

Leaf Phenology Type

TRY DB Extraction

```
In [ ]: # IMPORTS
import numpy as np
import pandas as pd
from TRYpros import *
```

```
In [43]: # EXTENSIONS
%load_ext autoreload
%autoreload 2
```

The autoreload extension is already loaded. To reload it, use:
%reload_ext autoreload

```
In [44]: # SETTINGS
pd.set_option("display.max_columns", None) # Set pandas to show all columns.
```

```
In [45]: # CONSTANTS
DATA_DIR = "../../__data/try/raw"
PATH_SRC = f"{DATA_DIR}/37522_LeafPhenologyType_27nov24/37522.txt"
FEATURE_NAME = "leaf_phenology_type"
FEATURE_TYPE = "str"
PATH_REF = "./dataset_ids.txt"
DROP_COLS = ["LastName", "FirstName", "Dataset", "Reference",
             "Replicates", "SpeciesName", "ValueKindName",
             "OrigUncertaintyStr", "UncertaintyName",
             "RelUncertaintyPercent"]

# DEFINE FEATURE HANDLER
FH = FeatureHandler(path_src=PATH_SRC, path_ref=PATH_REF,
                   feature_name=FEATURE_NAME, d_type=FEATURE_TYPE)
```

1. Load Data

```
In [46]: # LOAD DATA
        FH.load_big_data(drop_cols=DROP_COLS)
```

```
LOADING BIG DATA:
Reading data ... done.
Cleaning data ... done.
Prepping data ... done.
Noting down unique TraitIDs ... done.
Noting down unique DatasetIDs for referencing later ... done.
ALL DONE :3
```

2. Process Trait Data

2.1. ID Review

```
In [47]: # VIEW KNOWN IDS
        print(f"Keep IDs = {FH.keep_ids['trait']}")
        FH.known_ids['trait']
```

```
Keep IDs = [37.0, 1251.0]
```

```
Out[47]:
```

	id	name
0	37.0	Leaf phenology type
1	1251.0	Plant vegetative phenology (leaf phenology)

2.2. Extract Data

```
In [48]: FH.extract_trait_covariate_data()
```

```
EXTRACTING TRAIT & COVARIATE DATA:
Separating trait and covariate data ... done.
Noting down unique LonLat DataIDs ... done.
Noting down unique Year DataIDs ... done.
Separating standardized and non-standardized data ... done.
Adding column 'value_form' ... done.
Total no. of raw data points. = 3865715
No. of trait data points. = 212793
No. of standardized trait data points. = 0
No. of non standardized trait data points. = 211372
No. of covariate data points with ObservationID matching that of at least 1 row of trait data. = 3652922
No. of standardized such covariate data points. = 484810
No. of non standardized such covariate data points. = 3166862
No. of data points with no data. = 2671
Loaded TraitIDs: [37.0, 1251.0]
Identified LonLat DataIDs: [60, 506, 524, 6959, 4705, 4707, 4710, 4711, 8302, 59, 505, 523, 6958, 4704, 4706, 4708, 4709, 8301]
Identified Year DataIDs: [241, 211, 212, 695, 696, 697, 698, 699, 2571, 6601, 1664, 4688, 4689, 4690, 4691, 4692, 4693]
ALL DONE :3
```

2.3. Manual Investigation

```
In [49]: # Are values numeric/categorical? What forms are they in?
FH.view_units_value_forms(data_type="trait")
```

```

Trait Standardised Units: []
Trait Non-Standardised Units: ['months', 'cm', 'yr', 'days', '-']
Trait Standardised Value Forms: []
Trait Non-Standardised Value Forms: ['y', 'n', 'w', 'd', 'e', 'evergreen', '@', 'semi', 'y/n', 'miduous', 'always per
sistent green', 'always s green', 'always oversing green', 'aestival', 'partly evergreen', 'vernal', 'hibernal', 'win
t', 'summ', 's and s', 'yes', 'no', 'year round', 's, s, s', 's', 's, s and s', 'sd', 'dc', 's, s', 'any time with ra
in, s', 'any time with rain', 'any time with rain, s, s', 's, s, s, s', 'any time with rain, s, s, s', 'm - m', '?',
'semimiduous', 'miduous/evergreen', 'n.d.', 'variable miduous', 'semi-miduous', 'miduousecimiduousuous', 'm-m', 'm-m
(m)', 'evergreen type @', 'miduous type @', 'evergreen type @', 'semi-evergreen', 'megaphanerophyte', 'sessile', 's
g', 'evergreen scale-like', 'evergreen broad-leaved', 'miduous broad-leaved', 'evergreen needle-leaved', 'miduous or
evergreen broad-leaved', 'miduous needle-leaved', 'miduous scale-like', 'miduous or evergreen needle-leaved', 'phyllo
dium', 'miduous or semi-evergreen broad-leaved', 'miduous or evergreen scale-like', 'other', 'cladophylls', 'evergree
n grass', 'ev', 'd_ev', '@.@', 'briefly miduous', 'nonevergreen', 'semi-evergreen/drought-miduous', 'drought-miduou
s', 'evergreen', 'semi-evergreen/evergreen', 'ed', 'db', 'aphyllous', 'exchanger', '<@', 'nd', 'x', 'broadleaved', 'ev
ergreen/miduous', 'http://www.photomazza.com/?bauhinia-integrifolia&lang=en', 'semievergreen', 'everggreen', 'http://
www.sunshine-seeds.de/indigofera-heterophylla-@p.html', 'http://plants.jstor.org/stable/@.@/al.ap.flora.ftea@', 'htt
p://plants.jstor.org/compilation/blepharispermum.zanguebaricum', 'http://tropical.theferns.info/viewtropical.php?id=b
orrichia+arborescens', 'https://en.wikipedia.org/wiki/colutea', 'http://www.efloras.org/florataxon.aspx?flora_id=@&ta
xon_id=@', 'miduous/semi-evergreen', 'https://en.wikipedia.org/wiki/phygelius_aequalis', 'miduous to semi-evergreen',
'evergreen/semievergreen', 'evergreen/semi-miduous', 'http://tropical.theferns.info/viewtropical.php?id=bauhinia+hild
ebrandtii', 'http://www.rogerstreesandshrubs.com/gallery/displayblock~bid~@~gid~~source~gallerydefault.asp', 'miduou
s', 'miduous/needlelike', 'miduous-semimiduous', 'muduou', 'semi-miduous to evergreen', 'evergreen/semi-evergreen',
'miduous/semimiduous', 'https://de.wikipedia.org/wiki/mopane', 'http://tropical.theferns.info/viewtropical.php?id=alb
izia+pedicellaris', 'semimicuous', 'semi-evergreen/miduous', 'semimiduoushttps://sites.unicentro.br/wp/manejofloresta
l/@-@/', 'evergren', 'semi-evergreen/evergren', 'semimiduous/evergreen', 'drought semi-miduous', 's miduous', 's semi
-miduous', 'perennifolio', 'caducifolio', 'bd', 'drought miduous', 'cold miduous', 'leafless', 'stem succulent', 'sgr
een', 'seasonal biomass peak']

```

```

In [50]: # View all forms.
         FH.data_trait["non_std"].value_form.unique()

```



```

Out[50]: array(['y', 'n', 'w', 'd', 'e', 'evergreen', '@', 'semi', 'y/n',
'miduous', 'always persistent green', 'always s green',
'always oversing green', 'aestival', 'partly evergreen', 'vernal',
'hibernal', 'wint', 'summ', 's and s', 'yes', 'no', 'year round',
's, s, s', 's', 's, s and s', 'sd', 'dc', 's, s',
'any time with rain, s', 'any time with rain',
'any time with rain, s, s', 's, s, s, s',
'any time with rain, s, s, s', 'm - m', '?', 'semimiduous',
'miduous/evergreen', 'n.d.', 'variable miduous', 'semi-miduous',
'miduousecimiduousuous', 'm-m', 'm-m(m)', 'evergreen type @',
'miduous type @', 'evergreen type @', 'semi-evergreen',
'megaphanerophyte', 'sessile', 'sg', 'evergreen scale-like',
'evergreen broad-leaved', 'miduous broad-leaved',
'evergreen needle-leaved', 'miduous or evergreen broad-leaved',
'miduous needle-leaved', 'miduous scale-like',
'miduous or evergreen needle-leaved', 'phyllodium',
'miduous or semi-evergreen broad-leaved',
'miduous or evergreen scale-like', 'other', 'cladophylls',
'evergreen grass', 'ev', 'd_ev', '@.@', 'briefly miduous',
'nonevergreen', 'semi-evergreen/drought-miduous',
'drought-miduous', 'evergeen', 'semi-evergreen/evergreen', 'ed',
'db', 'aphyllous', 'exchanger', '<@', 'nd', 'x', 'broadleaved',
'evergreen/miduous',
'http://www.photomazza.com/?bauhinia-integrifolia&lang=en',
'semievergreen', 'everggreen',
'http://www.sunshine-seeds.de/indigofera-heterophylla-@p.html',
'http://plants.jstor.org/stable/@.@/al.ap.flora.ftea@',
'http://plants.jstor.org/compilation/blepharispermum.zanguebaricum',
'http://tropical.theferns.info/viewtropical.php?id=borrichia+arborescens',
'https://en.wikipedia.org/wiki/colutea',
'http://www.efloras.org/florataxon.aspx?flora_id=@&taxon_id=@',
'miduous/semi-evergreen',
'https://en.wikipedia.org/wiki/phygelius_aequalis',
'miduous to semi-evergreen', 'evergreen/semievergreen',
'evergreen/semi-miduous',
'http://tropical.theferns.info/viewtropical.php?id=bauhinia+hildebrandtii',
'http://www.rogerstreesandshrubs.com/gallery/displayblock~bid~@~gid~~source~gallerydefault.asp',
'mioduous', 'miduous/needlelike', 'miduous-semimiduous', 'muduuous',
'semi-miduous to evergreen', 'evergreen/semi-evergreen',
'miduous/semimiduous', 'https://de.wikipedia.org/wiki/mopane',
'http://tropical.theferns.info/viewtropical.php?id=albizia+pedicellaris',
'semimicuous', 'semi-evergreen/miduous',

```

```
'semimiduoushttps://sites.unicentro.br/wp/manejoflorestal/@-@/',
'evergren', 'semi-evergreen/evergren', 'semimiduous/evergreen',
'drought semi-miduous', 's miduous', 's semi-miduous',
'perennifolio', 'caducifolio', 'bd', 'drought miduous',
'cold miduous', 'leafless', 'stem succulent', 'sgreen',
'seasonal biomass peak'], dtype=object)
```

```
In [51]: # View context.
FH.get_context(FH.data_trait["non_std"])
# OBSERVATION: Columns "OrigName", and "Comment" provide useful context information.
# Also, there are many values that are placeholders and should be replaced
# with their expanded meaning.
```

Out[51]:

	DataName	OrigName	OrigUnitStr	Comment
0	Leaf phenology type	Evergreen	NaN	Phenological type: evergreen: Y deciduous: N
1612	Leaf phenology type	Phenological type	NaN	D=deciduous, E=evergreen, W=wintergreen (D/E)
1647	Leaf phenology type	phenological type	NaN	NaN
7288	Leaf phenology: seasonal timing of foliage	Shoot phenology	NaN	Ordinal for green biomass peak: 1 = no evident...
14324	Leaf phenology type	evergreen/dec	NaN	evergreen/decidious; these are used for classi...
...
3814187	Leaf phenology type	Type	NaN	Phenological group
3815269	Leaf phenology type	phenology	NaN	NaN
3817563	Leaf phenology type	Phenology Class	NaN	0= evergreen coniferous, 1= deciduous broadlea...
3818500	Leaf phenology: evergreen	Evergreen	NaN	0 = no, 1 = yes
3818501	Leaf phenology: deciduous	Deciduous	NaN	0 = no, 1 = yes

