

Using regular expressions

Lauren Ponisio

Conservation/ecology Topics

- Species distributions

Computational Topics

- Use regular expressions to clean and categorize data

Part 1: Oregon bee atlas data exploration

Import the OBA data using your favorite parsing function, name the data oba.

a.

```
path <- "data/OBA_2018-2023.csv"
oba <- read.csv(path)
head(oba)
```

```
##              Observation.No. Voucher.No. user_id    user_login
## 1 Andony_Melathopoulos:18.001.001      429964 amelathopoulos
## 2 Andony_Melathopoulos:18.002.001      429964 amelathopoulos
## 3 Andony_Melathopoulos:18.002.002      429964 amelathopoulos
## 4 Andony_Melathopoulos:18.002.003      429964 amelathopoulos
## 5 Andony_Melathopoulos:18.002.004      429964 amelathopoulos
## 6 Andony_Melathopoulos:18.002.005      429964 amelathopoulos
## Collector...First.Name Collector...First.Initial Collector...Last.Name
## 1              Andony              A.              Melathopoulos
## 2              Andony              A.              Melathopoulos
## 3              Andony              A.              Melathopoulos
## 4              Andony              A.              Melathopoulos
## 5              Andony              A.              Melathopoulos
## 6              Andony              A.              Melathopoulos
## Collectors taxon_kingdom_name Associated.plant...genus..species url
## 1 A.Melathopoulos
## 2 A.Melathopoulos
## 3 A.Melathopoulos
## 4 A.Melathopoulos
```

```

## 5 A.Melathopoulos
## 6 A.Melathopoulos
## Sample.ID Specimen.ID Collection.Day.1 Month.1 MonthJul MonthAb Year.1
## 1 NA 18 iii March 3 2018
## 2 NA 20 iii March 3 2018
## 3 NA 20 iii March 3 2018
## 4 NA 20 iii March 3 2018
## 5 NA 2 ix September 9 2018
## 6 NA 2 ix September 9 2018
## Collection.Date Time.1 Collection.Day.2 Month.2 Year.2 Collection.Day.2.Merge
## 1 3/18/2018
## 2 3/20/2018
## 3 3/20/2018
## 4 3/20/2018
## 5 9/2/2018
## 6 9/2/2018
## Time.2 Collection.ID Position.of.1st.digit Collection.No. Sample.No.
## 1 A Melathopoulos 1 1
## 2 A Melathopoulos 2 1
## 3 A Melathopoulos 2 2
## 4 A Melathopoulos 2 3
## 5 A Melathopoulos 2 4
## 6 A Melathopoulos 2 5
## Country State County Location
## 1 USA Oregon Benton Corvallis, NW Orchard Ave
## 2 USA Oregon Benton Corvallis, NW Orchard Ave
## 3 USA Oregon Benton Corvallis, NW Orchard Ave
## 4 USA Oregon Benton Corvallis, NW Orchard Ave
## 5 USA Oregon Clatsop Clatskanie, Big Creek Mainline, Knob Point Road
## 6 USA Oregon Clatsop Clatskanie, Big Creek Mainline, Knob Point Road
## Abbreviated.Location Collection.Site.Description Team
## 1 Astoria Maggie Johnson Rd Melathopoulos
## 2 Big Crk. Mainline Knob Pt Rd Melathopoulos
## 3 Big Crk. Mainline Knob Pt Rd Melathopoulos
## 4 Big Crk. Mainline Knob Pt Rd Melathopoulos
## 5 Big Crk. Mainline Knob Pt Rd Melathopoulos
## 6 Big Crk. Mainline Knob Pt Rd Melathopoulos
## Habitat Elevation..m. Dec..Lat. Dec..Long. X Collectionmethod
## 1 44.556 -123.285 NA Net
## 2 44.567 -123.283 NA Net
## 3 44.567 -123.283 NA Net
## 4 44.567 -123.283 NA Net
## 5 46.102 -123.506 NA Net
## 6 46.102 -123.506 NA Net
## Collection.method.merge.field Associated.plant...family
## 1
## 2
## 3
## 4
## 5
## 6
## Associated.plant...genus..species.1 Associated.plant...Inaturalist.URL
## 1
## 2

