

# 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 24 datasets. After cleaning and harmonization, there were 2027 unique taxa.

FAMILY	GENUS	SPECIES	SUBSPECIES	VARIETY
15	241	1640	120	11

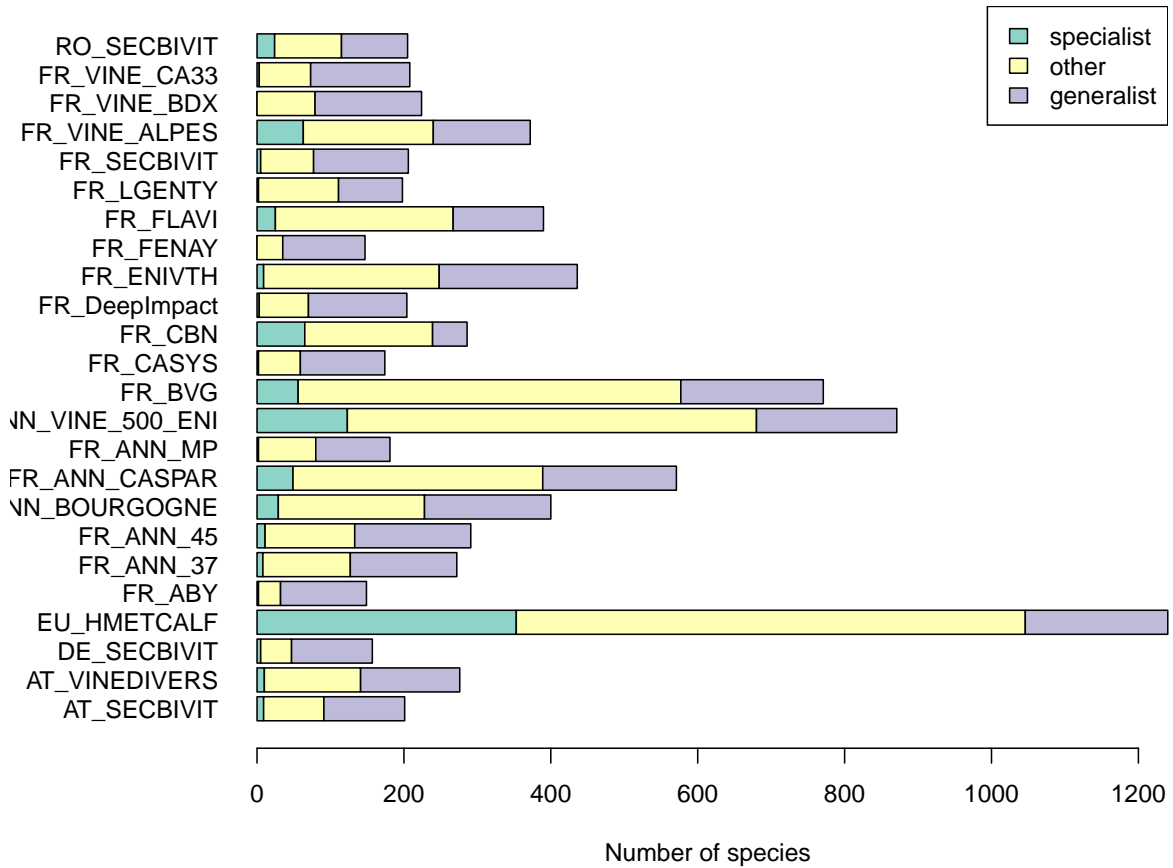
Let's define:

*specialist*: a taxa that occurred only in a single database

*generalist*: a taxa that is listed in 50% of the databases (12 out of 24)

