# 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

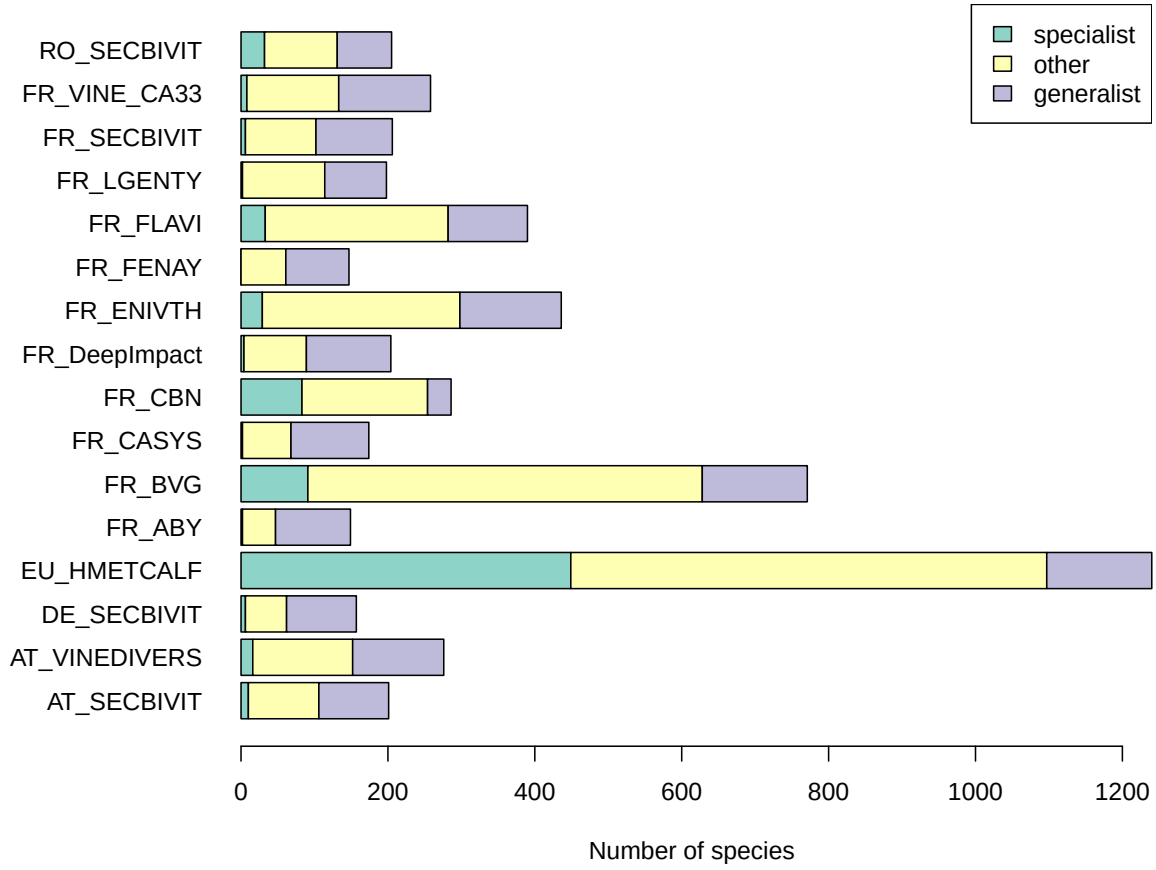
## Description of the species list

We compiled the species lists from 16 datasets. After cleaning and harmonization, there were 1706 unique taxa.

