

Problems analysis

Daniela Linero

25/5/2020

Analysis Walkthrough

Load filtered table

```
diversity <- readRDS("./output/cleaned_data/01_Filter_data_PREDICTS_Frugivores_and_Endoplants.rds")
diversity$Sample_start_earliest <- as.Date(diversity$Sample_start_earliest)
diversity$Sample_end_latest <- as.Date(diversity$Sample_end_latest)
```

Table 1: Example of measurements

	Measurement	Sampling_effort	Diversity_metric_is_effort_sensitive	SS
7759	2	48	TRUE	CM2_2007__Sambuichi 1
20590	1	34	TRUE	CM2_2007__Sambuichi 1
67668	5	52	TRUE	CM2_2007__Sambuichi 1

Now correct the sampling efforts

```
diversity <- yarg::CorrectSamplingEffort(diversity)

## Correcting 0 missing sampling effort values
## Rescaling sampling effort
## Correcting 84929 values for sensitivity to sampling effort
```

Table 2: Effort corrected measurements

	Measurement	Sampling_effort	Diversity_metric_is_effort_sensitive	SS
7759	2	1	TRUE	CM2_2007__Sambuichi 1
20590	1	1	TRUE	CM2_2007__Sambuichi 1
67668	5	1	TRUE	CM2_2007__Sambuichi 1

Next we'll merge any sites that are within the same land-use type and that have identical coordinates, start and end dates.

```
diversity <- yarg::MergeSites(diversity,
                              silent = TRUE,
                              merge.extra = "Wilderness_area")
```

```
## Error in yarg::MergeSites(diversity, silent = TRUE, merge.extra = "Wilderness_area"): length(overwri
```

The complete error is: Error in yarg::MergeSites(diversity, silent = TRUE, merge.extra = "Wilderness_area") : length(overwrite) == length(measurements) is not TRUE

Rename the `Predominant_habitat` column, since it's not really "habitat" that we're looking at. This will also keep things a little more consistent with the public version of the PREDICTS database.

```
diversity <- rename(diversity,  
                    Predominant_land_use = Predominant_habitat)
```

Now you can calculate site-level diversity metrics:

```
sites <- diversity %>%  
  
  # add Diversity_metric_is_valid column  
  mutate(Diversity_metric_is_valid = TRUE) %>%  
  
  # The extra.cols parameter is used for columns that we want to  
  # transferred to the final site-level data frame and that the function  
  # does not add automatically  
  yarg::SiteMetrics(extra.cols = c("SSB", "SSBS", "Predominant_land_use")) %>%  
  
  # calculate the maximum abundance within each study  
  group_by(SS) %>%  
  mutate(MaxAbundance = ifelse(Diversity_metric_type == "Abundance",  
                               max(Total_abundance),  
                               NA)) %>%  
  
  ungroup() %>%  
  
  # now calculate the rescaled abundance (abundance divided by the maximum within each study)  
  mutate(RescaledAbundance = ifelse(Diversity_metric_type == "Abundance",  
                                     Total_abundance/MaxAbundance,  
                                     NA))
```

```
## Computing site metrics for 99617 measurements  
## The data contain 99 sources, 125 studies and 4991 sites  
## TODO fix within-site uniqueness check  
## Computing site-level values  
##Computing total abundance  
##Computing species richness  
##Computing Simpson's diversity  
##Computing Chao  
##Computing Rarefied Species Richness  
  
## Warning in min(x[x > 0]): ningún argumento finito para min; retornando Inf  
  
## Assembling site-level values
```

Table 3: Example of the resulting table

SS	SSS	SSBS	Total_abundance	Species_richness
BS1_2010__Page 1	BS1_2010__Page 1 1	BS1_2010__Page 1 1	10	10
BS1_2010__Page 1	BS1_2010__Page 1 2	BS1_2010__Page 1 2	9	8

But I'm not getting the same results if I do the operations only for one site

1) Example of different species richness results

```
TotaAbundance <- diversity %>%
  filter(SSS == "BS1_2010__Page 1 1") %>%
  select(Best_guess_binomial, Diversity_metric,
    Measurement, SSS) %>%
  mutate(Total_Abundance = sum(Measurement),
    Species_richness = length(unique(Best_guess_binomial)))
kable(TotaAbundance, "pandoc", caption = "Example of the resulting table", align=rep('c', 8))
```

Table 4: Example of the resulting table

Best_guess_binomial	Diversity_metric	Measurement	SSS	Total_Abundance	Species_richness
Caryota urens	abundance	2	BS1_2010__Page 1 1	10	3
Ficus hispida	abundance	1	BS1_2010__Page 1 1	10	3
Xanthophyllum flavescens	abundance	7	BS1_2010__Page 1 1	10	3

2) Example of different abundance and species richness results

```
TotaAbundance2 <- diversity %>% filter(SSS == "BS1_2010__Page 1 2") %>% select(Best_guess_binomial, Diversity_metric,
  Measurement, SSS) %>%
  mutate(Total_Abundance = sum(Measurement), Species_richness = length(unique(Best_guess_binomial)))
kable(TotaAbundance2, "pandoc", caption = "Example of the resulting table", align=rep('c', 8))
```

Table 5: Example of the resulting table

Best_guess_binomial	Diversity_metric	Measurement	SSS	Total_Abundance	Species_richness
Caryota urens	abundance	1	BS1_2010__Page 1 2	4	2
Xanthophyllum flavescens	abundance	3	BS1_2010__Page 1 2	4	2

The Predominant_land_use column holds the land use categories. They often need some tidying up.

```
sites <- sites %>%
  mutate(
    # collapse primary forest and non-forest together into primary vegetation as these aren't well dist
    Predominant_land_use = recode_factor(Predominant_land_use,
      "Primary forest" = "Primary",
      "Primary non-forest" = "Primary"),
```

```

# indeterminate secondary veg and cannot decide get NA
Predominant_land_use = na_if(Predominant_land_use, "Secondary vegetation (indeterminate age)",
Predominant_land_use = na_if(Predominant_land_use, "Cannot decide"),
Use_intensity = na_if(Use_intensity, "Cannot decide"),

# set reference levels
Predominant_land_use = factor(Predominant_land_use),
Predominant_land_use = relevel(Predominant_land_use, ref = "Primary"),
Use_intensity = factor(Use_intensity),
Use_intensity = relevel(Use_intensity, ref = "Minimal use")
)

# take a look at the LandUse/Use intensity split
table(sites$Predominant_land_use, sites$Use_intensity)

```

```

##
##               Minimal use Intense use Light use
## Primary               954         170       394
## Cropland              112         126        21
## Intermediate secondary vegetation  174          4       156
## Mature secondary vegetation      84          6       147
## Pasture               261         13       158
## Plantation forest      266         62       706
## Young secondary vegetation  173          0        48
## Urban                 0          1        40

```