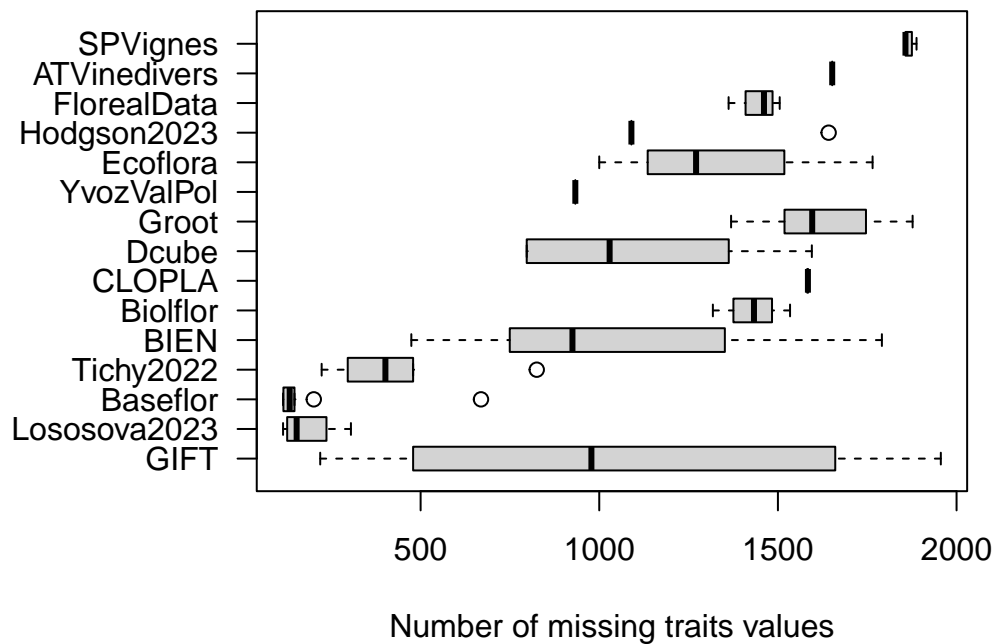
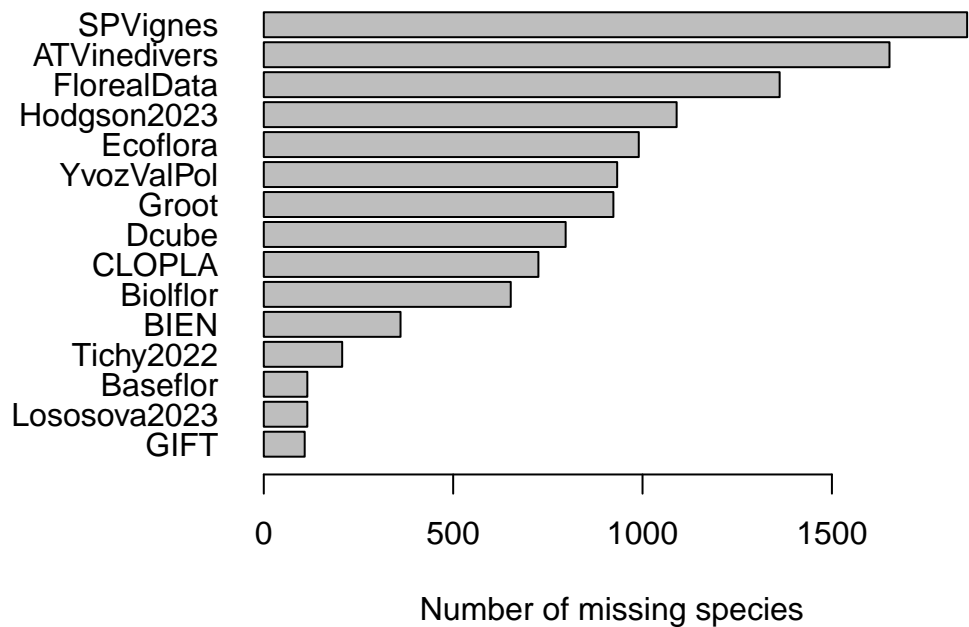
sp\_class