| FAMILY | GENUS | SPECIES | SUBSPECIES | VARIETY |
|---|---|---|---|---|
| 12 | 197 | 1430 | 62 | 5 |

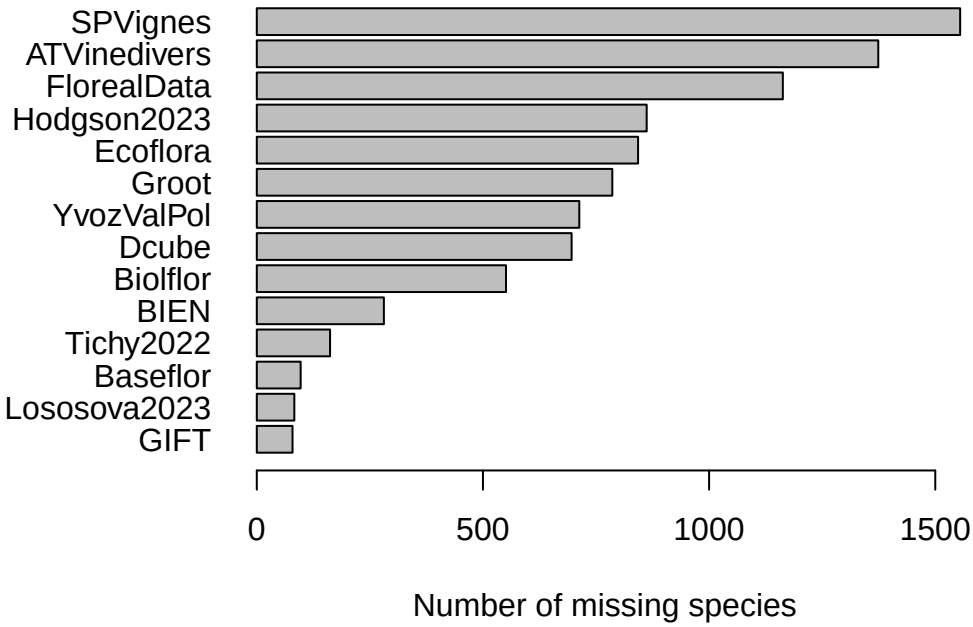
Let's define:
*specialist*: a taxa that occured only in a singe database
*generalist*: a taxa that is listed in 50% of the databases (8 out of 16)

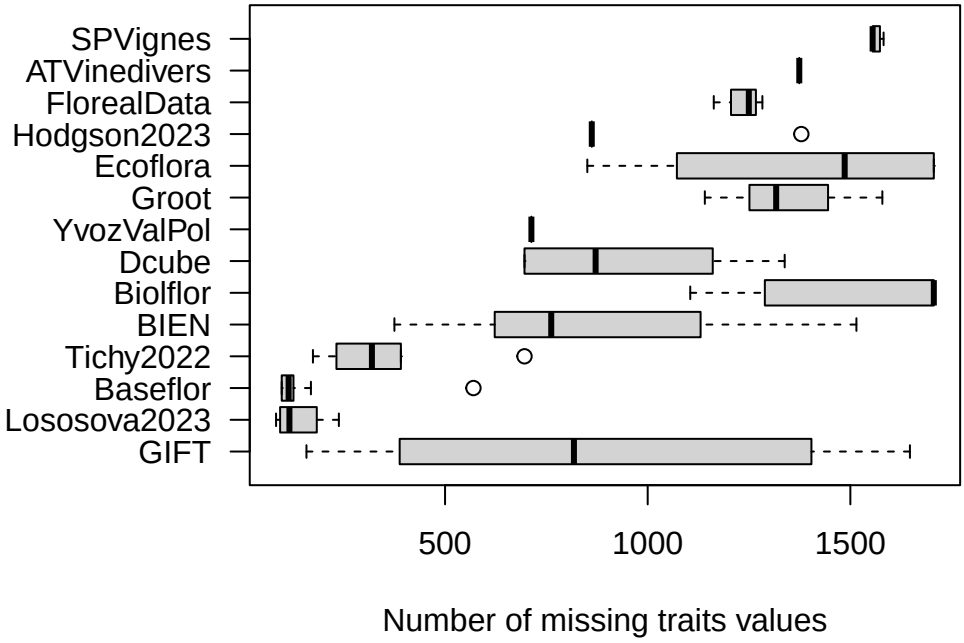
```
sp_class
specialist      other generalist
       773        789        144
```

