# Exploration of FELLOW trait dataset

The objective of this document is to:

- visually explore the trait database
- understand trait coverage / data gaps
- check for possible inconsistencies

Description of trait databases

So far, we compiled 162 traits for 1706 taxa gathered from 15 trait databases. But there are many missing values.

Number of missing species

Number of missing traits values
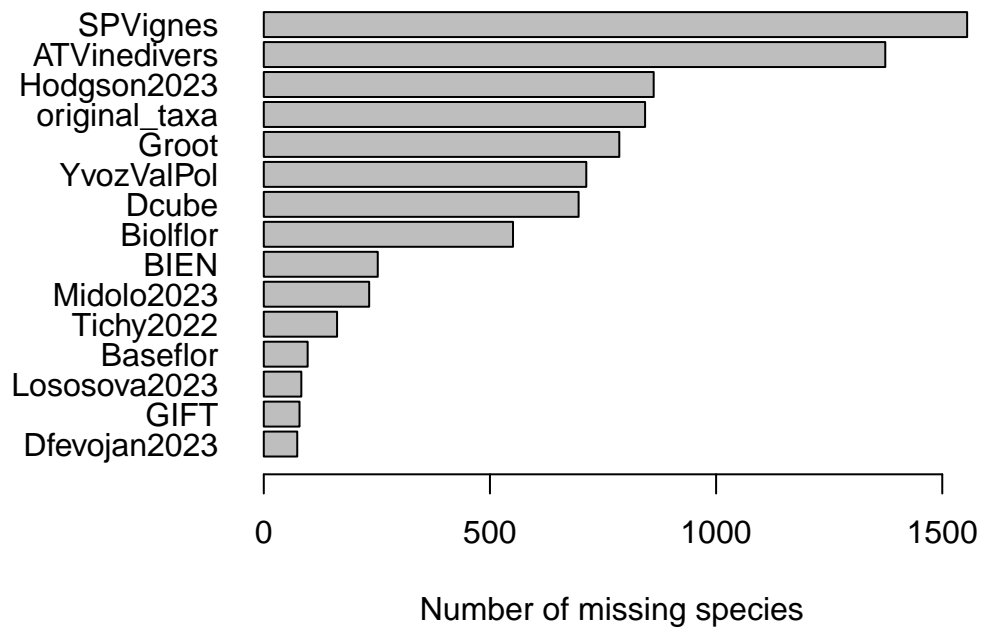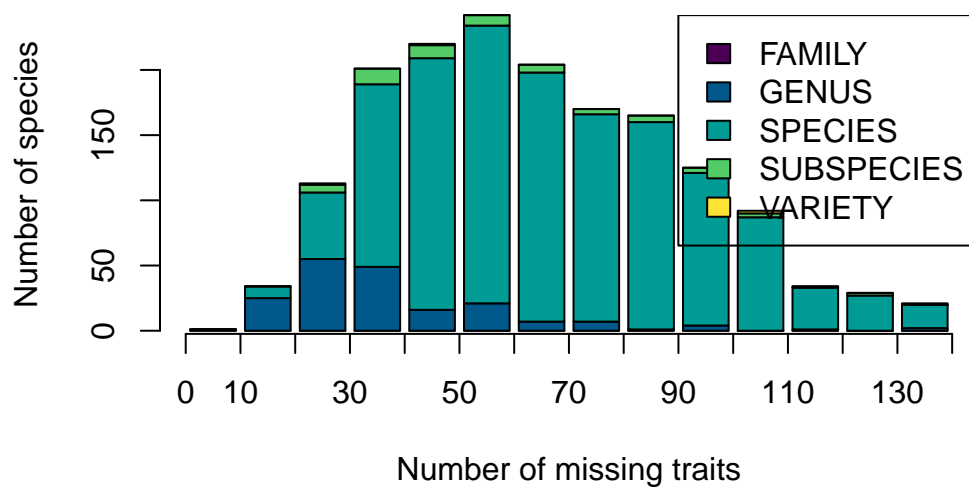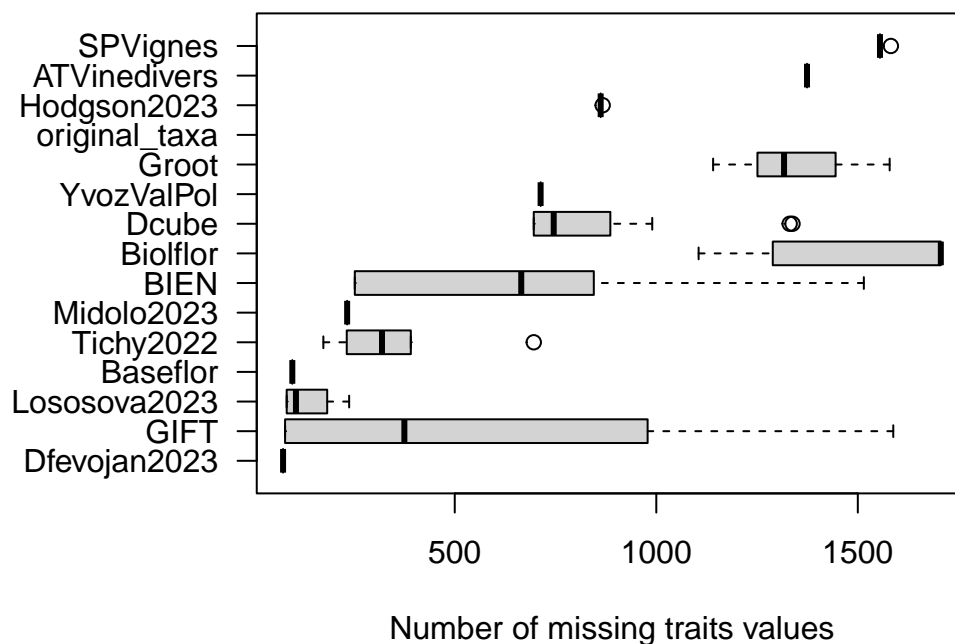
Taxa with no or limited trait information (N=22).

```
 [1] "Acacia"              "Agrimonia agrimonoides" "Agropyron"
 [4] "Amaranthaceae"       "Apiaceae"               "Asparagaceae"
 [7] "Brassicaceae"        "Caryophyllaceae"        "Crambe abyssinica"
[10] "Dysphania aristata"  "Glyceria"               "Lamiaceae"
[13] "Liliaceae"           "Paronychia"             "Piptatherum"
[16] "Poaceae"             "Roemeria hispida"       "Rosaceae"
[19] "Rubiaceae"           "Salix"                  "Sanguisorba"
[22] "Viburnum"
```
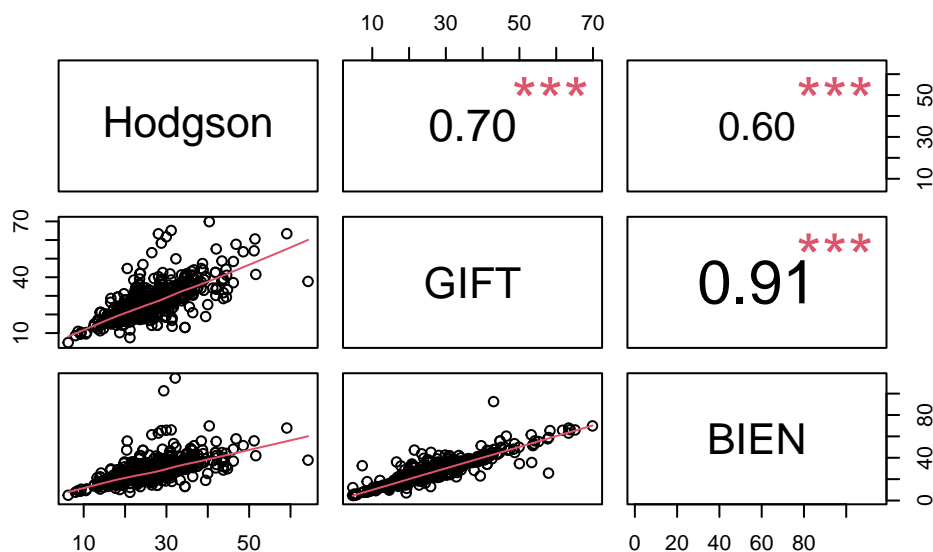
Open question: **How to deal with families taxa?**
How to deal with missing trait values? Trait imputation, discarding taxa, ...

