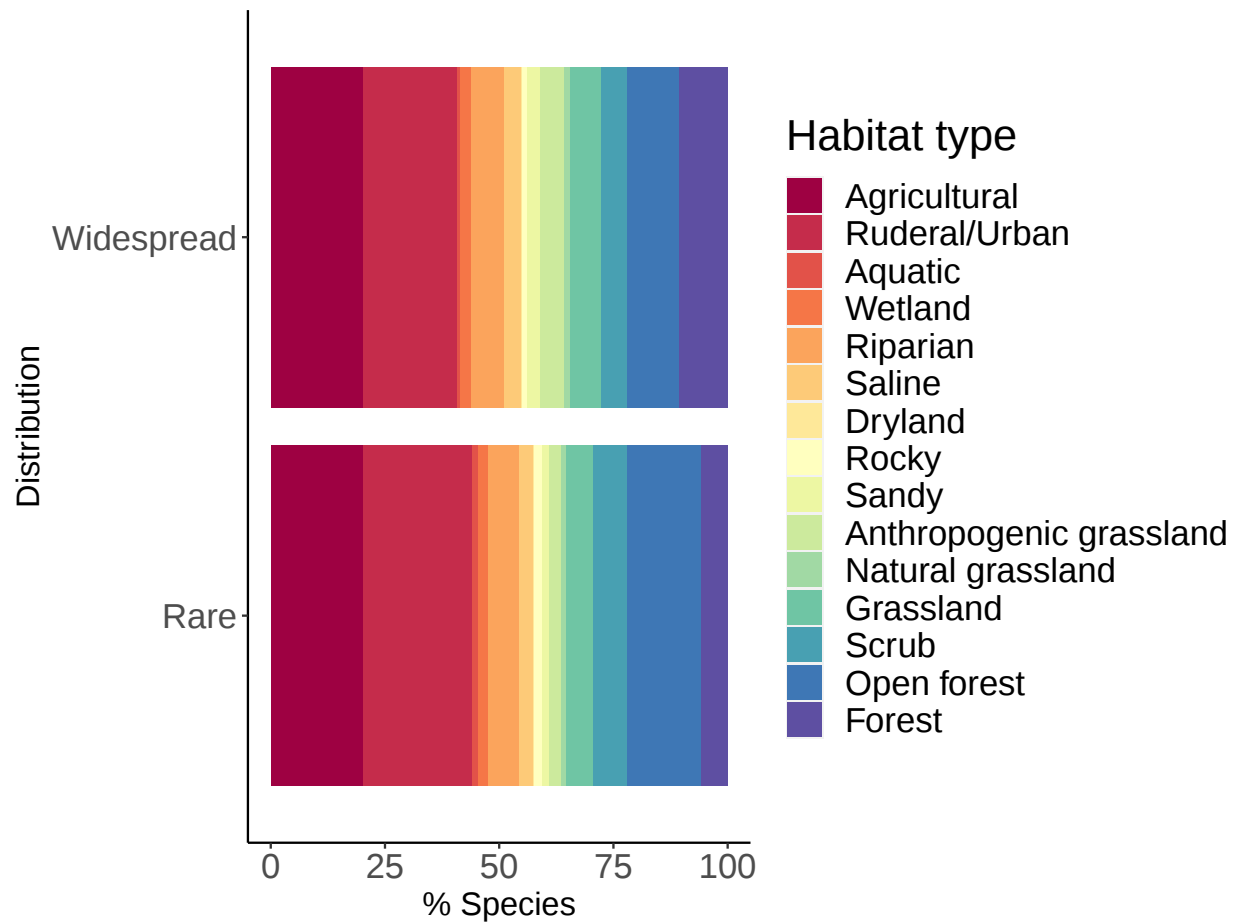
cols <- 15
mycolors <- colorRampPalette(brewer.pal(11, "Spectral"))(cols)
mycolours <- rev(mycolors)

# plot

p_hab_stack_freq <- freq_hab %>%
ggplot(aes(fill = fct_relevel(Habitat, "Forest", "Open forest", "Scrub", "Grassland",
                             "Natural grassland", "Anthropogenic grassland",
                             "Sandy", "Rocky", "Dryland", "Saline",
                             "Riparian", "Wetland", "Aquatic",
                             "Ruderal/Urban", "Agricultural"),
        y= perc_hab_records,
        x= Distribution)) +
geom_bar(position="stack", stat="identity") +
theme (text = element_text(size = 20),
      # Hide panel borders and remove grid lines
      panel.border = element_blank(),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      panel.background = element_blank(),
      # Change axis line
      axis.line = element_line(colour = "black"),
      axis.title = element_text(size = 15)) +
scale_fill_manual(values = mycolours) +
guides(fill=guide_legend(title="Habitat type", reverse = TRUE)) +
ylab("% Species") +
xlab("Distribution") +
coord_flip()

```

```
p_hab_stack_freq # use as fig 6c.
```



```
ggarrange(mal_hab_stack, p_hab_stack , p_hab_stack_freq, nrow = 3, align = "hv",
  common.legend = TRUE, legend = "right")
```