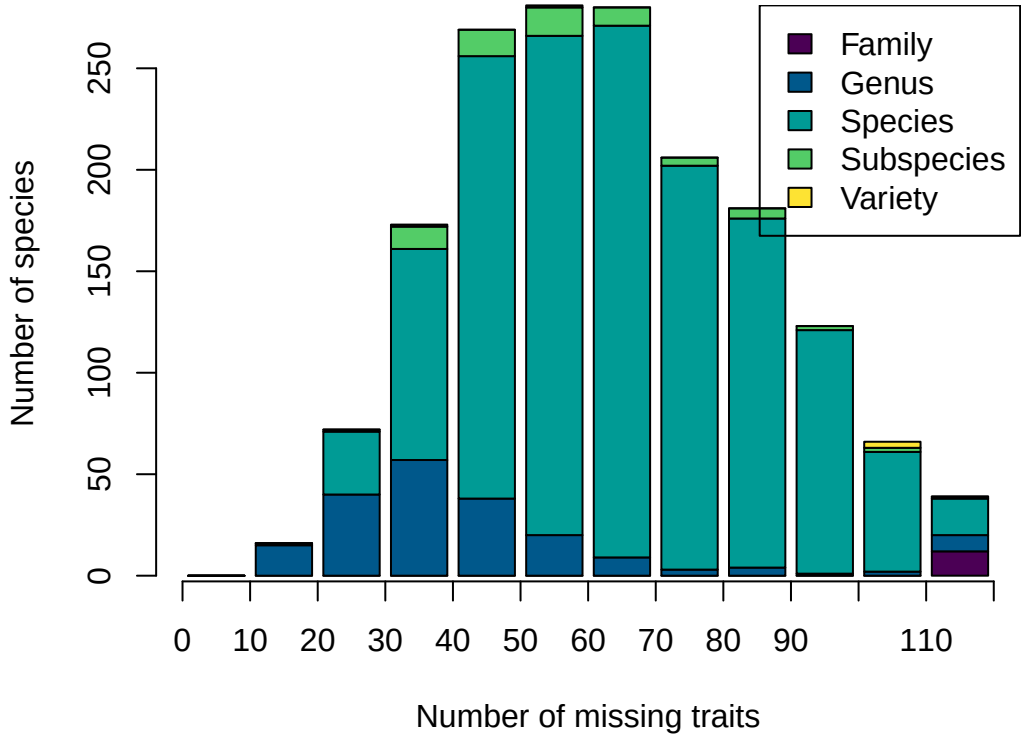


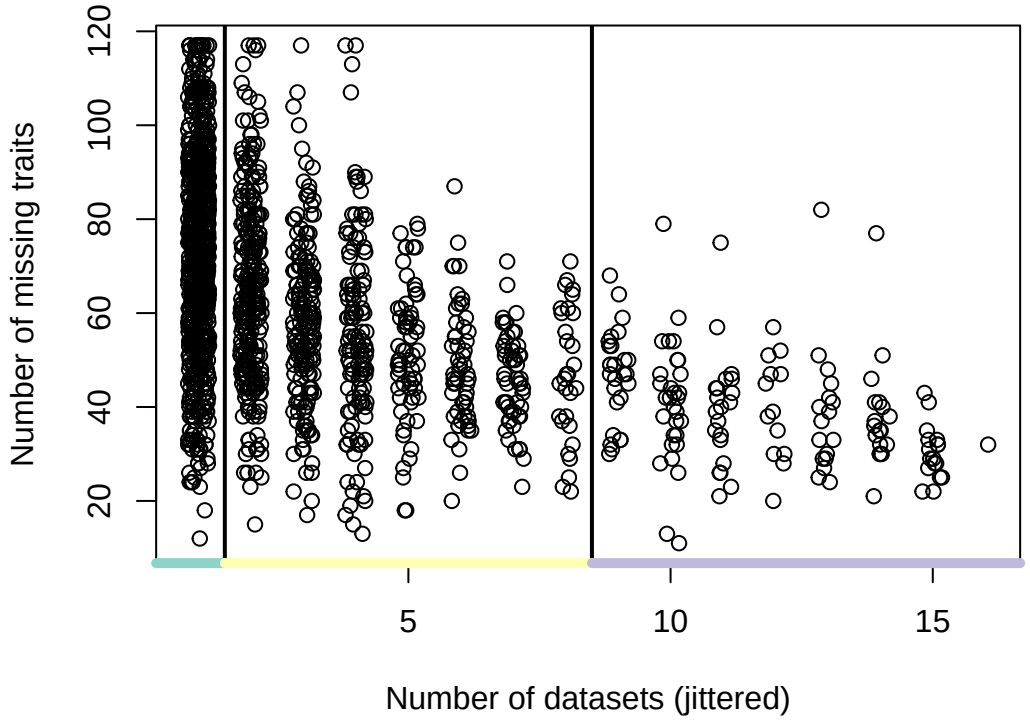
## Description of trait databases

So far, we compiled 117 traits for 1706 taxa gathered from 14 trait databases. But there are many missing values.