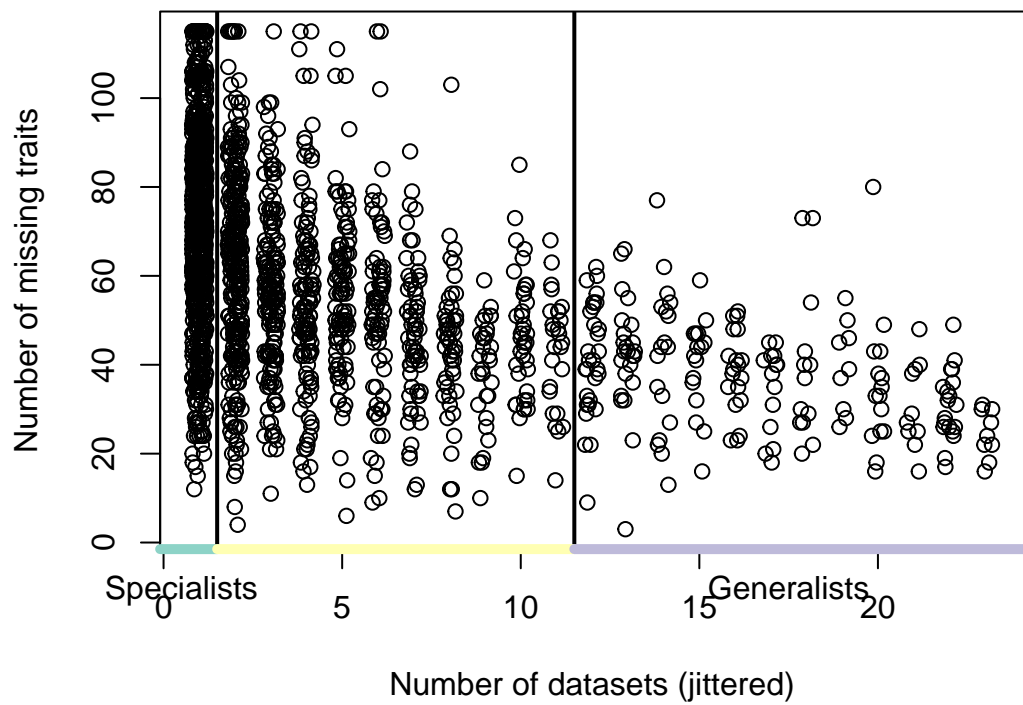
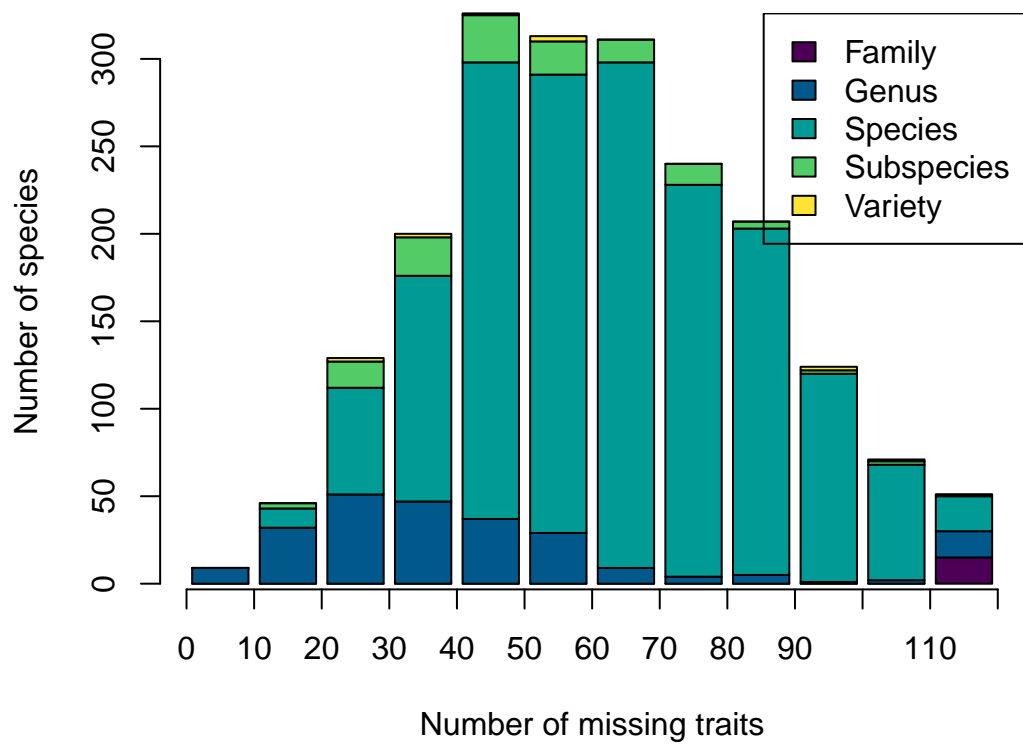
specialist	other	generalist
858	973	196



## Description of trait databases

So far, we compiled 115 traits for 2027 taxa gathered from 15 trait databases. But there are many missing values.





Taxa with no or limited trait information (N=39).