79 rows × 4 columns

```
In [52]: # View OriglName + comments.  
for n, c, f in FH.data_trait["non_std"][[  
    "OriglName", "Comment", "value_form"  
]].drop_duplicates().values: print(f"{n} | {c} | {f}")
```

```

Evergreen | Phenological type: evergreen: Y deciduous: N | y
Evergreen | Phenological type: evergreen: Y deciduous: N | n
Phenological type | D=deciduous, E=evergreen, W=wintergreen (D/E) | w
Phenological type | D=deciduous, E=evergreen, W=wintergreen (D/E) | d
Phenological type | D=deciduous, E=evergreen, W=wintergreen (D/E) | e
phenological type | nan | evergreen
Shoot phenology | Ordinal for green biomass peak: 1 = no evident peak; 2 = winter, autumn, early spring peak; 3 = late spring, spring, spring-summer; 4 = late summer-early autumn; 5 = late spring-summer, summer | @
evergreen/decid | evergreen/deciduous; these are used for classifying species in some dynamic veg models; please note if a species is semi-evergreen; | e
evergreen/decid | evergreen/deciduous; these are used for classifying species in some dynamic veg models; please note if a species is semi-evergreen; | d
evergreen/decid | evergreen/deciduous; these are used for classifying species in some dynamic veg models; please note if a species is semi-evergreen; | semi
evergreen/decid | evergreen/deciduous; these are used for classifying species in some dynamic veg models; please note if a species is semi-evergreen | d
evergreen/decid | evergreen/deciduous; these are used for classifying species in some dynamic veg models; please note if a species is semi-evergreen | e
Evergreen | nan | y
Evergreen | nan | y/n
Evergreen | nan | n
decidEver | nan | evergreen
decidEver | nan | miduous
Phenology | nan | always persistent green
Phenology | nan | always s green
Phenology | nan | always oversing green
Decid/Egreen | nan | evergreen
Decid/Egreen | nan | miduous
Leaf phenology | nan | evergreen
Leaf phenology | nan | miduous
D/E | nan | evergreen
D/E | nan | miduous
evergreen/decid | E= evergreen D=deciduous | e
evergreen/decid | E= evergreen D=deciduous | d
evergreen/decid | E= evergreen D=deciduous | semi
Leaf phenology | Leaf phenology | aestival
Leaf phenology | Leaf phenology | evergreen
Leaf phenology | Leaf phenology | partly evergreen
Leaf phenology | Leaf phenology | vernal
Leaf phenology | Leaf phenology | hibernal
Seasonality of growth | (wint-winter growing; summ-summer growing) | wint
Seasonality of growth | (wint-winter growing; summ-summer growing) | summ

```

```

phenological type | nan | miduous
Leaf habit | nan | d
Leaf habit | nan | e
Active Growth Period | nan | s and s
Leaf Retention | nan | yes
Leaf Retention | nan | no
Active Growth Period | nan | year round
Active Growth Period | nan | s, s, s
Active Growth Period | nan | s
Active Growth Period | nan | s, s and s
D_E | deciduous and evergreen | d
D_E | deciduous and evergreen | e
D_E | deciduous and evergreen | sd
D/E | nan | e
D/E | nan | d
D/E | nan | dc
evergreen / deciduous | nan | evergreen
GR, main growth period | nan | s, s
GR, main growth period | nan | any time with rain, s
GR, main growth period | nan | any time with rain
GR, main growth period | nan | s
GR, main growth period | nan | s, s, s
GR, main growth period | nan | any time with rain, s, s
GR, main growth period | nan | s, s, s, s
GR, main growth period | nan | any time with rain, s, s, s
growth priod | nan | m - m
evergreen-decidious | nan | d
evergreen-decidious | nan | e
evergreen-decidious | nan | s
evergreen-decidious | nan | semi
Leaf Habit | leaf phenology | miduous
Leaf Habit | leaf phenology | ?
Leaf Habit | leaf phenology | evergreen
Leaf habit | nan | evergreen
Leaf habit | nan | miduous
Leaf habit | nan | semimiduous
Deciduous/Evergreen | based on TRY and expert knowledge, not done for grasses and sedges (n.d.) | miduous
Deciduous/Evergreen | based on TRY and expert knowledge, not done for grasses and sedges (n.d.) | evergreen
Deciduous/Evergreen | based on TRY and expert knowledge, not done for grasses and sedges (n.d.) | miduous/evergreen
Deciduous/Evergreen | based on TRY and expert knowledge, not done for grasses and sedges (n.d.) | n.d.
Leaf Phenological type | evergreen, winter deciduous, drought deciduous, semideciduous | miduous
Leaf Phenological type | evergreen, winter deciduous, drought deciduous, semideciduous | evergreen

```

Leaf Phenological type | evergreen, winter deciduous, drought deciduous, semideciduous | variable miduous
 DecEv | Deciduous D / evergreen E / semi-deciduous SD | miduous
 DecEv | Deciduous D / evergreen E / semi-deciduous SD | evergreen
 DecEv | Deciduous D / evergreen E / semi-deciduous SD | semi-miduous
 DecEv | DECIDUOUS / EVERGREEN | e
 DecEv | DECIDUOUS / EVERGREEN | d
 Shoot phenology (seasonality of maximum production of photosynthetic tissue) | 1=no clear peak, 2=winter,autumn,early spring, 3=late spring, spring, spring/summer, 4=late spring/summer, summer | @
 Leaf habit | E=evergreen, D=deciduous | d
 Leaf habit | E=evergreen, D=deciduous | e
 Dec/Egreen | Deciduous / evergreen | miduous
 Dec/Egreen | Deciduous / evergreen | evergreen
 Dec/Egreen | Deciduous / evergreen | miduousecimiduousuous
 Dec | Deciduousness as adult | miduous
 Dec | Deciduousness as adult | evergreen
 Phenological type | nan | evergreen
 Phenological type | PFT 2,4,6,8: deciduous / 1,3,5,7:evergreen | miduous
 Phenological type | PFT 2,4,6,8: deciduous / 1,3,5,7:evergreen | evergreen
 dec_ev_new | nan | e
 dec_ev_new | nan | d
 growing period | nan | m-m
 growing period | nan | m-m(m)
 Pheno | nan | evergreen
 Pheno | nan | miduous
 EB/EC/DB | nan | d
 EB/EC/DB | nan | e
 Leaf Phenology | nan | miduous
 Leaf Phenology | nan | evergreen
 Decid/Ever | nan | evergreen
 Decid/Ever | nan | miduous
 Decid/Ever | nan | semi-miduous
 Leaf persistence | 1, aestival green; 2, partial evergreen; 3, evergreen | @
 Evergreen broadleaf/Deciduous broadleaf | nan | miduous
 Evergreen broadleaf/Deciduous broadleaf | nan | evergreen
 WholePlant kType | foliage type | evergreen type @
 WholePlant kType | foliage type | miduous type @
 WholePlant kType | foliage type | evergreen type @
 Leaf longevity: 1. type | 3.22 | evergreen
 Leaf longevity: 1. type | 3.22 | aestival
 Leaf longevity: 1. type | 3.22 | vernal
 Leaf longevity: 1. type | 3.22 | hibernal
 Leaf longevity: 1. type | 3.22 | semi-evergreen

```

Leaf longevity: 1. type | 3.22 | megaphanerophyte
Leaf longevity: 1. type | 3.22 | sessile
Evergreen | nan | evergreen
Evergreen | nan | miduous
Leaf phenology: deciduous | nan | yes
Leaf phenology: semi-deciduous | nan | no
Leaf phenology: leaf exchanger | nan | no
Leaf phenology: evergreen | nan | no
Leaf phenology: deciduous | nan | no
Leaf phenology: evergreen | nan | yes
Leaf phenology: semi-deciduous | nan | yes
Phenology | E = evergreen, W = winter deciduous, D = drought deciduous | w
Phenology | E = evergreen, W = winter deciduous, D = drought deciduous | e
Phenology | E = evergreen, W = winter deciduous, D = drought deciduous | d
Phenology2 | nan | sd
Phenology2 | nan | sg
leaf habit | nan | semimiduous
leaf habit | nan | miduous
leaf habit | nan | evergreen
Leaf phenology | nan | evergreen
Leaf phenology | nan | miduous
SpeciesLeafType | nan | evergreen scale-like
SpeciesLeafType | nan | evergreen broad-leaved
SpeciesLeafType | nan | miduous broad-leaved
SpeciesLeafType | nan | evergreen needle-leaved
SpeciesLeafType | nan | miduous or evergreen broad-leaved
SpeciesLeafType | nan | miduous needle-leaved
SpeciesLeafType | nan | miduous scale-like
SpeciesLeafType | nan | miduous or evergreen needle-leaved
SpeciesLeafType | nan | phyllodium
SpeciesLeafType | nan | miduous or semi-evergreen broad-leaved
SpeciesLeafType | nan | miduous or evergreen scale-like
SpeciesLeafType | nan | other
SpeciesLeafType | nan | cladophylls
SpeciesLeafType | nan | evergreen grass
Phenology | nan | ev
Phenology | nan | d
Phenology | nan | d_ev
DecidEvgreen_woodyonly | leaf phenology. D: deciduous; E: evergreen. Data assigned to woody species only. | d
DecidEvgreen_woodyonly | leaf phenology. D: deciduous; E: evergreen. Data assigned to woody species only. | e
Deciduous | nan | yes
Evergreen | nan | no

```

```

Deciduous | nan | no
Evergreen | nan | yes
DecEv | Deciduous or evergreen species | e
DecEv | Deciduous or evergreen species | d
DecEv | Deciduous or evergreen species | n
leaf phenology (months without leaves) | trait measurements according to Cornelssen et al. 2003 | @
leaf phenology (months without leaves) | trait measurements according to Cornelssen et al. 2003 | @.@
LAL link average length | nan | @.@
Phenology | nan | miduous
Phenology | nan | briefly miduous
Phenology | nan | evergreen
leaf_habit | nan | miduous
leaf_habit | nan | evergreen
Phenology | Indicator of deciduousness of individual (evergreen or nonevergreen) | evergreen
Phenology | Indicator of deciduousness of individual (evergreen or nonevergreen) | nonevergreen
Leaf Phenology | nan | semi-evergreen/drought-miduous
Leaf Phenology | nan | drought-miduous
Leaf Phenology | nan | evergeen
Leaf Phenology | nan | semi-miduous
Leaf Phenology | nan | semi-evergreen
Leaf Phenology | nan | semi-evergreen/evergreen
EveDec | Deciduous or Evergreen | d
EveDec | Deciduous or Evergreen | e
EveDec | Deciduous or Evergreen | ed
DecEv | Deciduous or Evergreen | e
DecEv | Deciduous or Evergreen | d
DecEv | Deciduous or Evergreen | ed
DecEv | Deciduous or Evergreen | db
Leaf phenology | nan | aphyllous
Leaf phenology | nan | semi-miduous
Leaf phenology | nan | exchanger
Leaf retention time (yr) | nan | <@
Leaf retention time (yr) | nan | @
Leaf retention time (yr) | nan | nd
Leaf phenology type | Phenology | @
Leafspan | nan | miduous
Leafspan | nan | evergreen
Growth form_Leaf habit | Whether plant is deciduous or evergreen. Data are from the categorical traits data package f
rom the TRY database: Kattge, J., Bonisch, G., GUNther, A., Wright, I., Zanne, A., Wirth, C., Reich, P.B. and the TRY
Consortium (2012) TRY - Categorical Traits Dataset. Data from: TRY - a global database of plant traits. TRY File Arch
ive https://www.try-db.org/TryWeb/Data.php#3 | miduous
Growth form_Leaf habit | Whether plant is deciduous or evergreen. Data are from the categorical traits data package f

```


rom the TRY database: Kattge, J., Bonisch, G., GUnther, A., Wright, I., Zanne, A., Wirth, C., Reich, P.B. and the TRY Consortium (2012) TRY - Categorical Traits Dataset. Data from: TRY - a global database of plant traits. TRY File Archive <https://www.try-db.org/TryWeb/Data.php#3> | evergreen

Growth form_Leaf habit | Whether plant is deciduous or evergreen. Data are from the categorical traits data package from the TRY database: Kattge, J., Bonisch, G., GUnther, A., Wright, I., Zanne, A., Wirth, C., Reich, P.B. and the TRY Consortium (2012) TRY - Categorical Traits Dataset. Data from: TRY - a global database of plant traits. TRY File Archive <https://www.try-db.org/TryWeb/Data.php#3> | x

Growth form_Leaf habit | Whether plant is deciduous or evergreen. Data are from the categorical traits data package from the TRY database: Kattge, J., Bonisch, G., GUnther, A., Wright, I., Zanne, A., Wirth, C., Reich, P.B. and the TRY Consortium (2012) TRY - Categorical Traits Dataset. Data from: TRY - a global database of plant traits. TRY File Archive <https://www.try-db.org/TryWeb/Data.php#3> | miduous/evergreen