Family
Genus
Species
Subspecies
Variety
Number of species
0
50
100
150
200
250
0
10
20
30
40
50
60
70
80
90
110
Number of missing traits


Number of missing traits
20
40
60
80
100
120
5
10
15
Number of datasets (jittered)

Taxa with no or limited trait information (N=20).

```
 [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"             "Viburnum"
```

Open question:

**How to deal with families taxa?**

**How to deal with missing trait values?** Trait imputation, discarding taxa, …

**Summary of trait completness**

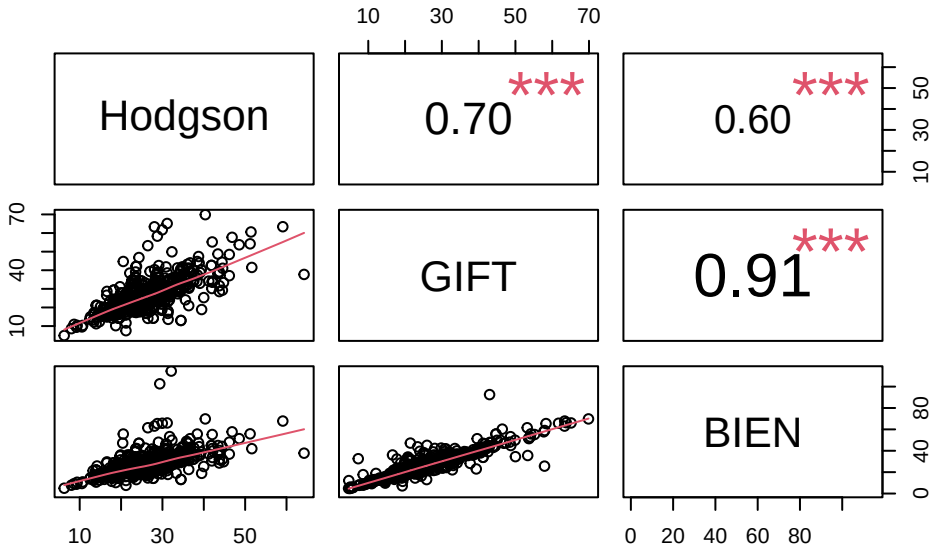
| Trait | N database | N taxa | Completness (%) |
|---|---|---|---|
| Growth.form | 5 | 1671 | 98 |
| Dispersal.mode | 4 | 1660 | 97 |
| Plant.height | 8 | 1623 | 95 |
| Chorology | 1 | 1609 | 94 |
| Habitat | 1 | 1609 | 94 |
| Dispersal.distance | 1 | 1603 | 94 |
| Sexuality | 1 | 1600 | 94 |
| Fruit.type | 2 | 1599 | 94 |
| Pollination | 1 | 1599 | 94 |
| Flower.color | 6 | 1591 | 93 |
| Inflorescence | 2 | 1586 | 93 |
| Lifecycle | 2 | 1570 | 92 |
| Ellenberg.Salinity | 1 | 1532 | 90 |
| Seed.mass | 5 | 1530 | 90 |
| Ellenberg.Light | 1 | 1474 | 86 |
| Flowering | 4 | 1467 | 86 |
| Ellenberg.Reaction | 1 | 1446 | 85 |
| Photosynthetic.pathway | 1 | 1445 | 85 |
| Ellenberg.Moisture | 1 | 1327 | 78 |
| Ellenberg.Nutrients | 1 | 1315 | 77 |
| SLA | 3 | 1231 | 72 |
| Pollination.syndrome | 3 | 1196 | 70 |
| Flower.UV.reflectance | 3 | 1118 | 66 |
| Diaspore.exposure | 1 | 1010 | 59 |
| Diaspore.type | 1 | 1010 | 59 |
| Ellenberg.Temperature | 1 | 1010 | 59 |
| Leaf.area | 2 | 1009 | 59 |
| Floral.symmetry | 1 | 993 | 58 |
| Flower.class | 1 | 993 | 58 |
| Flower.type | 1 | 993 | 58 |
| Nectar.quantity | 1 | 993 | 58 |
| Pollen.quantity | 1 | 993 | 58 |
| PV.Bees | 1 | 993 | 58 |
| PV.Bumblebees | 1 | 993 | 58 |
| PV.butterflies | 1 | 993 | 58 |
| PV.Hoverflies | 1 | 993 | 58 |
| Leaf.dry.mass.content | 1 | 891 | 52 |
| Anemochory | 1 | 885 | 52 |
| Leaf.dry.mass | 1 | 861 | 50 |