[1] "Abies"	"Acacia"
[3] "Agrimonia agrimonoides"	"Agropyron"
[5] "Amaranthaceae"	"Apiaceae"
[7] "Asparagaceae"	"Aster"
[9] "Boraginaceae"	"Brassicaceae"
[11] "Bryum dichotomum"	"Caryophyllaceae"
[13] "Chaenomeles x superba"	"Chrysanthemum"
[15] "Cochlearia"	"Cosmos"
[17] "Crambe abyssinica"	"Dysphania aristata"
[19] "Geraniaceae"	"Glyceria"
[21] "Imbriobryum subapiculatum"	"Lamiaceae"
[23] "Lavandula"	"Leontodon autumnale"
[25] "Liliaceae"	"Lunaria"
[27] "Moehringia"	"Orchis"
[29] "Paronychia"	"Piptatherum"
[31] "Poaceae"	"Primulaceae"
[33] "Pulmonaria"	"Rhizogemma staphylina"
[35] "Riccia sorocarpa"	"Riccia warnstorffii"
[37] "Roemeria hispida"	"Rosaceae"
[39] "Rubiaceae"	

Open question:

**How to deal with families taxa?**

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

## Summary of trait completeness

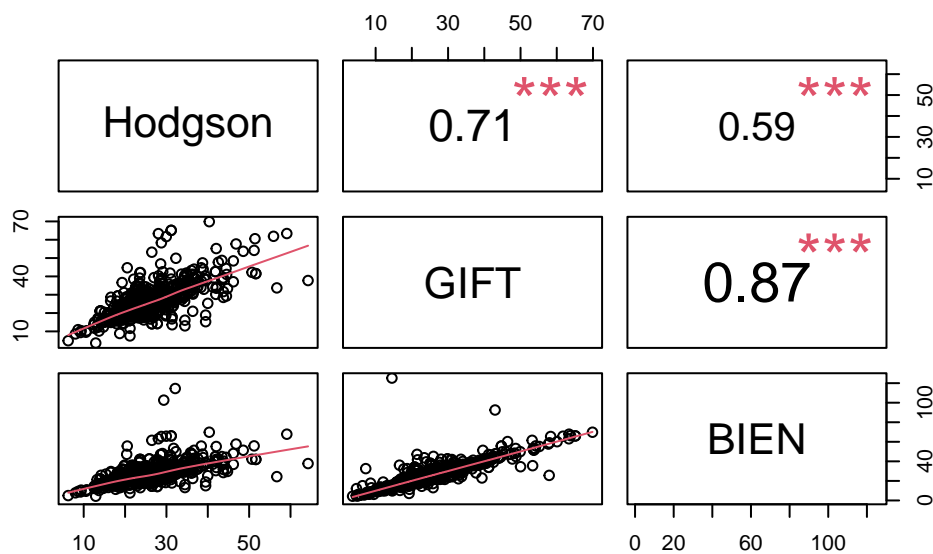
Trait	N database	N taxa	Completeness (%)
Growth.form	5	1974	97
Dispersal.mode	4	1961	97
Plant.height	8	1916	95
Chorology	1	1912	94
Habitat	1	1911	94
Sexuality	1	1902	94
Fruit.type	2	1901	94
Pollination	1	1901	94
Dispersal.distance	1	1889	93
Flower.color	5	1889	93
Inflorescence	2	1884	93
Lifecycle	2	1841	91
Ellenberg.Salinity	1	1804	89
Seed.mass	5	1799	89
Flowering	3	1734	86
Ellenberg.Light	1	1731	85
Ellenberg.Reaction	1	1691	83
Photosynthetic.pathway	1	1691	83
Ellenberg.Moisture	1	1560	77
Ellenberg.Nutrients	1	1548	76
SLA	3	1438	71
Pollination.syndrome	3	1404	69
Flower.UV.reflectance	3	1279	63
Diaspore.exposure	1	1230	61
Diaspore.type	1	1230	61
Ellenberg.Temperature	1	1202	59
Leaf.area	2	1188	59
Floral.symmetry	1	1094	54
Flower.class	1	1094	54
Flower.type	1	1094	54
Nectar.quantity	1	1094	54
Pollen.quantity	1	1094	54
PV.Bees	1	1094	54
PV.Bumblebees	1	1094	54
PV.butterflies	1	1094	54
PV.Hoverflies	1	1094	54
Anemochory	1	1058	52
Leaf.dry.mass.content	1	1043	51
Leaf.dry.mass	1	1000	49