```

```

## 3
## 4
## 5
## 6
## Associated.plant Assoc.plant.merge.field Collectors.1
## 1 Andony Melathopoulos
## 2 Andony Melathopoulos
## 3 Andony Melathopoulos
## 4 Andony Melathopoulos
## 5 Andony Melathopoulos
## 6 Andony Melathopoulos
## Collector.1.abreviation Collector.2 Collector.3 Genus Species sex caste
## 1 A Melathopoulos NA NA
## 2 A Melathopoulos NA NA
## 3 A Melathopoulos NA NA
## 4 A Melathopoulos NA NA
## 5 A Melathopoulos NA NA
## 6 A Melathopoulos NA NA
## vol.det.Genus vol.det.Species vol.det.sex.caste Determined.By Date.Determined
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
## Verified.By Other.Determiner.s. Other.Dets.Sci..Name.s. Other.Dets..Date.s.
## 1 NA NA NA
## 2 NA NA NA
## 3 NA NA NA
## 4 NA NA NA
## 5 NA NA NA
## 6 NA NA NA
## Additional.Notes X.1
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA

```

- b. Examine the unique entries of 'Associated.plant' using any function you find useful. What are at least two patterns in the associated taxa string what should be removed if we want consistent plant names? (Make a list together as a class). Only print the first 10 here to avoid having a giant output.

```
unique(oba$Associated.plant)[1:10]
```

```

## [1] "" "Salix"
## [3] "Arctostaphylos densiflora" "Lithodora diffusa"
## [5] "Ceanothus gloriosus" "Ceanothus sp."
## [7] "Berberidaceae" "Ceanothus cuneatus"
## [9] "Prunus sp." "Camassia quamash"

```

```
# Two patterns in the associated taxa string that should be removed:
# "sp."
# ""
```

1. Sometimes there is only one word, often family, genus, or common name. We will sort these out by creating a column for plant resolution.
2. The common name is sometimes listed after the scientific name in (), we can strip this out.
3. Sometimes there is a list of plant names, cannot do much with those but drop them.
4. Some are blank (no name), we will drop these.
5. One says “net”
6. Sometimes a genus is followed by an sp. and sometimes not. I assume this can be solved with a regular expression, ignoring the rows with a family name, but I could not work out how.
7. There are a few with “genus XX” or “genus XX”

There may be more I am missing. :/

In week in lecture last I used a brute force pattern to remove some of these issues so we could plot them as a network. Now that we are familiar with regular expressions we can do better.

- c. Work together as a class to resolve the issues you listed with the associated taxa column using any function combination that uses regular expressions. You can reassign the contents of the column Associated.plant or create a new column. Return the sorted, unique values, ex: `sort(unique(oba$Associated.plant))`. Leave the plants resolved only to genus or family for later.

I have removed a really strange issue with special characters (R converted an apostrophe into a special character) to start things off.

Hint: You must escape any special characters. For example to use (matches any space, tab, or newline) you must use `\s` in your pattern.

```
# 1. Remove common names in parentheses
oba$Associated.plant <- str_replace_all(oba$Associated.plant, "\\(.*?\\)", "")

# 2. Remove rows with lists of plant names
oba$Associated.plant <- ifelse(grepl(",", oba$Associated.plant), NA, oba$Associated.plant)

# 3. Drop blank entries
oba <- oba[!is.na(oba$Associated.plant) & oba$Associated.plant != "", ]

# 4. Remove the entry "net"
oba$Associated.plant <- ifelse(oba$Associated.plant == "net", NA, oba$Associated.plant)
oba <- oba[!is.na(oba$Associated.plant), ]

# 5. Standardize genus entries with sp.
oba$Associated.plant <- str_replace_all(oba$Associated.plant, "^((\\w+)$", "\\1 sp.")

# 6. Remove entries like "genus XX"
oba$Associated.plant <- ifelse(grepl("^genus", oba$Associated.plant, ignore.case = TRUE), NA, oba$Associated.plant)
oba <- oba[!is.na(oba$Associated.plant), ]

## Remove the special character
oba$Associated.plant <- str_replace_all(oba$Associated.plant, "\\x92", "")

## To check that it worked
sort(unique(oba$Associated.plant))[1:10]
```

```
## [1] "Abelia sp."           "Abronia latifolia"      "Acer circinatum"
## [4] "Acer macrophyllum"     "Acer palmatum"          "Acer sp."
## [7] "Achillea millefolium"  "Achillea millefolium "  "Achillea sp."
## [10] "Aclepias speciosa"
```