Growth form_Leaf habit | Whether plant is deciduous or evergreen. Data are from the categorical traits data package from the TRY database: Kattge, J., Bonisch, G., GUnther, A., Wright, I., Zanne, A., Wirth, C., Reich, P.B. and the TRY Consortium (2012) TRY - Categorical Traits Dataset. Data from: TRY - a global database of plant traits. TRY File Archive <https://www.try-db.org/TryWeb/Data.php#3> | broadleaved

LeafPhenology | nan | evergreen

LeafPhenology | nan | miduous

LeafPhenology | nan | semi-evergreen

LeafPhenology | nan | miduous/evergreen

LeafPhenology | nan | evergreen/miduous

LeafPhenology | nan | <http://www.photomazza.com/?bauhinia-integrifolia&lang=en>

LeafPhenology | nan | semievergreen

LeafPhenology | nan | semimiduous

LeafPhenology | nan | evergreen

LeafPhenology | nan | semi-miduous

LeafPhenology | nan | <http://www.sunshine-seeds.de/indigofera-heterophylla-@p.html>

LeafPhenology | nan | <http://plants.jstor.org/stable/@.@/al.ap.flora.ftea@>

LeafPhenology | nan | <http://plants.jstor.org/compilation/blepharispermum.zanguebaricum>

LeafPhenology | nan | <http://tropical.theferns.info/viewtropical.php?id=borrichia+arborescens>

LeafPhenology | nan | <https://en.wikipedia.org/wiki/colutea>

LeafPhenology | nan | http://www.efloras.org/florataxon.aspx?flora_id=@&taxon_id=@

LeafPhenology | nan | miduous/semi-evergreen

LeafPhenology | nan | https://en.wikipedia.org/wiki/phygelius_aequalis

LeafPhenology | nan | miduous to semi-evergreen

LeafPhenology | nan | evergreen/semievergreen

LeafPhenology | nan | evergreen/semi-miduous

LeafPhenology | nan | <http://tropical.theferns.info/viewtropical.php?id=bauhinia+hildebrandtii>

LeafPhenology | nan | <http://www.rogerstreesandshrubs.com/gallery/displayblock~bid~@~gid~source~gallerydefault.asp>

LeafPhenology | nan | miduous

LeafPhenology | nan | miduous/needlelike

LeafPhenology | nan | miduous-semimiduous

LeafPhenology | nan | miduous

```

LeafPhenology | nan | semi-miduous to evergreen
LeafPhenology | nan | evergreen/semi-evergreen
LeafPhenology | nan | miduous/semimiduous
LeafPhenology | nan | https://de.wikipedia.org/wiki/mopane
LeafPhenology | nan | http://tropical.theferns.info/viewtropical.php?id=albizia+pedicellaris
LeafPhenology | nan | semimicuous
LeafPhenology | nan | semi-evergreen/miduous
LeafPhenology | nan | semimiduoushttps://sites.unicentro.br/wp/manejoflorestal/@-@/
LeafPhenology | nan | evergren
LeafPhenology | nan | semi-evergreen/evergren
LeafPhenology | nan | semimiduous/evergreen
leaf phenology | nan | miduous
leaf phenology | nan | evergreen
Leafspan | Evergreen or Deciduous | evergreen
Leafspan | Evergreen or Deciduous | miduous
LeafPhenology | nan | drought semi-miduous
LeafPhenology | nan | s miduous
LeafPhenology | nan | s semi-miduous
DecEv | nan | evergreen
DecEv | nan | miduous
Leaf habit | native= native to site (New Zealand endemic), "invasive" = introduced, non-native to site, with spreadin
g populations at site | miduous
Leaf habit | native= native to site (New Zealand endemic), "invasive" = introduced, non-native to site, with spreadin
g populations at site | semi-miduous
Leaf habit | native= native to site (New Zealand endemic), "invasive" = introduced, non-native to site, with spreadin
g populations at site | evergreen
Fenologia | nan | perennifolio
Fenologia | nan | caducifolio
Type | Phenological group | dc
Type | Phenological group | ev
Type | Phenological group | bd
phenology | nan | evergreen
phenology | nan | drought miduous
phenology | nan | cold miduous
phenology | nan | leafless
phenology | nan | stem succulent
Phenology Class | 0= evergreen coniferous, 1= deciduous broadleaved, 2= deciduous coniferous. | evergreen
Phenology Class | 0= evergreen coniferous, 1= deciduous broadleaved, 2= deciduous coniferous. | sgreen
phenology | nan | seasonal biomass peak
Evergreen | 0 = no, 1 = yes | @
Deciduous | 0 = no, 1 = yes | @

```

Evergreen | 0 = no, 1 = yes | ?
Deciduous | 0 = no, 1 = yes | ?

```
In [53]: # View values with specific form.  
FH.get_unique_matches(  
    data=FH.data_trait["non_std"],  
    match_col='OriglName',  
    to_match=["Leaf phenology type"],  
    keep=["OrigValueStr"])
```

Out[53]:

OrigValueStr	
1166721	147
1166729	144
1166745	117
1166769	91
1166777	126
1166785	140
1166825	112
1166833	84
1166841	49
1166849	105
1166857	68
1166865	56
1166873	103
1166889	88
1166897	175
1166905	12
1166921	35
1166929	172
1166945	79
1166953	63
1166961	110
1166969	51

OrigValueStr	
1167001	86
1167009	168
1167041	37
1167049	61

```
In [54]: # View data with specific OrigValueStr.
spec_form = FH.data_trait["non_std"][(
    (FH.data_trait["non_std"].value_form == "n") &
    (FH.data_trait["non_std"].OriglName == "DecEv")
)]
num_species = 0
for n in spec_form.AccSpeciesName.unique():
    print(n)
    num_species += 1
print(f"Total no. of species = {num_species}.")
```

Coreopsis suaveolens
Lythrum salicaria
Asclepias syriaca
Euthamia graminifolia
Solidago virgaurea
Dennstaedtia punctilobula
Deparia acrostichoides
Dryopteris goeldiana
Dryopteris intermedia
Dryopteris marginalis
Onoclea sensibilis
Polystichum acrostichoides
Osmundastrum cinnamomeum
Osmunda claytoniana
Osmunda regalis
Polypodium virginianum
Adiantum pedatum
Phegopteris hexagonoptera
Parathelypteris noveboracensis
Lespedeza davurica
Artemisia subdigitata
Artemisia gmelinii
Potentilla acaulis
Pulsatilla chinensis
Periploca sepium
Agropyron cristatum
Imperata cylindrica
Carex lanceolata
Glycyrrhiza uralensis
Euphorbia humifusa
Bothriochloa ischaemum
Astragalus laxmannii
Medicago sativa
Trigonella alba
Themeda triandra
Lespedeza cyrtobotrya
Lespedeza bicolor
Onobrychis viciifolia
Medicago lupulina
Potentilla multicaulis
Freycinetia arborea
Epipactis helleborine

Pilosella floribunda
Lupinus angustifolius
Bistorta plumosa
Calamagrostis canadensis
Carex aquatilis
Carex bigelowii
Eriophorum angustifolium
Eriophorum vaginatum
Hierochloa alpina
Petasites frigidus
Bistorta officinalis
Rubus chamaemorus
Colobanthus quitensis
Deschampsia antarctica
Achillea erba
Doronicum clusii
Erigeron uniflorus
Geum montanum
Geum reptans
Leucanthemopsis alpina
Ligusticum mutellina
Luzula alpinopilosa
Oxyria digyna
Pedicularis asplenifolia
Bistorta vivipara
Potentilla crantzii
Primula glutinosa
Oxygraphis glacialis
Taraxacum alpinum
Achillea millefolium
Carum carvi
Leucanthemum vulgare
Geum rivale
Potentilla anserina
Primula elatior
Ranunculus acris
Taraxacum campylodes
Trifolium repens
Desmodium sp
Brachiaria brizantha
Adenophyllum porophylloides
Krascheninnikovia ceratoides

Salvia dorrii
Festuca arundinaria
Elymus repens
Anemone cylindrica
Bromus inermis
Lupinus perennis
Poa pratensis
Solidago rigida
Oenothera perennis
Podophyllum peltatum
Panicum acuminatum
Aquilegia micrantha
Cirsium rydbergii
Solidago velutina
Nardus stricta
Ehrharta stipoides
Isachne distichophylla
Crocosmia pottsii
Astelia menziesiana
Conyza canadensis
Pseudognaphalium sandwicense
Hypochaeris radicata
Argyroxiphium kauense
Plantago lanceolata
Plantago hawaiiensis
Melinis repens
Heteropogon contortus
Holcus lanatus
Deschampsia nubigena
Paspalum urvillei
Eragrostis variabilis
Brachypodium phoenicoides
Lotus pentaphyllus
Aegilops geniculata
Helictotrichon bromoides
Brachypodium distachyon
Bromus erectus
Bromus lanceolatus
Bupleurum rigidum
Carex halleriana
Catananche caerulea
Dactylis glomerata

Daucus carota
Lotus hirsutus
Echinops ritro
Eryngium campestre
Medicago minima
Phleum pratense
Sanguisorba minor
Viola alba
Xeranthemum inapertum
Drimys maritima
Helictotrichon crassifolium
Crepis triasii
Delphinium pictum
Dianthus rupicola
Urtica atrovirens
Viola jaubertiana
Viola odorata
Viola balearica
Amaranthus blitoides
Amaranthus retroflexus
Beta vulgaris
Chenopodium album
Cichorium intybus
Convolvulus arvensis
Datura stramonium
Malva multiflora
Paeonia cambessedesii
Verbascum sinuatum
Leymus secalinus
Anemone cathayensis
Astragalus porphyrocalyx
Bistorta macrophylla
Przewalskia tangutica
Saussurea hieracioides
Kobresia royleana
Festuca rubra
Cremanthodium discoideum
Kobresia kansuensis
Potentilla nivea
Rhodiola crenulata
Saussurea medusa
Carex capillifolia

Gentiana lawrencei
Koeleria pyramidata
Artemisia sieversiana
Delphinium caeruleum
Pedicularis alaschanica
Gnaphalium affine
Dasiphora davurica
Eremogone kansuensis
Kobresia pusilla
Lagotis glauca
Pedicularis integrifolia
Primula tangutica
Rumex acetosa
Lonicera hispida
Kobresia sp
Marmoritis pharica
Saussurea glanduligera
Drosera rotundifolia
Pinguicula vulgaris
Andropogon gerardii
Lespedeza capitata
Dalea villosa
Schizachyrium scoparium
Sorghastrum nutans
Ambrosia psilostachya
Artemisia ludoviciana
Asclepias viridiflora
Asclepias viridis
Symphyotrichum oblongifolium
Callirhoe involucrata
Panicum oligosanthos
Echinacea angustifolia
Brickellia eupatorioides
Panicum virgatum
Physalis pumila
Pediomelum tenuiflorum
Ruellia humilis
Salvia azurea
Solidago missouriensis
Sporobolus heterolepis
Vernonia baldwinii
Abronia villosa

Eremalche rotundifolia
Geraea canescens
Chylismia claviformis
Chylismia brevipes
Abutilon theophrasti
Ambrosia trifida
Persicaria pensylvanica
Goeppertia inocephala
Pleiostachya pruinosa
Sporobolus wrightii
Geum rosii
Bistorta bistortoides
Caltha leptosepala
Ageratina liebmanna
Nemophila aphylla
Helianthus microcephalus
Veratrum parviflorum
Atriplex canescens
Pterocaulon pycnostachyum
Baptisia bracteata
Caulophyllum thalictroides
Thelypodium laciniatum
Echinacea purpurea
Erythronium americanum
Sanguinaria canadensis
Sarracenia purpurea
Silphium integrifolium
Silphium terebinthinaceum
Trillium grandiflorum
Oxalis oregana
Desmodium tortuosum
Elymus canadensis
Lactuca ludoviciana
Koenigia campanulata
Anemone rupicola
Bistorta amplexicaulis
Cremanthodium pinnatifidum
Lagotis kunawurensis
Picrorhiza scrophulariiflora
Argentina contigua
Argentina peduncularis
Rheum acuminatum