Trait	N database	N taxa	Completeness (%)
Epizoochory	1	938	46
Canopy.diameter	1	937	46
Canopy.height	1	937	46
Diaspore.mass	1	864	43
Leaf.width	1	746	37
Lifeform	2	716	35
Root.mycorrhizal.colonization	1	658	32
Root.mass.fraction	1	609	30
Seed.length	1	594	29
Grassland.specialization	1	570	28
Specific.root.length	1	535	26
Flower.length	2	515	25
Root.diameter	1	482	24
Root.lateral.spread	1	473	23
Diaspore.height	1	465	23
Root.tissue.density	1	456	22
lateral.spread	1	447	22
offspring	1	443	22
offspring.wsmall	1	443	22
clonal.index	1	442	22
Hydrochory	1	432	21
Strategy	2	423	21
Root.N.concentration	1	406	20
Vegetative.propagation	1	385	19
Plant.lifespan	1	375	19
Root.depth	2	371	18
Leaf.nitrogen.content	1	351	17
Root.C.concentration	1	315	16
Root.dry.matter	1	294	15
Leaf.length	1	290	14
Root.C.N.ratio	1	267	13
Leaf.carbon.to.nitrogen.content	1	236	12
Germination	2	170	8
Root.length.density	1	150	7
Flower.width	2	145	7
Fruit.color	1	71	4

## Comparison

### SLA

There are three sources of information for Specific leaf area (SLA) : Hodgson et al. 2023 (in mm<sup>2</sup>/mg), GIFT (in cm<sup>2</sup>/g) and BIEN (in m<sup>2</sup>/kg = mm<sup>2</sup>/mg).



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

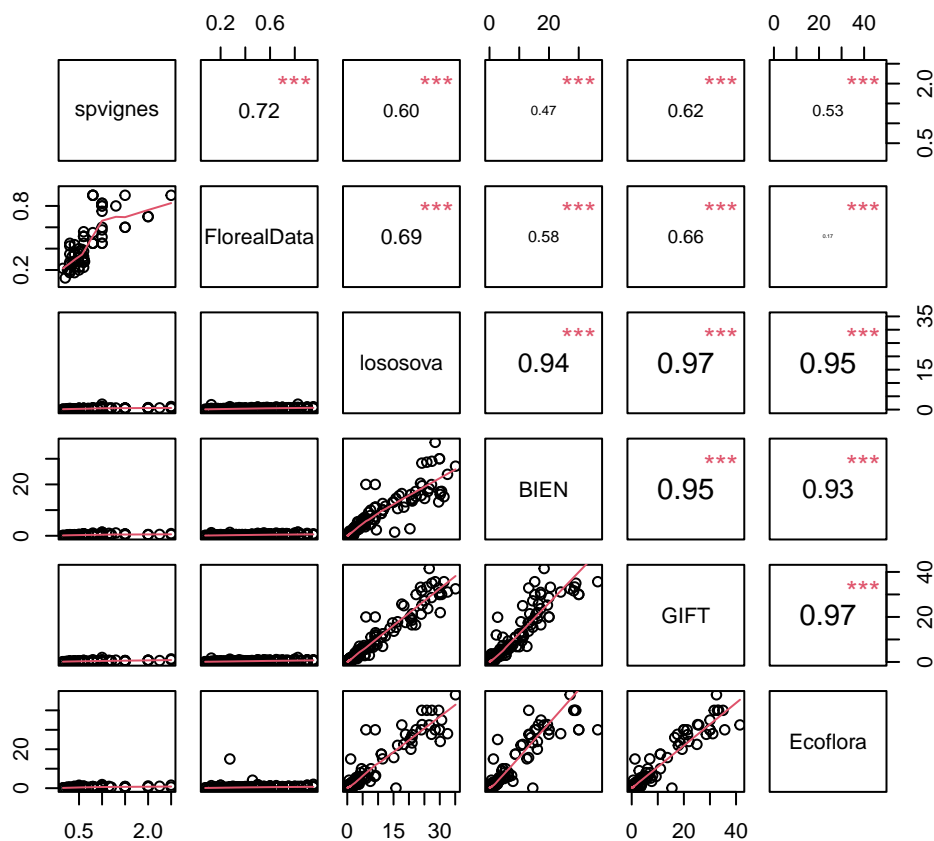
#### Number of NAs:

Hodgson	GIFT	BIEN	filled
1090	875	866	589



## Plant height

There are six sources of information for plant height.