Comparison

SLA

There are three sources of information for Specific leaf area (SLA) : Hodgson et al. 2023 (in mm2/mg), GIFT (in cm2/g) and BIEN (in m2/kg = mm2/mg).
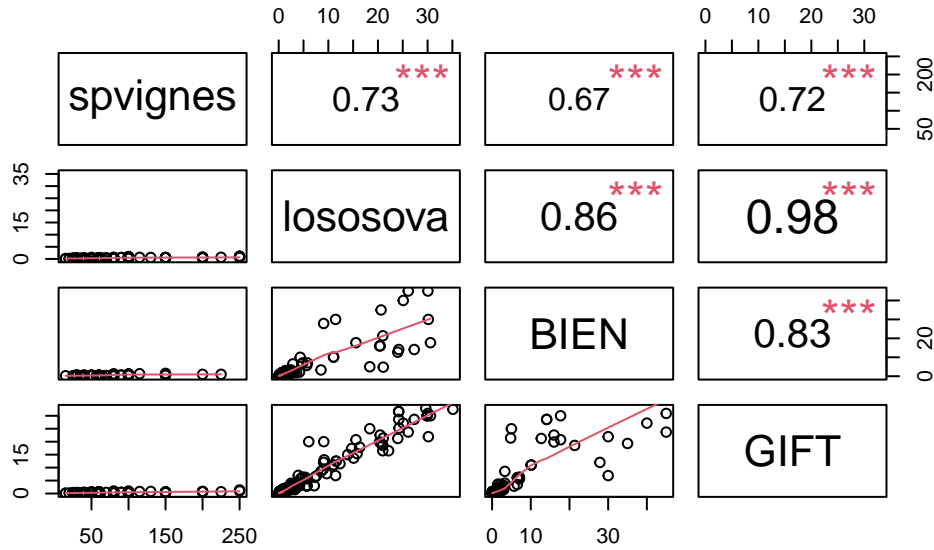

Values are highly correlated, so we could imagine filling the missing values (using preferred data sources or averaging them).

Number of NAs:

```
Hodgson     GIFT     BIEN  filled
    862      731      720     475
```

Plant height

There are four sources of information for plant height.

Not sure what is the unit of `Plant_Height` in SPVignes (cm?), and why it is limited to only small plants (<1m). Yet we could potentially fill the height trait values.

Number of NAs:

```
spvignes lososova    BIEN    GIFT  filled
    1582      129    1276     492    109
```