Survey123 plot data

Norwegian institute for nature research

2024-09-18

This script converts data imported from Survey123 into the format used in the masterfile: Herminium_monorchis_masterfile_v_2024.

Before getting started, we can run this line to make sure R can read Norwegian letters.

```
Sys.setlocale(locale='no_NB.utf8')
```

```
[1] "LC_COLLATE=no_NB.utf8;LC_CTYPE=no_NB.utf8;LC_MONETARY=no_NB.utf8;LC_NUMERIC=C;LC_TIME=no
```

First, we need to load the required packages.

```
library(tidyverse)
library(readxl)
library(vegan)
library(car)
library(goeveg)
library(ggrepel)

#NB use install.packages("package name") if the package is not installed yet.
#Using qmd requires the "rmarkdown" package installed.
```

We can import the raw data from Survey123.

```
env <- read_excel("P:/154027_effektovervaking_av_trua_arter_og_naturtyper_2024/04 Honningblom
species <- read_excel("P:/154027_effektovervaking_av_trua_arter_og_naturtyper_2024/04 Honning
```

The job at hand is to combine these two datasets, and create one dataset that we can analyze in R.

The column named ParentGlobalID in the species data matches with GlobalID in the env data. GlobalID in the species data is a unique identifier in that dataset, and we do not need it at this point. We can thus use GlovalID (env) and ParentGlobalID (species) as identifiers to merge the data. We can call it ParentGlobalID in both datasets for ease of joining. Multiple rows of data per GlobalParent may be created (e.g., multiple species per GlobalParentID).

```
#First we need to fix the plot names
env <- env %>%
tidyr::separate("Rute-ID", into = c("Locality_ID", "Plot_number"), "(?<=[A-Z])(?=[0-9])", fi
mutate(Locality_ID = if_else(`Polygon ID` == "Filletassen", "FI",
                             if_else(`Polygon ID` == "Skipstadsand", "SKI",
                                     if_else(`Polygon ID` == "Skjellvik", "SKJ",
                                             if_else(`Polygon ID` == "Teneskjær", "TEN", NA)
unite(Plot_ID_new, Locality_ID:Plot_number, sep = "", remove = FALSE)  #creates new Plot_ID

#Now we can rename the GlobalID so we can join the datafiles

env <- env %>%
  rename(ParentGlobalID = GlobalID)

#Join the species dataset to the environmental dataset

df_plot_plants <- env %>%
  left_join(species, by = "ParentGlobalID")
```

Now we have a datafile with a lot of columns we do not need. To match the master datafile for Herminium monorchis, we only need:

- Date (date of registration)
- Locality (full locality name)
- Locality_ID (abbreviation for the locality)
- Plot_ID (plot number)
- Plot_ID_new (consists of the locality ID and plot number)
- Count Herminium monorchis, vegetative (we do not strictly need this in this dataset, but it is needed to filter out obligatory H. orchis registrations in a quadrat for plots that did not have H. monorchis. At these plots, no full species composition was surveyed, so we want to remove them or substitute with NAs)
- Count Herminium monorchis, fertile
- Species_norwegian (norwegian species name)
- Presence_quadrat (the quadrats the species is registered in, formatted as #,#,#,etc)
- Quadrats (the number of quadrats the species is present in - this one does not exist in the S123 data, so needs to be created later)
- Cover_vascular.plants