| Trait | N database | N taxa | Completness (%) |
|---|---|---|---|
| Canopy.diameter | 1 | 844 | 49 |
| Canopy.height | 1 | 844 | 49 |
| Epizoochory | 1 | 784 | 46 |
| Diaspore.mass | 1 | 716 | 42 |
| Leaf.width | 1 | 653 | 38 |
| Lifeform | 2 | 591 | 35 |
| Root.mycorrhizal.colonization | 1 | 565 | 33 |
| Root.mass.fraction | 1 | 515 | 30 |
| Seed.length | 1 | 498 | 29 |
| Grassland.specialization | 1 | 458 | 27 |
| Specific.root.length | 1 | 455 | 27 |
| Flower.length | 2 | 454 | 27 |
| Root.diameter | 1 | 412 | 24 |
| Root.tissue.density | 1 | 390 | 23 |
| Root.lateral.spread | 1 | 388 | 23 |
| Diaspore.height | 1 | 375 | 22 |
| Strategy | 3 | 375 | 22 |
| Hydrochory | 1 | 368 | 22 |
| Root.N.concentration | 1 | 338 | 20 |
| Plant.lifespan | 2 | 332 | 19 |
| Vegetative.propagation | 1 | 327 | 19 |
| Root.depth | 2 | 310 | 18 |
| Leaf.nitrogen.content | 1 | 290 | 17 |
| Root.C.concentration | 1 | 261 | 15 |
| Leaf.length | 1 | 253 | 15 |
| Root.C.N.ratio | 1 | 225 | 13 |
| Leaf.carbon.to.nitrogen.content | 1 | 191 | 11 |
| Germination | 2 | 151 | 9 |
| Flower.width | 2 | 132 | 8 |
| Fruit.color | 1 | 59 | 3 |

## 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).



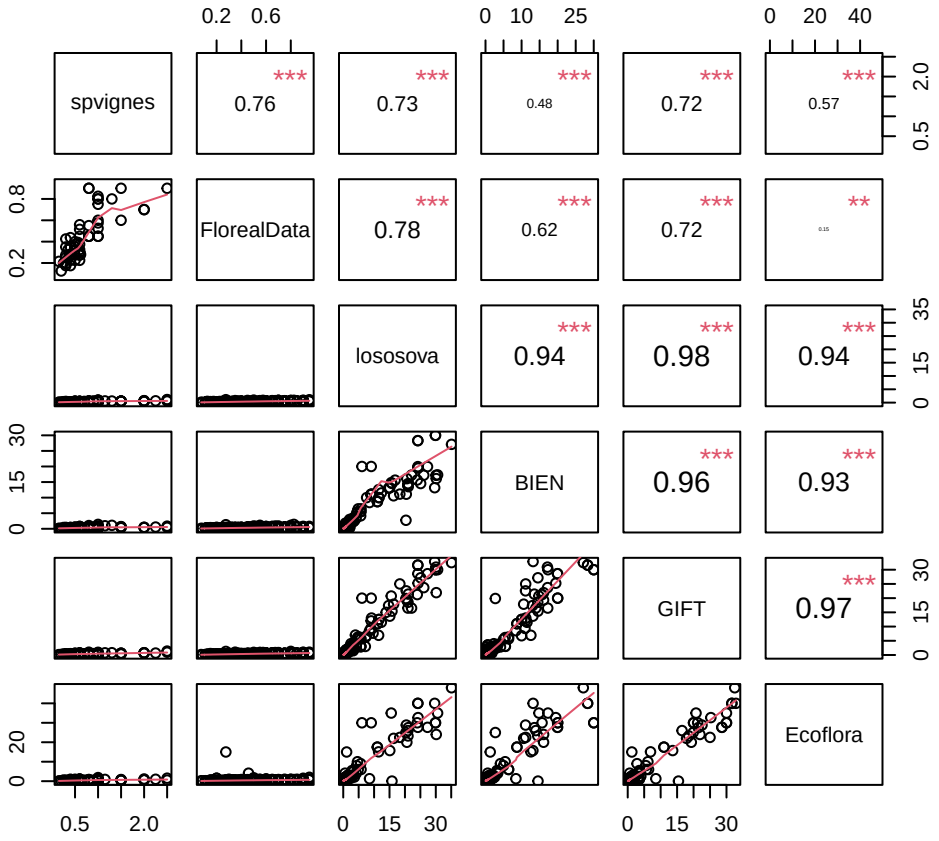

Values are highly correlated, so we could imagine filling the missing values (using preferred data sources or averaging them).