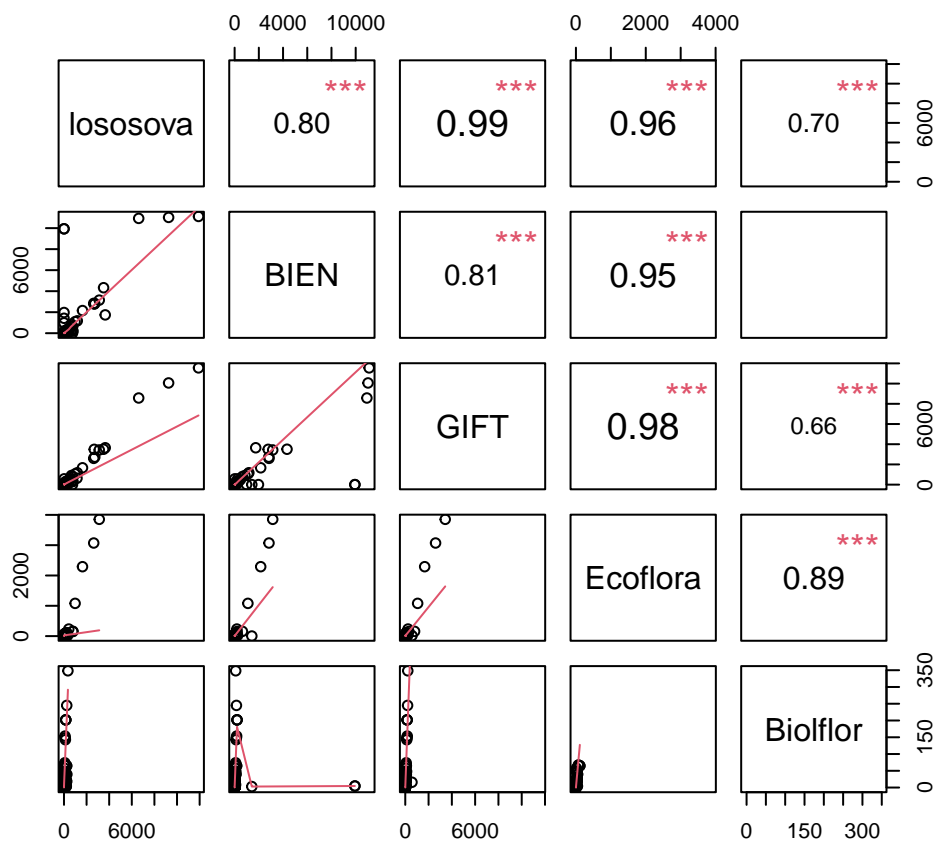
SPVignes and FlorealData are limited to small plants (<1m) (no trees) *but FlorealData>100cm must be clean.*

## Number of NAs:

spvignes	FlorealData	lososova	BIEN	GIFT	Ecoflora
1888	1578	168	732	593	1000
filled					
132					

## Seed mass

There are five sources of information for seed mass



## Number of NAs:

lososova	BIEN	GIFT	Ecoflora	Biolflor	filled
305	728	365	1271	1534	228

## Flower colour

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

```
[1] "Flower.color_Baseflor"          "Flower.color_YvozValPol"
[3] "Flower.color_BIEN"              "Fruit.colour_GIFT"
[5] "Flower.color_GIFT"              "Flower.color_FlorealData"
[7] "Root.mycorrhizal.colonization_Groot"
```

Baseflor	BIEN	GIFT	FlorealData	YvozValPol
201	1769	1731	1465	933

baseflor

Blanc	Blanc, jaune	Blanc, jaune, bleu	Blanc, jaune, rose
347	31	3	4
Blanc, rose	Blanc, vert, rose	Bleu	Bleu, blanc
52	1	148	11
Bleu, blanc, rose	Bleu, jaune	Bleu, jaune, rose	Bleu, rose
11	5	2	10
Jaune	Jaune, rose	Jaune, vert	Marron
470	3	1	23
Noir	Rose	Vert	Vert, bleu
2	306	249	13
Vert, jaune, rose	Vert, rose	<NA>	
1	35	299	