Rheum nobile
Saussurea gossypiphora
Swertia acaulis
Alternanthera crucis
Agastache foeniculum
Agrostis scabra
Ambrosia artemisiifolia
Asclepias tuberosa
Symphyotrichum oolentangiense
Symphyotrichum ericoides
Symphyotrichum novae-angliae
Astragalus canadensis
Bouteloua curtipendula
Calamovilfa longifolia
Coreopsis palmata
Desmodium canadense
Liatris aspera
Panicum capillare
Penstemon grandiflorus
Dalea purpurea
Drymocallis arguta
Echinacea serotina
Solidago nemoralis
Stipa spartea
Lyginia barbata
Patersonia occidentalis
Anigozanthos humilis
Drosera sp
Gladiolus caryophyllaceus
Pterostylis sp
Stylidium sp
Triodia scabra
Mikania cordata
Mikania micrantha
Verbena officinalis
Actinidia chinensis
Phlomoides macrophylla
Brachybotrys paridiformis
Carex filipes
Carex arnellii
Adenophora elata
Isodon amethystoides

Potentilla fragarioides
Veratrum nigrum
Asplenium cristatum
Syneilesis aconitifolia
IMPATIENS NOLI-TANGERE
Cardamine macrophylla
Vicia amoena
Hemerocallis citrina
Thalictrum aquilegiifolium
Dioscorea nipponica
Doellingeria scabra
Circaea canadensis
Geum aleppicum
Arabidopsis halleri
Parasenecio hastatus
Carex siderosticta
Maianthemum bifolium
Ligularia lapathifolia
Aconitum carmichaelii
Veronica spuria
Filipendula palmata
Euphorbia helioscopia
Vicia unijuga
Aster tataricus
Viola variegata
Actaea cimicifuga
Polygonatum odoratum
Veronica longipetiolata
Heteropappus tataricus
Bassia scoparia
Afroaster hispidus
Aster alpinus
Leymus chinensis
Carex enervis
Taraxacum ussuriense
Arundinella hirta
Sanguisorba tenuifolia_x
Limonium bicolor
Artemisia desertorum
Astragalus compressus
Equisetum hyemale
Microtoena omeiensis

Thalictrum scabrifolium
Carex duriuscula
Iris lactea
Phragmites australis
Allium schoenoprasum
Saussurea amara
Hordeum brevisubulatum
Pennisetum flaccidum
Chenopodium acuminatum
Kali collinum
Puccinellia distans
Cleistogenes squarrosa
Corispermum chinganicum
Suaeda glauca
Rhaponticum centauroides
Halerpestes salsuginosa
Stipa gigantea
Arthraxon hispidus
Saussurea macrota
Xanthium strumarium subsp. sibiricum
Chrysanthemum indicum
Atriplex patens
Plantago major
Trigonella officinalis
Enneapogon desvauxii
Sibbaldianthe bifurca
Potentilla chrysantha
Artemisia mongolica
Artemisia scoparia
Phlomis umbrosa
Calamagrostis epigejos
Bupleurum chinense
Astragalus melilotoides
Oxytropis cachemiriana
Polygala tenuifolia
Thalictrum squarrosum
Veronica incana
Artemisia sibirica
Iris ruthenica
Carex idzuroei
Thermopsis lanceolata
Saposhnikovia divaricata

Chloris virgata
Glaux maritima
Sporobolus aculeatus
Potentilla chinensis
Potentilla freyniana
Equisetum arvense
Koenigia alpina
Persicaria lapathifolia
Atriplex oblongifolia
Oxybasis glauca
Tribulus terrestris
Iris gracilipes
Tragus berteronianus
Artemisia annua
Astragalus scaberrimus
Sanguisorba officinalis
Taraxacum mongolicum
Astragalus strictus
Corispermum mongolicum
Carex korshinskyi
Plantago depressa
Allium ramosum
Setaria viridis
Grubovia dasyphylla
Ligularia mongolica
Lomelosia comosa
Nepeta tenuifolia
Psammochloa villosa
Allium tenuissimum
Astragalus kuschakewiczii
Kalimeris altaica
Stipa mongolorum
Echinops latifolius
Bromus pumpellianus
Tournefortia sibirica
Asparagus dauricus
Echinops gmelinii
Agriophyllum squarrosum
Crepidiastrum denticulatum
Carex callitrichos
Orostachys fimbriata
Total no. of species = 419.

```
In [55]: # ChatGPT was used to identify which species in the list above
# are "deciduous", "evergreen" or "semi-evergreen" because 419
# values are too many to ignore if it can be helped. It was
# reported that most of these species are deciduous except
# ["Osmundastrum cinnamomeum", "Phlegopteris hexagonoptera",
# "Polystichum acrostichoides"] which are "evergreen" and
# ["Dryopteris intermedia", "Onoclea sensibilis", "Carex lanceolata"]
# which are semi-evergreen.
```

```
In [56]: FH.data_trait["non_std"].OriglName.unique()
```

```
Out[56]: array(['Evergreen ', 'Phenological type', 'phenological type',
'Shoot phenology', 'evergreen/decid', 'Evergreen', 'decidEver',
'Phenology', 'Decid/Egreen', 'Leaf phenology', 'D/E',
'Seasonality of growth ', 'Leaf habit', 'Active Growth Period',
'Leaf Retention', 'D_E', 'evergreen / deciduous',
'GR, main growth period', 'growth priod', 'evergreen-deciduous',
'Leaf Habit', 'Deciduous/Evergreen', 'Leaf Phenological type',
'DecEv',
'Shoot phenology (seasonality of maximum production of photosynthetic tissue) ',
'Dec/Egreen', 'Dec', 'dec_ev_new', 'growing period', 'Pheno',
'EB/EC/DB', 'Leaf Phenology', 'Decid/Ever', 'Leaf persistence',
'Evergreen broadleaf/Deciduous broadleaf', 'WholePlant kType',
'Leaf longevity: 1. type', 'Leaf phenology: deciduous',
'Leaf phenology: semi-deciduous', 'Leaf phenology: leaf exchanger',
'Leaf phenology: evergreen', 'Phenology2', 'leaf habit',
'Leaf\phenology', 'SpeciesLeafType', 'DecidEvgreen_woodyonly',
'Deciduous', 'leaf phenology (months without leaves)',
'LAL link average length', 'leaf_habit', 'EveDec',
'Leaf retention time (yr)', 'Leaf phenology type', 'Leafspan',
'Growth form_Leaf habit', 'LeafPhenology', 'leaf phenology',
'Fenologia', 'Type', 'phenology', 'Phenology Class'], dtype=object)
```

```
In [57]: # View data with specific OriglNames.
spec_form = FH.get_unique_matches(
    data=FH.data_trait["non_std"],
    match_col='OriglName',
    to_match=["Leaf phenology type", "Shoot phenology"],
    keep=["AccSpeciesName", "OrigValueStr", "Comment"])
print("No. of rows =", len(spec_form))
print("No. of unique species =", len(
```



```
spec_form.AccSpeciesName.unique()))  
for name in spec_form.AccSpeciesName.unique(): # View species  
    print(name)
```

No. of rows = 435
No. of unique species = 404
Acer monspessulanum
Achillea millefolium
Adonis aestivalis
Aegilops kotschy
Agropyron cristatum
Agrostis canina
Alchemilla erythropoda
Alchemilla fluminea
Alchemilla sericata
Alliaria petiolata
Allium rubellum
Alopecurus apiatus
Alopecurus textilis
Malva hirsuta
Alyssum desertorum
Prunus feniziana
Androsace villosa
Cotoneaster integerrimus
Arrhenatherum elatius
Artemisia absinthium
Artemisia armeniaca
Asperula glomerata
Astragalus alyssoides
Astragalus microcephalus
Berberis integerrima
Bothriochloa ischaemum
Briza media
Bromus japonicus
Bromus ramosus
Bromus scoparius
Bromus tectorum
Buglossoides arvensis
Bupleurum falcatum
Bupleurum gerardii
Callipeltis cucullaris
Campanula glomerata
Carduus seminudus
Carex divulsa
Carex halleriana
Carex nigra

Carpinus betulus
Catabrosa aquatica
Orlaya platycarpus
Celtis australis
Centaurea aziziana
Centaurea solstitialis
Cerastium glomeratum
Cerastium holosteoides
Asplenium ceterach
Clinopodium vulgare
Colutea cilicica
Convolvulus arvensis
Convolvulus lineatus
Cornus sanguinea
Cornus mas
Coronilla scorpioides
Securigera varia
Cotinus coggygria
Cotoneaster nummularioides
Cousinia belangeri
Crepis foetida
Dactylis glomerata
Delphinium tuberosum
Deschampsia caespitosa
Dianthus cretaceus
Dianthus orientalis
Erodium cicutarium
Eryngium caucasicum
Festuca ovina
Festuca rupicola
Filipendula vulgaris
Fraxinus excelsior subsp. coriariifolia
Galium aparine
Galium tenuissimum
Galium verum
Geranium lucidum
Geranium pratense
Geranium rotundifolium
Geum rivale
Geum urbanum
Helianthemum nummularium
Helianthemum salicifolium

Helichrysum plicatum
Helictotrichon pubescens
Herniaria hirsuta
Hesperis hyrcana
Pilosella officinarum
Hippocrepis unisiliquosa
Hordeum brevisubulatum
Hypericum hirsutum
Hypericum linarioides
Hypericum perforatum
Pentanema conyzae
Jasminum fruticans
Juncus articulatus
Juniperus communis
Bassia prostrata
Koeleria pyramidata
Lamium amplexicaule
Lappula barbata
Lapsana communis
Lathyrus pratensis
Ligustrum vulgare
Linum corymbulosum
Linum nodiflorum
Lolium rigidum
Lonicera caucasica
Lonicera iberica
Lotus corniculatus
Luzula multiflora
Malva neglecta
Marrubium astracanicum
Medicago minima
Melica persica
Mentha longifolia
Myosotis sylvatica
Neslia paniculata subsp. thracica
Onobrychis bungei
Origanum vulgare
Achnatherum virescens
Paliurus spina-christi
Phleum paniculatum
Phragmites australis
Pimpinella tragium

Pistacia atlantica
Plantago afra
Plantago lanceolata
Poa nemoralis
Poa pratensis
Potentilla aucheriana
Potentilla inclinata
Potentilla szovitsii
Primula veris
Prunella vulgaris
Prunus spinosa
Punica granatum
Quercus macranthera
Quercus petraea
Rhamnus cathartica
Rhamnus pallasii
Rhus coriaria
Rosa canina
Rubus sp
Rumex acetosa
Rumex tuberosus
Salvia sclarea
Salvia verticillata
Salvia viridis
Sanguisorba minor
Satureja macrantha
Scorzonera laciniata
Scrophularia umbrosa
Scrophularia variegata
Scutellaria pinnatifida
Phedimus spurius
Klasea quinquefolia
Stachys montana
Silene italica
Silene persica
Stachys byzantina
Stellaria media
Stipa arabica
Stipa sp
Tanacetum parthenium
Teucrium chamaedrys
Thalictrum minus

Thlaspi arvense
Thymus kotschyanus
Torilis leptophylla
Brachypodium distachyon
Tribulus terrestris
Trifolium arvense
Trifolium hybridum
Trifolium pratense
Trifolium repens
Trigonella calliceras
Trisetum flavescens
Ulmus minor
Urtica dioica
VERONICA ANAGALLIS-AQUATICA
Veronica polita
Viburnum lantana
Vicia dumetorum
Vitis vinifera
Ziziphora capitata
Ziziphora clinopodioides
Agrostis stolonifera
Anacyclus clavatus
Anagallis arvensis
Helosciadium nodiflorum
Artemisia herba-alba
Aster squamatus
Atractylis humilis
Atriplex halimus
Atriplex patula
Atriplex prostrata
Avena sterilis
Helictotrichon bromoides
Bassia scoparia
Brachypodium phoenicoides
Brachypodium retusum
Bupleurum fruticescens
Calystegia sepium
Lepidium draba
Centaurea aspera
Chenopodium album
Chenopodium vulvaria
Cichorium intybus