**Number of NAs:**

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

### Plant height

There are six sources of information for plant height.



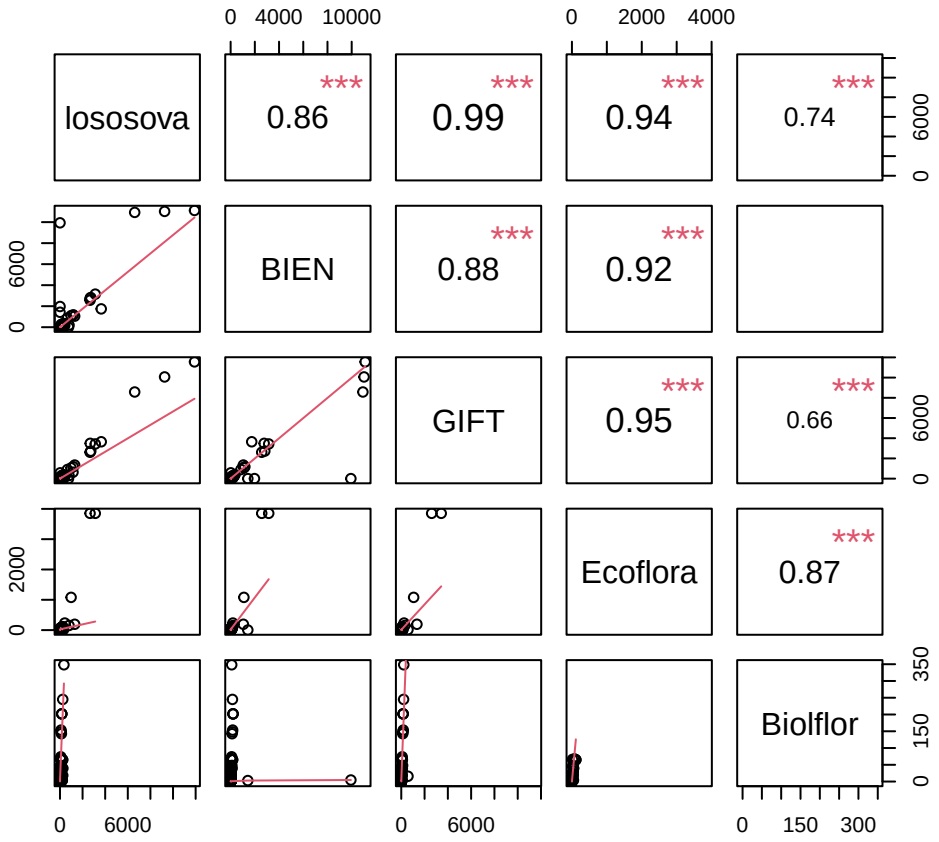

SPVignes and FlorealData are limited to small plants (<1m) (no trees) *but FlorealData>100cm must be clean.*

**Number of NAs:**

```
   spvignes FlorealData    lososova        BIEN        GIFT    Ecoflora
       1582        1345         129         610         492         851
     filled
        102
```

### Seed mass

There are five sources of information for seed mass




**Number of NAs:**

```
lososova      BIEN      GIFT Ecoflora Biolflor    filled
     238       589       284      1072      1289       176
```

### Flower colour

There are five sources of information for flower colour, but it must be cleaned

| Baseflor | BIEN | GIFT | FlorealData | Biolflor | YvozValPol |
|---|---|---|---|---|---|
| 169 | 1483 | 1444 | 1251 | 1706 | 713 |

baseflor

| | | | |
|---|---|---|---|
| Blanc | Blanc, jaune | Blanc, jaune, bleu | Blanc, jaune, rose |
| 279 | 27 | 3 | 3 |
| Blanc, rose | Blanc, vert, rose | Bleu | Bleu, blanc |
| 40 | 1 | 130 | 10 |
| Bleu, blanc, rose | Bleu, jaune | Bleu, jaune, rose | Bleu, rose |
| 8 | 4 | 2 | 8 |
| Jaune | Jaune, rose | Jaune, vert | Marron |
| 412 | 3 | 1 | 19 |
| Noir | Rose | Vert | Vert, bleu |
| 2 | 251 | 208 | 11 |
| Vert, jaune, rose | Vert, rose | <NA> | |
| 1 | 30 | 253 | |