- Cover_DOM
- Cover_bare.soil (we do not have this on in S123)
- Cover_bryophytes
- Cover_woody_fieldlayer
- Cover_shrubs
- Height_1
- Height_2
- Height_3
- Height_4
- Comments (any comments to species registrations)

```
#NB use `variabe name` when the variable name (column) contains spaces or other signs that a

df_plot_plants <- df_plot_plants %>%
  select(`Registreringsdato:`,
         `Polygon ID.x`,
         Locality_ID,
         Plot_number,
         Plot_ID_new,
         `Antall honningblom:vegetative planter`,
         `Antall honningblom:fertile planter`,
         sp_species,
         `Smårutene nr.`,
         `Kommentarer:...10`,
         `Kommentarer:...11`,
         `Kommentarer:...18`,
         `Kommentarer:...29`,
         `Dekning % av karplanter i feltsjikt`,
         `Dekning % av strø`,
         `Dekning % av bunnsjikt`,
         `Dekning % av vedplanter i feltsjikt`,
         `Dekning % av vedplanter i busksjikt`,
         `Vegetasjonshøyde måling 1`,
         `Vegetasjonshøyde måling 2`,
         `Vegetasjonshøyde måling 3`,
         `Vegetasjonshøyde måling 4`)

#Now we can rename the non-compatible column names

df_plot_plants <- df_plot_plants %>%
  dplyr::rename("Date" = `Registreringsdato:`, #rename uses format new name = old name
         "Locality" = `Polygon ID.x`,
         "Locality_ID" = Locality_ID,
         "Plot_ID" = Plot_number,
         "Plot_ID_new" = Plot_ID_new,
         "Herminium_count_vegetative" = `Antall honningblom:vegetative planter`,
         "Herminium_count_fertile" =  `Antall honningblom:fertile planter`,
         "Species_both" =  sp_species, #to be splitted in Norwegian and scientific names late
         "Presence_quadrat" = `Smårutene nr.`,
         "Comment_1" = `Kommentarer:...10`,
         "Comment_2" = `Kommentarer:...11`,
         "Comment_3" = `Kommentarer:...18`,
         "Comment_4" = `Kommentarer:...29`,
         "Fieldlayer_cover" = `Dekning % av karplanter i feltsjikt`,
         "Litter_cover" = `Dekning % av strø`,
         "Groundlayer_cover" = `Dekning % av bunnsjikt`,
         "Woody_fieldlayer_cover" = `Dekning % av vedplanter i feltsjikt`,
         "Shrub_cover" = `Dekning % av vedplanter i busksjikt`,
         "Vegetation_height_1" = `Vegetasjonshøyde måling 1`,
         "Vegetation_height_2" = `Vegetasjonshøyde måling 2`,
         "Vegetation_height_3" = `Vegetasjonshøyde måling 3`,
         "Vegetation_height_4" = `Vegetasjonshøyde måling 4`)
```

There are some difficulties in the data set. Problem 1: some plots were not sampled for plant community composition, but only for the ecological variables. To be able to complete the S123 form, we had to register a species in one of the quadrats: H. monorchis in square 1 (in hindsight, a completely different species should have been used) . Problem 2: some plots were sampled for community composition by a masterstudent, while the S123 contains counts of H. monorchis and ecological variables for these plots.

For problem 1, the "false" registration in the ruter should be removed. This can be done by finding the plots where H. monorchis was registered in a quadrat, but counts were zero.

```
df_plot_plants <- df_plot_plants %>%
  mutate(Presence_quadrat = if_else((Species_both %in% "Herminium monorchis (honningblom)") &
```

For problem 2, we want to create the column "Sampling" will describe whether vegetation composition data was: not recorded, recorded in S123, or recorded by MSc student in analog format. This can be used to sort the dataset later.

```
#The MSc student has sampled all plots at Skjellvik and Skipstadsand, at the other localitie

df_plot_plants <- df_plot_plants %>%
  mutate(Sampling = if_else(Locality %in% "Skipstadsand" | Locality %in% "Skjellvik", "MSc",
#To find plots that were not sampled for vegetation composition, and assign that they were no
df_plot_plants <- df_plot_plants %>%
    group_by(Plot_ID_new) %>%
    mutate(Sampling = if_else(n() == 1 & (Locality %in% "Filletassen" | Locality %in% "Tenesk
    ungroup()

#And now lets assign the plots that were sampled in S123
df_plot_plants <- df_plot_plants %>%
  ungroup() %>%
  group_by(Plot_ID_new) %>%
  mutate(Sampling = if_else(n() >= 2 & (Locality %in% "Filletassen" | Locality %in% "Teneskja
```

Count the number of quadrats with registrations (required: stringr).

```
#| label: Count quadrat registrations
#| warning: false

df_plot_plants <- df_plot_plants %>%
  mutate(Quadrats = str_count(Presence_quadrat,",")+1)
```

Split up the species names in latin and norwegian. We want a column with the norwegian name, and one with the scientific name

```
df_plot_plants <- df_plot_plants %>%
  mutate(Species_norwegian = sapply(str_extract_all(Species_both, "(?<=\\()[^)(]+(?=\\))"), p
         Species_scientific = gsub("\\([^)(]+\\)", "", Species_both))
```