Cirsium arvense
Clematis vitalba
Cynodon dactylon
Lotus dorycnium
Echinochloa crus-galli
Echinops ritro
Elymus pungens
Epilobium hirsutum
Eryngium campestre
Euphorbia serrata
Festuca arundinacea
Filago pyramidata
Foeniculum vulgare
Fumana ericoides
Genista scorpius
Gypsophila struthium
Hedera helix
Helianthemum marifolium
Helianthemum pilosum
Helianthemum squamatum
Helianthemum syriacum
Helichrysum stoechas
Hypochaeris radicata
Juncus maritimus
Juniperus oxycedrus
Kickxia spuria
Lathyrus cicera
Launaea pumila
Rhaponticum coniferum
Linum narbonense
Linum strictum
Linum suffruticosum
Lotus tenuis
Lygeum spartum
Lythrum salicaria
Medicago lupulina
Medicago polymorpha
Medicago sativa
Ononis pusilla
Ononis spinosa
Paspalum distichum
Phlomis lychnitis

Helminthotheca echiioides
Pinus halepensis
Plantago albicans
Plantago coronopus
Plantago major
Poa annua
Polygonum aviculare
Persicaria maculosa
Populus nigra
Quercus coccifera
Quercus faginea
Salvia rosmarinus
Rubus ulmifolius
Kali turgidum
Samolus valerandi
Santolina chamaecyparissus
Bolboschoenus maritimus
Senecio vulgaris
Sonchus maritimus
Stipa offneri
Suaeda vera
Thymus vulgaris
Trifolium fragiferum
Typha domingensis
Xanthium orientale subsp. italicum
Acer campestre
Acer pseudoplatanus
Agrostis capillaris
Allium vineale
Alnus glutinosa
Anemone nemorosa
Bromus sterilis
Anthoxanthum odoratum
Anthriscus sylvestris
Aphanes arvensis
Arenaria serpyllifolia
Artemisia vulgaris
Atriplex littoralis
Betula pendula
Brachypodium pinnatum
Brachypodium sylvaticum
Bromus hordeaceus

Buddleja davidii
Callitriche stagnalis
Calluna vulgaris
Campanula rotundifolia
CAPSELLA BURSA-PASTORIS
Cardamine pratensis
Centaurea nigra
Cerastium fontanum
Epilobium angustifolium
Cirsium vulgare
Cochlearia danica
Corylus avellana
Crataegus monogyna
Crepis capillaris
Deschampsia cespitosa
Deschampsia flexuosa
Dryopteris dilatata
Elodea canadensis
Elymus repens
Epilobium montanum
Fagus sylvatica
Fallopia convolvulus
Lolium giganteum
Festuca rubra
Fraxinus excelsior
Galium palustre
Galium saxatile
Geranium robertianum
Glyceria fluitans
Heracleum sphondylium
Holcus lanatus
Holcus mollis
HYACINTHOIDES NON-SCRIPTA
Ilex aquifolium
Juncus effusus
Koeleria macrantha
Lamium galeobdolon
Larix decidua
Lemna minor
Leontodon hispidus
Linum catharticum
Lolium perenne

Lonicera periclymenum
Luzula campestris
Matricaria discoidea
Mentha aquatica
Mercurialis perennis
Myosotis scorpioides
Nardus stricta
Oxalis acetosella
Picea sitchensis
Pinus sylvestris
Poa trivialis
Potamogeton natans
Potentilla erecta
Pteridium aquilinum
Puccinellia distans
Ranunculus acris
Ranunculus bulbosus
Ficaria verna
Ranunculus penicillatus
Ranunculus repens
Rhododendron ponticum
Nasturtium microphyllum
Rubus fruticosus
Rumex acetosella
Rumex obtusifolius
Salix caprea
Sambucus nigra
Sedum acre
Hylotelephium telephium
Jacobaea vulgaris
Senecio squalidus
Silene dioica
Solanum dulcamara
Sonchus oleraceus
Sorbus aucuparia
Sparganium erectum
Spergula arvensis
Spergularia salina
Stellaria alsine
Taraxacum campylodes
Taxus baccata
Teucrium scorodonia

Tripleurospermum inodorum
Tussilago farfara
Ulex europaeus
Ulmus glabra
Umbilicus rupestris
Vaccinium myrtillus
Veronica chamaedrys
Veronica persica
Viburnum opulus
Viola riviniana
Calamagrostis epigejos
Alopecurus pratensis
Aegopodium podagraria
Arctium lappa
Digitalis purpurea
Digitaria sanguinalis
Pulicaria dysenterica
Lamium purpureum
Cirsium acaule
APERA SPICA-VENTI
Alopecurus geniculatus
Centaureum erythraea
Phalaris arundinacea
Filipendula ulmaria
Anthyllis vulneraria
Succisa pratensis
Molinia caerulea
Scirpus sylvaticus

2.4. Define & Apply Transformation Functions

```
In [58]: ai_evergreen = [  
    'Rhododendron kiusianum',  
    'Pinus tabulaeformis',  
    'Alnus glutinosa',  
    'Pinus armandii',  
    'Elymus pungens',  
    'Agrostis canina',  
    'Hedera helix',  
    'Rhododendron smirnowii',  
    'Juniperus communis',
```

```
'Hedera nepalensis',  
'Rhododendron yedoense',  
'Rhododendron catawbiense',  
'Rhododendron tashiroi',  
'Acer monspessulanum',  
'Ilex pernyi',  
'Crataegus azarolusRhododendron simsii',  
'Pinus sylvestris',  
'Rhododendron arboreum',  
'Tsuga heterophylla',  
'Rhododendron calendulaceum',  
'Pinus sylvestris var. mongholica',  
'Picea sitchensis',  
'Rhododendron tomentosum',  
'Rhododendron ponticum',  
'Rhododendron keiskei',  
'Rhododendron triflorum',  
'Vaccinium myrtillus',  
'Larix decidua',  
'Ilex aquifolium',  
'Pinus halepensis',  
'Vaccinium vitis-idaea',  
'Rhododendron brachycarpum',  
'Rhododendron hyperythrum',  
'Stauntonia angustifolia',  
'Rhododendron lutescens',  
'Rhododendron indicum',  
'Rhododendron maximum',  
'Rhododendron lanigerum',  
'Taxus baccata',  
'Brachypodium sylvaticum',  
'Tsuga chinensis',  
'Rhododendron makinoi',  
'Betula pendula',  
'Ulex europaeus',  
'Picea glauca'  
]  
ai_semi = [  
    'Celastrus orbiculatus',  
    'Equisetum arvense',  
    'Quercus aliena var. acutiserrata',  
    'Betula platyphylla',
```

```
'Sorbaria sorbifolia',  
'Rhododendron simsii',  
'Smilax discotis',  
'Juniperus oxycedrus',  
'Solanum dulcamara',  
'Styrax hemsleyanus',  
'Ilex pernyi',  
'Rosa acicularis',  
'Rhododendron arboreum',  
'Elaeagnus umbellata',  
'Calluna vulgaris',  
'Pinus sylvestris var. mongholica',  
'Alnus hirsuta',  
'Alopecurus textilis',  
'Ziziphora clinopodioides',  
'Rhododendron tomentosum',  
'Rhododendron ponticum',  
'Cornus mas',  
'Agropyron cristatum',  
'Paederia foetida',  
'Litsea pungens',  
'Vaccinium vitis-idaea',  
'Eleutherococcus septemlobus',  
'Alopecurus apiatus',  
'Ledum palustre subsp. groenlandicum',  
'Larix gmelinii',  
'Betula fruticosa',  
'Rubus fruticosus',  
'Dasiphora fruticosa',  
'Smilax stans',  
'Vicia sepium'  
]  
ai_deciduous = [  
    'Neslia paniculata subsp. thracica',  
    'Rhododendron austrinum',  
    'Hedera helix',  
    'Helianthemum pilosum',  
    'Potentilla aucheriana',  
    'Fraxinus excelsior subsp. coriariifolia',  
    'Bupleurum fruticescens',  
    'Achillea millefolium',  
    'Calystegia sepium',
```

```
'Callipeltis cucullaris',  
'Lotus dorycnium',  
'Securigera varia',  
'Allium rubellum',  
'Eryngium caucasicum',  
'Onobrychis bungei',  
'Populus davidiana',  
'Ulmus minor',  
'Trifolium arvense',  
'Lygeum spartum',  
'Alopecurus apiatus',  
'Pentanema conyzae',  
'Helichrysum plicatum',  
'Marrubium astracanicum',  
'Silene italica',  
'Helictotrichon pubescens',  
'Geranium lucidum',  
'Betula platyphylla',  
'Quercus faginea',  
'Euonymus alatus',  
'Cousinia belangeri',  
'Achnatherum virescens',  
'Rhododendron molle',  
'Helminthotheca echioides',  
'Salvia verticillata',  
'Phleum paniculatum',  
'Mentha longifolia',  
'Linum strictum',  
'Scrophularia variegata',  
'Cornus kousa',  
'Lonicera caucasica',  
'Celtis australis',  
'Alliaria petiolata',  
'Hypericum perforatum',  
'Malva hirsuta',  
'Carpinus tschonoskii',  
'Bromus scoparius',  
'Atractylis humilis',  
'Sorbaria sorbifolia',  
'Juniperus oxycedrus',  
'Plantago lanceolata',  
'Persicaria maculosa',
```

```
'Centaurea aziziana',  
'Bupleurum falcatum',  
'Cichorium intybus',  
'Ziziphora capitata',  
'Alopecurus textile',  
'Bothriochloa ischaemum',  
'Centaurea aspera',  
'Salvia rosmarinus',  
'Phedimus spurius',  
'Acer oliverianum',  
'Lythrum salicaria',  
'Plantago coronopus',  
'Trigonella calliceras',  
'Rhododendron prinophyllum',  
'Lotus tenuis',  
'Helianthemum syriacum',  
'Launaea pumila',  
'Lathyrus pratensis',  
'Gypsophila struthium',  
'Lamium amplexicaule',  
'Salvia sclarea',  
'Geranium pratense',  
'Ononis pusilla',  
'Juncus articulatus',  
'Carex nigra',  
'Alyssum desertorum',  
'Trifolium repens',  
'Erodium cicutarium',  
'Helianthemum salicifolium',  
'Berberis integerrima',  
'Alchemilla fluminea',  
'Hypericum linarioides',  
'Quercus aliena var. acutiserrata',  
'Alnus glutinosa',  
'Cotinus coggygria',  
'Stachys byzantina',  
'Cirsium arvense',  
'Trifolium pratense',  
'Larix decidua',  
'Linum nodiflorum',  
'Lolium rigidum',  
'Clematis vitalba',
```

```
'Klasea quinquefolia',  
'Tanacetum parthenium',  
'Lonicera periclymenum',  
'Koeleria pyramidata',  
'Galium verum',  
'Acer campestre',  
'Foeniculum vulgare',  
'Fumana ericoides',  
'Viburnum lantana',  
'Stellaria media',  
'Ribes manshuricum',  
'Alchemilla sericata',  
'Lathyrus cicera',  
'Geranium rotundifolium',  
'Rumex tuberosus',  
'Bromus japonicus',  
'Fraxinus excelsior',  
'Vicia dumetorum',  
'Campanula glomerata',  
'Salvia viridis',  
'Brachypodium phoenicoides',  
'Toxicodendron vernicifluum',  
'Rhododendron degronianum',  
'Eryngium campestre',  
'Coronilla scorpioides',  
'Aster squamatus',  
'Sorbus aucuparia',  
'Pilosella officinarum',  
'Deschampsia caespitosa',  
'Agrostis stolonifera',  
'Geum urbanum',  
'Quercus macranthera',  
'Colutea cilicica',  
'Pinus halepensis',  
'Phragmites australis',  
'Rumex acetosa',  
'Bromus tectorum',  
'Acer pseudoplatanus',  
'Vicia sativa',  
'Populus nigra',  
'Linum narbonense',  
'Helosciadium nodiflorum',
```


'Delphinium tuberosum',
'Anagallis arvensis',
'Hesperis hyrcana',
'Vaccinium uliginosum',
'Rhododendron calendulaceum',
'Atriplex patula',
'Helictotrichon bromoides',
'Echinops ritro',
'Scorzonera laciniata',
'Atriplex prostrata',
'Sanguisorba minor',
'Helichrysum stoechas',
'Crepis foetida',
'Adonis aestivalis',
'Chenopodium vulvaria',
'Linum corymbulosum',
'Teucrium chamaedrys',
'Thalictrum minus',
'Festuca arundinacea',
'Acer monspessulanum',
'Juncus maritimus',
'Polygonum aviculare',
'Quercus petraea',
'Carpinus betulus',
'Hypochaeris radicata',
'Rhododendron atlanticum',
'Stipa sp',
'Silene persica',
'Artemisia absinthium',
'Cornus mas',
'Bassia scoparia',
'Alchemilla erythropoda',
'Thlaspi arvense',
'Bupleurum gerardii',
'Paliurus spina-christi',
'Lotus corniculatus',
'Helianthemum squamatum',
'Crataegus monogyna',
'Arrhenatherum elatius',
'Styrax hemsleyanus',
'Centaurea solstitialis',
'Dactylis glomerata',

```
'Rosa acicularis',  
'Poa nemoralis',  
'Cota triumfettii',  
'Hypericum hirsutum',  
'Poa annua',  
'Lepidium draba',  
'Agropyron cristatum',  
'Briza media',  
'Cornus sanguinea',  
'Prunus spinosa',  
'Clinopodium vulgare',  
'Buglossoides arvensis',  
'Cornus controversa',  
'Salix caprea',  
'Castanea sativa',  
'Ligustrum vulgare',  
'Larix gmelinii',  
'Cotoneaster nummularioides',  
'Luzula multiflora',  
'Plantago major',  
'Lespedeza buergeri',  
'Astragalus alyssoides',  
'Helianthemum nummularium',  
'Elymus pungens',  
'Melica persica',  
'Agrostis canina',  
'Myosotis sylvatica',  
'Rhododendron yedoense',  
'Avena sterilis',  
'Sambucus nigra',  
'Tribulus terrestris',  
'Scrophularia umbrosa',  
'Carduus seminudus',  
'Trifolium hybridum',  
'Ononis spinosa',  
'Artemisia herba-alba',  
'Alnus hirsuta',  
'Medicago sativa',  
'Androsace villosa',  
'Euphorbia serrata',  
'Poa pratensis',  
'Rhododendron serpyllifolium',
```

```
'Rhododendron arborescens',  
'Satureja hortensis',  
'Medicago lupulina',  
'Linum suffruticosum',  
'Rubus sp',  
'Rhus coriaria',  
'Primula veris',  
'Fagus sylvatica',  
'Brachypodium distachyon',  
'Herniaria hirsuta',  
'Dianthus cretaceus',  
'Origanum vulgare',  
'Rhododendron yunnanense',  
'Pistacia atlantica',  
'Stachys montana',  
'Cynodon dactylon',  
'Medicago minima',  
'Genista scorpius',  
'Festuca ovina',  
'Prunus fenzliana',  
'Filipendula vulgaris',  
'Rosa canina',  
'Kickxia spuria',  
'Epilobium hirsutum',  
'Lapsana communis',  
'Pimpinella tragium',  
'Orlaya platycarpus',  
'Rhododendron catawbiense',  
'Prunella vulgaris',  
'Convolvulus lineatus',  
'Phlomis lychnitis',  
'Galium aparine',  
'Galium tenuissimum',  
'Elaeagnus umbellata',  
'Schisandra sphenanthera',  
'Cerastium holosteoides',  
'Lappula barbata',  
'Cerastium glomeratum',  
'Malva neglecta',  
'Juglans mandshurica',  
'Neillia sinensis',  
'Hordeum brevisubulatum',
```

```
'Betula pendula',  
'Atriplex halimus',  
'Jasminum fruticans',  
'Lonicera iberica',  
'Trisetum flavescens',  
'Vitis vinifera',  
'Chenopodium album',  
'Ziziphora clinopodioides',  
'Carex halleriana',  
'Artemisia armeniaca',  
'Anacyclus clavatus',  
'Viburnum opulus',  
'Convolvulus arvensis',  
'Carex divulsa',  
'Catabrosa aquatica',  
'Dianthus orientalis',  
'Rubus ulmifolius',  
'Veronica polita',  
'Ulex europaeus',  
'Asplenium ceterach',  
'Stipa arabica',  
'Plantago albicans',  
'Bromus ramosus',  
'Ulmus glabra',  
'Asperula glomerata',  
'Thymus kotschyanus',  
'Paspalum distichum',  
'VERONICA ANAGALLIS-AQUATICA',  
'Helianthemum marifolium',  
'Filago pyramidata',  
'Echinochloa crus-galli',  
'Quercus coccifera',  
'Aegilops kotschyi',  
'Potentilla inclinata',  
'Vaccinium myrtillus',  
'Plantago afra',  
'Hippocrepis unisiliquosa',  
'Satureja macrantha',  
'Geum rivale',  
'Betula fruticosa',  
'Salix raddeana',  
'Rhamnus cathartica',
```

```

'Rhaponticum coniferum',
'Astragalus microcephalus',
'Potentilla szovitsii',
'Rhamnus pallasii',
'Festuca rupicola',
'Medicago polymorpha',
'Brachypodium retusum',
'Scutellaria pinnatifida',
'Torilis leptophylla',
'Bassia prostrata',
'Urtica dioica',
'Corylus avellana',
'Punica granatum'
]

```

```

In [59]: # Value transformation.
def map_to_3categories(row):
    """ Maps values to "evergreen", "semi-evergreen", or "deciduous". """
    value = row.OrigValueStr
    if value == value: # Not NaN.
        value = str(value).lower()
        form = str(row.value_form).lower()
        name = str(row.OrigName).lower()

    # Hard coded mapping for some species based on predictions from
    # ChatGPT 3.5.
    if row.AccSpeciesName in ai_evergreen: return "evergreen"
    if row.AccSpeciesName in ai_semi: return "semi"
    if row.AccSpeciesName in ai_deciduous: return "deciduous"
    if name == "decev" and form == "n":
        if row.AccSpeciesName in [
            "Osmundastrum cinnamomeum",
            "Phegopteris hexagonoptera",
            "Polystichum acrostichoides"
        ]: return "evergreen"
        if value in [
            "Dryopteris intermedia",
            "Onoclea sensibilis",
            "Carex lanceolata"
        ]: return "semi"
        else: return "deciduous"

```

```

# Map numeric values to encoded categories based on info
# from metadata.
if form in ["@", "@.@"]:
    if name == "leaf persistence":
        if value == "1": return "deciduous"
        if value == "2": return "semi"
        if value == "3": return "evergreen"
    if "months without leaves" in name:
        value = float(value)
        if value == 0: return "evergreen"
        if value >= 2: return "deciduous"
        if value < 2: return "semi"
    if name == "evergreen" and value == "1":
        return "evergreen"
    if name == "deciduous" and value == "1":
        return "deciduous"

# Map to phenology type based on Leaf persistence time information.
if "always" in value:
    if "persistent" in value or "overwintering" in value:
        return "evergreen"
    if "summer" in value or "spring" in value:
        return "deciduous"

# Map yes/no label encodings to true meanings.
if form in ["y", "n"] and name in ["evergreen"]:
    value = value.replace("n", "deciduous")
    value = value.replace("y", "evergreen")
    value = value.replace("/", " ")
    return value
if value == "yes":
    if "semi" in name:
        return "semi"
    if "evergreen" in name or "leaf retention" in name:
        return "evergreen"
    if "deciduous" in name:
        return "deciduous"

# Map alternate notations to real meanings.
if (("semi" in value) or
    ("partly" in value) or
    ("deci" in value and "ever" in value)):

```

```

        return "semi"
    if value in ["sd", "sg", "bd", "semi", "d_ev", "ed"]:
        return "semi"
    if value in ["d", "dc", "w", "nonevergreen", "summergreen",
                "aestival", "vernal", "hibernal", "caducifolio"]:
        return "deciduous"
    if value in ["e", "ev", "aphyllous", "exchanger",
                "perennifolio", "leafless", "stem succulent"]:
        return "evergreen"
    if "duous" in value:
        return "deciduous"
    if "ever" in value:
        return "evergreen"

    # Map to class based on growth period information.
    if ("growth period" in name or
        "growing period" in name or
        "growth priod" in name):
        if "year round" in value or "any time with rain" in value:
            return "evergreen"
        if value in ["spring and summer",
                    "spring, summer, fall",
                    "summer and fall",
                    "spring, summer, autumn",
                    "spring, autumn, winter"]:
            return "semi"
        return "deciduous"

    return np.nan

```

```
t_val_std = DFColValTransformation(map_to_3categories, "OrigValueStr")
```

```

In [60]: # CONFIGURE TRANSFORMATIONS
FH.transforms["non_std"]["trait"] = [t_val_std]

```

```

In [61]: # TEST TRANSFORMATIONS.
res = FH.data_trait["non_std"]
res = t_val_std(res)
print(res.OrigValueStr.unique()) # Works!

```

```
[nan 'deciduous' 'evergreen' 'semi']
```

```
In [62]: # APPLY TRANSFORMATIONS
        FH.apply_transformations("trait")
```

```
APPLYING TRANSFORMATIONS:
Transforming non-standardized trait data ... done.
ALL DONE :3
```

```
In [63]: # ENSURE SUCCESSFUL APPLICATION
        FH.data_trait["non_std"].OrigValueStr.unique() # Correct!
```

```
Out[63]: array([nan, 'deciduous', 'evergreen', 'semi'], dtype=object)
```

3. Extract Locations

3.1. ID Review

```
In [64]: # VIEW KNOWN IDS
        print(f"Keep IDs = {FH.keep_ids['lonlat']}")
        FH.known_ids["lonlat"]
```

```
Keep IDs = [60, 506, 524, 6959, 4705, 4707, 4710, 4711, 8302, 59, 505, 523, 6958, 4704, 4706, 4708, 4709, 8301]
```


Out[64]:

	id	name	coordinate
0	60	Longitude	lon
1	506	Longitude of provenance	lon
2	524	Longitude of provenance of seed / transplant /...	lon
3	6959	Longitude grid	lon
4	4705	Longitude (decimal degrees)	lon
5	4707	Longitude estimated	lon
6	4710	Longitude minimum (decimal degrees)	lon
7	4711	Longitude maximum (decimal degrees)	lon
8	8302	UTM Y (longitude)	lon
9	59	Latitude	lat
10	505	Latitude of provenance	lat
11	523	Latitude of provenance of seed / transplant / ...	lat
12	6958	Latitude grid	lat
13	4704	Latitude (decimal degrees)	lat
14	4706	Latitude estimated	lat
15	4708	Latitude minimum (decimal degrees)	lat
16	4709	Latitude maximum (decimal degrees)	lat
17	8301	UTM X (latitude)	lat

In [65]:

```
# VIEW TRAITS CONSIDERED
FH.get_considered_traits("lonlat").name.to_list()
```

```
Out[65]: ['Longitude',
          'Longitude of provenance',
          'Longitude of provenance of seed / transplant / seedling',
          'Longitude grid',
          'Longitude (decimal degrees)',
          'Longitude estimated',
          'Longitude minimum (decimal degrees)',
          'Longitude maximum (decimal degrees)',
          'UTM Y (longitude)',
          'Latitude',
          'Latitude of provenance',
          'Latitude of provenance of seed / transplant / seedling',
          'Latitude grid',
          'Latitude (decimal degrees)',
          'Latitude estimated',
          'Latitude minimum (decimal degrees)',
          'Latitude maximum (decimal degrees)',
          'UTM X (latitude)']
```

3.2 Extract Data

```
In [66]: FH.extract_lonlat_data()
```

```
EXTRACTING LON LAT DATA ...
No. of std lon data rows. = 97332
No. of non_std lon data rows. = 69847
No. of std lat data rows. = 98371
No. of non_std lat data rows. = 69844
ALL DONE :3
```

3.3. Manual Investigation

```
In [67]: # What do value units and forms look like?
FH.view_units_value_forms(data_type="lonlat")
```

```

Lonlat Longitude Standardised Units: []
Lonlat Longitude Non-Standardised Units: ['dec', 'decimal_degrees']
Lonlat Longitude Standardised Value Forms: ['@.@', '-@.@']
Lonlat Longitude Non-Standardised Value Forms: ['@.@', '-@.@', '@', '@ @' w', '@ @' @' ' w', '@ @\' @" w', '@ @' e', '@ @ @ e', '@ @\'w', '@ @\'w', '@ @ @e', '@ @\' @" e', '@.@ e', '@ w', '@ e', '@.@ w', '@ @.@ w', '@ @.@w', '@.@e-', '-@', '@°@ e"]
Lonlat Latitude Standardised Units: []
Lonlat Latitude Non-Standardised Units: ['dec', 'decimal_degrees']
Lonlat Latitude Standardised Value Forms: ['@.@', '-@.@']
Lonlat Latitude Non-Standardised Value Forms: ['-@.@', '@.@', '@', '-@', '@ @' n', '@ @' @' ' n', '@ @\' @" n', '@ @' s', '@ @ @ s', '@ @\' @" s', '@.@ s', '@ s', '@.@ n', '@ n', '@ @ @ n', '@ @ ' n', '@ @\'n', '@°@s"]

```

```

In [68]: # Combine non_std Lon & Lat data for easier investigation.
         data_latlon_non_std = FH.get_combine_lonlat("non_std")

```

```

In [69]: # View context.
         FH.get_context(data_latlon_non_std)
         # OBSERVATION: Min and max Latitude and Longitude data can be averaged.

```

Out[69]:

	DataName	OrigName	OrigUnitStr	Comment
0	Longitude of provenance	Collection longitude	NaN	decimal degrees;
185	Longitude of provenance	Collection longitude	NaN	decimal degrees
591	Longitude of provenance of seed / transplant /...	Longitude-seed	dec	NaN
600	Longitude	Longitude	NaN	NaN
603	Longitude of provenance	csLong	NaN	NaN
1491	Longitude of provenance	Provenance Longitude	NaN	Provenance Longitude
1796	Longitude grid	Lon_grid	NaN	NaN
4318	Longitude (decimal degrees)	Longitude	decimal_degrees	Decimal degrees longitude of data collection l...
4490	Longitude estimated	Longitude_Estimated	decimal_degrees	When original source does not specify coordina...
4641	Longitude minimum (decimal degrees)	Min_Longitude	decimal_degrees	Minimum longitude of data collection location ...
4642	Longitude maximum (decimal degrees)	Max_Longitude	decimal_degrees	Maximum longitude of data collection location ...
32307	Longitude of provenance	seed_source_longitude	NaN	species' seed source longitude
69822	UTM Y (longitude)	COORDENADA Y	NaN	NaN
69847	Latitude of provenance	Collection Latitude	NaN	decimal degrees as above; google earth seems t...
70032	Latitude of provenance	Collection Latitude	NaN	decimal degrees as above; google earth seems t...
70438	Latitude of provenance of seed / transplant / ...	Latitude-for seed	dec	NaN
70447	Latitude of provenance	csLat	NaN	NaN
71335	Latitude of provenance	Provenance Latitude	NaN	Provenance Latitude

	DataName	OrigName	OrigUnitStr	Comment
71640	Latitude grid	Lat_grid	NaN	NaN
74162	Latitude (decimal degrees)	Latitude	decimal_degrees	Decimal degrees latitude of data collection lo...
74334	Latitude estimated	Latitude_Estimated	decimal_degrees	When original source does not specify coordina...
74485	Latitude minimum (decimal degrees)	Min_Latitude	decimal_degrees	Minimum latitude of data collection location w...
74486	Latitude maximum (decimal degrees)	Max_Latitude	decimal_degrees	Maximum latitude of data collection location w...
102151	Latitude of provenance	seed_source_latitude	NaN	species' seed source latitude
139612	Latitude	Latitude	NaN	NaN
139666	UTM X (latitude)	COORDENADA X	NaN	NaN

```
In [70]: # View context.
FH.get_context(data_latlon_non_std, context_cols=["DataID", "DataName"])
# OBSERVATION: Min max values can be averaged.
```

Out[70]:

	DataID	DataName
0	506	Longitude of provenance
591	524	Longitude of provenance of seed / transplant /...
600	60	Longitude
1796	6959	Longitude grid
4318	4705	Longitude (decimal degrees)
4490	4707	Longitude estimated
4641	4710	Longitude minimum (decimal degrees)
4642	4711	Longitude maximum (decimal degrees)
69822	8302	UTM Y (longitude)
69847	505	Latitude of provenance
70438	523	Latitude of provenance of seed / transplant / ...
71640	6958	Latitude grid
74162	4704	Latitude (decimal degrees)
74334	4706	Latitude estimated
74485	4708	Latitude minimum (decimal degrees)
74486	4709	Latitude maximum (decimal degrees)
139612	59	Latitude
139666	8301	UTM X (latitude)

```
In [71]: # View values in UTM format.
data_utm = FH.get_utm_data()
print(f"DatasetIDs = {data_utm.DatasetID.unique().tolist()}")
# OBSERVATION: Zone = 18. Hemisphere = N.
```

DatasetIDs = [653]

```
In [72]: data_latlon_non_std.DatasetID.unique()
```

```
Out[72]: array([ 1, 37, 87, 35, 89, 241, 443, 339, 469, 511, 653])
```

```
In [73]: # View Data with no unit.
print("LON LAT DATA WITH UNIT = NAN")
ll_unit_nan = data_latlon_non_std[data_latlon_non_std.OrigUnitStr.isna()]
print("value forms =", ll_unit_nan.value_form.unique())
# OBSERVATION: Seems to be in degrees.
# print("value range =", ll_unit_nan.OrigValueStr.unique().astype(float))
```

LON LAT DATA WITH UNIT = NAN

```
value forms = ['@.@' '-@.@' '@' '@ @' w' '@ @' @' w' '@ @\' @' w' '@ @' e' '@ @ @ e'
 '@ @\' w' '@ @\' w' '@ @ @e' '@ @\' @' e' '@.@ e' '@ w' '@ e' '@.@ w'
 '@ @.@' w' '@ @.@\' w' '@.@e-@' '-@' '@°@' e' '@ @' n' '@ @' @' n'
 '@ @\' @' n' '@ @' s' '@ @ @ s' '@ @\' @' s' '@.@ s' '@ s' '@.@ n' '@ n'
 '@ @ @ n' '@ @ ' n' '@ @\' n' '@°@\' s']
```

Initially, it was assumed that all units other than UTM related ones are decimal degrees. But testing transformations revealed that this assumption led to dataset id 35 values being > 180. So, it must be investigated whether here, degrees are in the 0 to 360 range instead of the standard -180 to 180 and -90 to 90 range.

```
In [74]: # View values in dataset ID 35.
ll_ds_35 = data_latlon_non_std[
    data_latlon_non_std.DatasetID.astype(float) == 35]
# OBSERVATION: Based on measurement related to this dataset observed here:
# https://www.try-db.org/de/DatasetDetails.php. It seems that Locations are
# in decimal degrees in the standard range with negative values present as well.
# So, the value of 200.335 here, may be a mistake.
# DECISION: Drop erroneous values in datasetID 35.
```

```
In [75]: # Examine standardized data.
data_lonlat_std = FH.get_combine_lonlat("std")
FH.view_range("lonlat", "std")
# OBSERVATION: Standardized Latitude values contain invalid measurements > 90.
```

Std Longitude Value Range = [-165.75, 176.2166667]

Std Latitude Value Range = [-65.599747, 236.6641]

```
In [76]: display(data_lonlat_std[(data_lonlat_std.DataName == "Latitude") &
                                (data_lonlat_std.StdValue > 90)])
# OBSERVATION: There is only one such value. It must be a mistake and can be dropped.
# display(data_lonlat_std[data_lonlat_std.DatasetID == 432])
```

	DatasetID	AccSpeciesID	AccSpeciesName	ObservationID	ObsDataID	TraitID	TraitName	DataID	DataName	OrigName
136010	432	34542	Lycium andersonii	2750112	25311962	NaN	NaN	59	Latitude	Latitude (degree)

3.4. Define & Apply Transformation Functions

```
In [77]: # Unit standardization.
def ll_unit_std(row):
    """ Adds zone and hemisphere information for UTM values
        in the "OrigValueStr" column of the dat frame. Maps
        unit "dec" to "decimal degrees".
    """
    # unit = row.OrigUnitStr
    # if unit == unit and unit in ["dec", "decimal_degrees"]:
    #     return "decimal degrees"
    dataset_id = row.DatasetID
    if dataset_id == dataset_id and int(dataset_id) == 653:
        return "utm zone_18 hemisphere_n"
    return "decimal degrees"

def ll_ds35(row):
    """ Drop values in DatasetID 35 that are outside valid range. """
    dataset_id = row.DatasetID
    name = row.DataName.lower()
    value = float(row.OrigValueStr)
    if dataset_id == dataset_id and name == name and value == value:
        if int(dataset_id) == 35:
            if "longitude" in name and value < -180 or value > 180:
                return np.nan
            elif "latitude" in name and value < -90 or value > 90:
                return np.nan
    return row.OrigValueStr
```



```
def ll_drop_gt180(row):
    """ Remove standardized values > 180. """
    v = row.StdValue
    if v == v and float(v) > 180:
        return np.nan
    return v

t_ll_drop_igt180 = DFColValTransformation(ll_drop_gt180, "StdValue")
t_ll_unit_std = DFColValTransformation(ll_unit_std, "OrigUnitStr")
t_ll_value_std = get_transformation_lonlat_std()
t_ll_ds35 = DFColValTransformation(ll_ds35, "OrigValueStr")
t_ll_form_recompute = get_transformation_get_value_form("OrigValueStr")
```

```
In [78]: # TEST TRANSFORMATIONS
res_ll = data_latlon_non_std
print("before:", res_ll.value_form.unique())
res_ll = t_ll_unit_std(res_ll)
res_ll = t_ll_value_std(res_ll)
res_ll = t_ll_ds35(res_ll)
res_ll = t_ll_form_recompute(res_ll)
res_ll.loc[:, "OrigValueStr"] = res_ll.OrigValueStr.astype(float)
print("after:", res_ll.value_form.unique())
res_ll_no_utm = res_ll[np.logical_not(
    res_ll.OrigUnitStr.astype(str).str.contains("utm"))]
res_ll_no_utm.OrigValueStr.astype(float).describe() # Success.
```

```
before: ['@.@' '-@.@' '@' '@ @' w' '@ @' '@' w' '@ @\' @' w' '@ @' e' '@ @ @ e'
 '@ @\' w' '@ @\' w' '@ @ @e' '@ @\' @' e' '@.@ e' '@ w' '@ e' '@.@ w'
 '@ @.@' w' '@ @.@\' w' '@.@e-@' '-@' '@°@' e' '@ @' n' '@ @' '@' n'
 '@ @\' @' n' '@ @' s' '@ @ @ s' '@ @\' @' s' '@.@ s' '@ s' '@.@ n' '@ n'
 '@ @ @ n' '@ @ ' n' '@ @\' n' '@°@\' s']
after: ['@.@' '-@.@' nan '@']
```

```
Out[78]: count    139638.000000
         mean      27.577313
         std       72.350832
         min      -165.750000
         25%      -31.100000
         50%       34.823000
         75%       57.978400
         max       176.216667
         Name: OrigValueStr, dtype: float64
```

```
In [79]: # CONFIGURE TRANSFORMATIONS
         FH.transforms["std"]["lonlat"] = [t_ll_drop_igt180, t_ll_form_recompute]
         FH.transforms["non_std"]["lonlat"] = [
             t_ll_unit_std, t_ll_value_std, t_ll_ds35, t_ll_form_recompute]
```

```
In [80]: # APPLY TRANSFORMATIONS
         FH.apply_transformations("lonlat")
```

APPLYING TRANSFORMATIONS:
 Transforming non-standardized lonlat data ... done.
 Transforming standardized lonlat data ... done.
 ALL DONE :3

```
In [81]: # CONVERT UTM VALUES
         FH.lonlat_utm_to_decimal_degrees()
```

```
In [82]: # AVERAGE MIN & MAX LON LAT DATA IDS
         FH.avg_trait_values(data_type="lonlat",
                             id1=4710, id2=4711) # Longitude
         FH.avg_trait_values(data_type="lonlat",
                             id1=4708, id2=4709) # Latitude
```

```
In [83]: # CHECK IF WORKED
         FH.view_range("lonlat", "std")
         FH.view_range("lonlat", "non_std")
         # Yes
```

Std Longitude Value Range = [-165.75, 176.216667]
 Std Latitude Value Range = [-65.599747, 78.88333333]
 Non Std Longitude Value Range = [-165.75, 176.216667]
 Non Std Latitude Value Range = [-51.59388889, 78.883]

4. Extract Years

4.1. ID Review

```
In [84]: # VIEW KNOWN IDS  
print(f"Keep IDs = {FH.keep_ids['year']}")  
FH.known_ids['year']
```

Keep IDs = [241, 211, 212, 695, 696, 697, 698, 699, 2571, 6601, 1664, 4688, 4689, 4690, 4691, 4692, 4693]

Out[84]:

	id	name
0	241	Measurement date / sampling date
1	211	Sampling date: month
2	212	Sampling date: year
3	695	Measurement date: season
4	696	Measurement date: year
5	697	Measurement date: month
6	698	Measurement date: day
7	699	Measurement date: hour
8	2571	Average number of ground frost days per year (...)
9	6601	Sampling or measurement date standardized
10	1664	Measurement date: Day of Year (DOY)
11	4688	Sampling date: first year of sample collection
12	4689	Sampling date: first month of sample collection
13	4690	Sampling date: first day of sample collection
14	4691	Sampling date: last year of sample collection
15	4692	Sampling date: last month of sample collection
16	4693	Sampling date: last day of sample collection

```
In [85]: # UPDATE KEEP_IDS
FH.keep_ids["year"] = [6601, 241, 212, 696, 4688, 4691]
FH.get_considered_traits("year")
```

Out[85]:

	id	name
0	241	Measurement date / sampling date
2	212	Sampling date: year
4	696	Measurement date: year
9	6601	Sampling or measurement date standardized
11	4688	Sampling date: first year of sample collection
14	4691	Sampling date: last year of sample collection

4.2. Extract Data

```
In [86]: # Data Extraction.
        FH.extract_year_data()
```

```
EXTRACTING YEAR DATA ...
No. of std data rows. = 85470
No. of non_std data rows. = 98082
ALL DONE :3
```

4.3. Manual Investigation

```
In [87]: # What do value units and forms Look Like?
        FH.view_units_value_forms(data_type="year")
        # OBSERVATION: Both standardized and non-standardized values
        # need to be value standardized.
```

```

Year Standardised Units: []
Year Non-Standardised Units: ['year/month']
Year Standardised Value Forms: ['@', '@-@-@', '(@/@)', '(@/@)-(@/@)', '(@/@)-s/s', '(@/@)-(@/@&@/@)', '@-(@/@)', '@-@', '@-s', '@-@-(@/@)', '@-m', '@-m-m', '@-m-@', 'm', '', 'm-@-@', '@ to @', '@-m to @-m', '@-m to @', '@-m-@ to @-m-@', '@-m-@ to @-m', 'm-@ to m-@', '@-@/@', '@-@/@-@']
Year Non-Standardised Value Forms: ['@', '@/@/@', '@-@.@.@', 'm. @', 's @', 'm @?', 'm @', '@/@', 'm & m @', 'm/m. @', '@-@', 'm/m. @/m. @', '@.m @', 'monthly measurements from spt. @ to m @', 'm-m @', '@?', 'm-m. @', 'mi-mi @', 'm t. @', 's @/s @', 'm/m @', '@-@?', 'spet. @/m @', 'm @, during mid morning hours (@-@h. solar time)', 'm-m, @-@', 'm - m @, end of rain season', 'm-m @ and @, m-m @', '@ m', 'm @ @', 'm', 'none', 'm @ @', '@-@ m', '@/@/@ @:@', '@.@. @', 'm-m-@', 'm-m/@', '@/m', '@/@-@/@']

```

```

In [88]: # View context.
         FH.get_context(FH.data_year["non_std"])

```

Out[88]:

	DataName	OrigName	OrigUnitStr	Comment
14319	Measurement date / sampling date	Harvest # (within study)	NaN	if there are multiple harvests of the same spe...
27206	Measurement date / sampling date	sampling date	NaN	NaN
120242	Sampling date: year	Collection.year	NaN	NaN
146450	Sampling date: year	year	NaN	NaN
400365	Measurement date / sampling date	Time	NaN	NaN
402380	Measurement date: year	yearstart	NaN	Beginning of measurements
469223	Measurement date / sampling date	Date	NaN	NaN
633148	Measurement date / sampling date	date of sampling	NaN	NaN
634277	Measurement date / sampling date	year	NaN	NaN
634291	Measurement date / sampling date	Year of survey	NaN	NaN
895431	Measurement date / sampling date	date	NaN	NaN
898466	Measurement date / sampling date	SampleDate	NaN	NaN
903683	Measurement date / sampling date	Sampling date	NaN	NaN
910929	Measurement date / sampling date	Date	NaN	Date of measurement
910930	Sampling date: year	Year	NaN	Year of measurement
1005021	Sampling date: year	Collection year	NaN	NaN

	DataName	OrigName	OrigUnitStr	Comment
1167087	Measurement date / sampling date	Date	NaN	Date of the year when measurement conducted
1964345	Measurement date: year	Sample collection_Measurement year	NaN	Year of sample collection.
1967349	Sampling date: first year of sample collection	Sample collection_Year beginning collection	NaN	The first year of sample collection for data c...
1967352	Sampling date: last year of sample collection	Sample collection_Year ending collection	NaN	The final year of sample collection for data c...
3049318	Measurement date / sampling date	Date	NaN	data collection date
3811571	Measurement date / sampling date	Year	NaN	NaN
3812364	Measurement date / sampling date	date	year/month	all measurements early-mid december 2013
3817684	Measurement date / sampling date	DateMeasurement	NaN	NaN

```
In [89]: # View rows with specific Comments.
FH.get_unique_matches(
    data=FH.data_year["non_std"], match_col="Comment",
    to_match=["all measurements early-mid december 2013"]).OrigValueStr.unique()
```

```
Out[89]: array(['2013/December'], dtype=object)
```

```
In [90]: # OBSERVATION: Averaging of years for all those rows that have both
# first and last year values during preprocessing is required.
```

4.4. Define & Apply Transformation Functions

```
In [91]: # DEFINE TRANSFORMATIONS
t_y_val_std_non_std_data = get_transformation_get_year("OrigValueStr")
t_y_val_std_std_data = get_transformation_get_year("StdValue")
```



```
t_y_form_recompute_std_data = get_transformation_get_value_form("StdValue")
t_y_form_recompute_non_std_data = get_transformation_get_value_form("OrigValueStr")
```

```
In [92]: # CHECK TRANSFORMATIONS
res_y = FH.data_year["non_std"]
res_y = t_y_val_std_non_std_data(res_y)
res_y = t_y_form_recompute_non_std_data(res_y)
print(res_y.value_form.unique().tolist()) # Success.
```

```
[nan, '@']
```

```
In [93]: # CONFIGURE TRANSFORMATIONS
FH.transforms["std"]["year"] = [t_y_val_std_std_data,
                                t_y_form_recompute_std_data]
FH.transforms["non_std"]["year"] = [t_y_val_std_non_std_data,
                                     t_y_form_recompute_non_std_data]
```

```
In [94]: # APPLY DEFINED TRANSFORMATIONS
FH.apply_transformations(data_type="year")
```

APPLYING TRANSFORMATIONS:

Transforming non-standardized year data ... done.

Transforming standardized year data ... done.

ALL DONE :3

```
In [95]: # CHECK TRANSFORMATION SUCCESS
pd.concat([FH.data_year["std"],
           FH.data_year["non_std"]]).value_form.unique() # Yes.
```

```
Out[95]: array(['@', nan], dtype=object)
```

```
In [ ]: # DECISION: Average first and last year data.
FH.avg_trait_values(data_type="year", id1=4688, id2=4691)

# CHECK TO ENSURE SUCCESS
FH.view_range(data_type="year", std_type="std")
FH.view_range(data_type="year", std_type="non_std")
```

Std Year Value Range = [1914.0, 2019.0]

Non Std Year Value Range = [1914.0, 2025.0]

5. Combine Data

In [97]: *# COMBINE DATA*

```
data_combined = FH.combine_data()  
data_combined.leaf_phenology_type.value_counts()
```

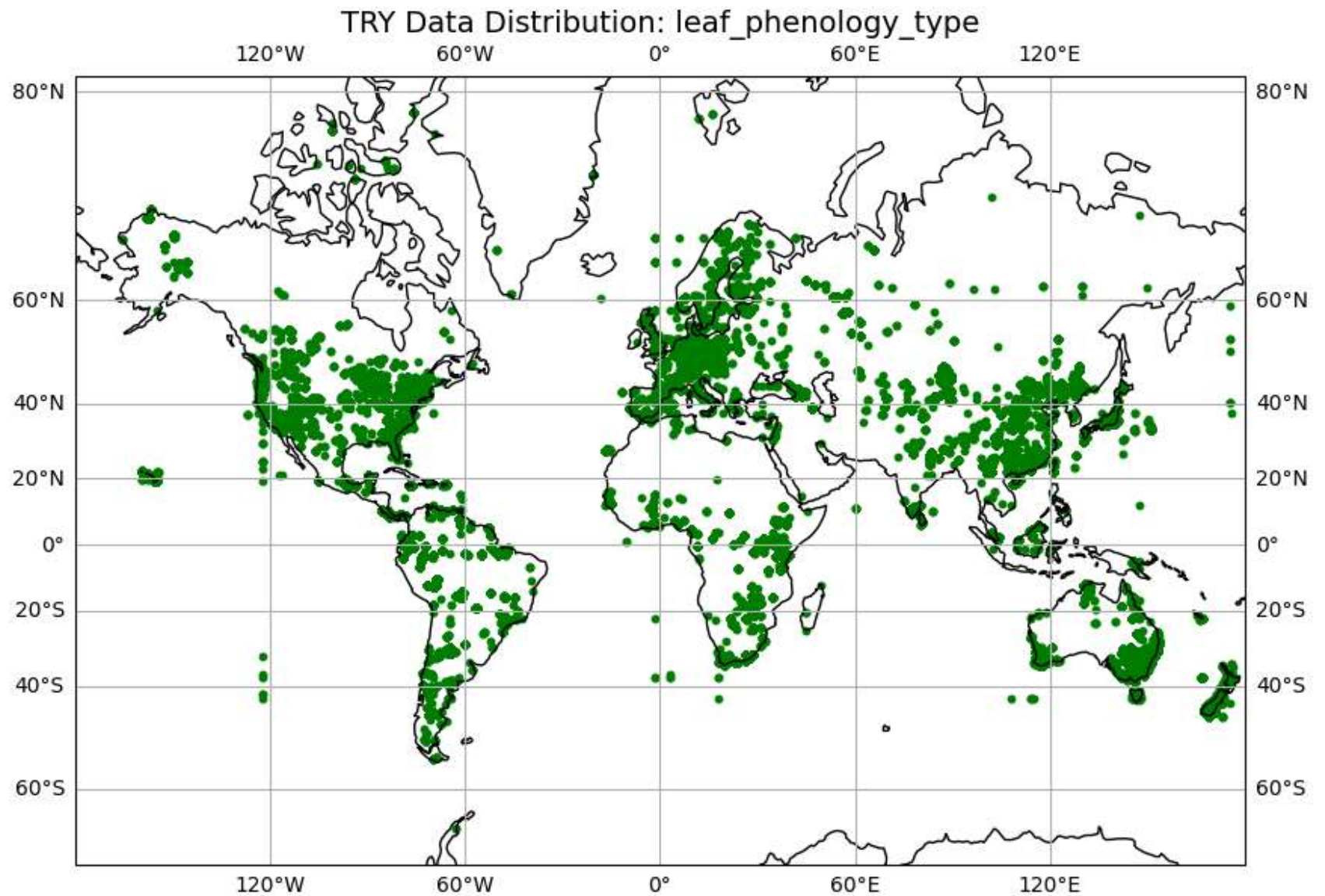
```
year value range = {'min': 1914.0, 'max': 2025.0}  
lon value range = {'min': -165.75, 'max': 176.2166667}  
lat value range = {'min': -65.6, 'max': 78.88}  
no. of rows = 70505  
no. of unique species IDs = 29214  
no. of unique (lon, lat) combinations = 6439  
no. of (lon, lat) combinations = 40459  
no. of unique leaf_phenology_type values = 3
```

Out[97]: leaf_phenology_type

```
evergreen    44060  
deciduous    23287  
semi         3158  
Name: count, dtype: int64
```

In [98]: *# View data distribution on map.*

```
map_plot(data_combined, title=f"TRY Data Distribution: {FH.feature_name}")
```



6. Save Processed Data

```
In [ ]: # Save data.  
save_data(  
    data = data_combined,
```

```
dest_fold = "../__data/try/extracted",  
feature_name = FH.feature_name,  
suffix="new")
```

Saved "leaf_phenology_type" data at "../__data/try/extracted/leaf_phenology_type_new.csv".