Part 2: Making a column for plant resolution

- Some plant species are resolved to species/subspecies, others to genus and others to family. If there are two or three words, we can assume the plant is resolved to species and subspecies, respectively, except if the string ends in “sp.” If there is only one word, this could be a genus or a family name. Family names always end in “aceae”, for example Lamiaceae (mints), Asteraceae (daisies).

We want to make a new column called `plantResolution` and assign it to “Family”, “Genus” or “Species” depending on the level of resolution associated taxa is resolved to. We will do this in two steps.

First use regular expressions to count up the number of words in each element of associated taxa. Assign the count to a new column called `plantTaxaWordCount`. Print the first 50 elements.

Hint: `str_count` may be useful.

```
oba$plantTaxaWordCount <- str_count(oba$Associated.plant, "\\S+")
head(oba[, c("Associated.plant", "plantTaxaWordCount")], 50)
```

```
##           Associated.plant plantTaxaWordCount
## 13           Salix sp.                2
## 14           Salix sp.                2
## 15           Salix sp.                2
## 16           Salix sp.                2
## 17           Salix sp.                2
## 18 Arctostaphylos densiflora          2
## 19           Lithodora diffusa          2
## 20           Ceanothus gloriosus          2
## 21           Ceanothus gloriosus          2
## 22           Ceanothus gloriosus          2
## 23           Ceanothus gloriosus          2
## 24           Ceanothus gloriosus          2
## 25           Ceanothus gloriosus          2
## 26           Salix sp.                2
## 27           Salix sp.                2
## 28           Salix sp.                2
## 29           Ceanothus sp.             2
## 30           Ceanothus sp.             2
## 31           Ceanothus sp.             2
## 32           Ceanothus sp.             2
## 33           Ceanothus sp.             2
## 34           Ceanothus sp.             2
## 35           Ceanothus sp.             2
## 36           Ceanothus sp.             2
## 37           Ceanothus sp.             2
## 38           Ceanothus sp.             2
## 39           Ceanothus sp.             2
## 40           Ceanothus sp.             2
```

```
## 41          Ceanothus sp.          2
## 42          Ceanothus sp.          2
## 43          Ceanothus sp.          2
## 44          Ceanothus sp.          2
## 45          Ceanothus sp.          2
## 58      Berberidaceae sp.          2
## 59      Berberidaceae sp.          2
## 60      Berberidaceae sp.          2
## 61      Berberidaceae sp.          2
## 62      Berberidaceae sp.          2
## 63      Berberidaceae sp.          2
## 64      Berberidaceae sp.          2
## 65      Berberidaceae sp.          2
## 66      Ceanothus cuneatus          2
## 67      Ceanothus cuneatus          2
## 68      Ceanothus cuneatus          2
## 69      Ceanothus cuneatus          2
## 70      Ceanothus cuneatus          2
## 71      Ceanothus cuneatus          2
## 72      Ceanothus cuneatus          2
## 73      Ceanothus cuneatus          2
## 74          Prunus sp.            2
```

- b. Write a for loop to assigned each entry of the column `plantResolution` to be “family”, “genus” or “species”. `table()` the final result. Hint: Don’t forget to initialize the new column. Starting with all NAs may be useful. Hint hint: The function `ifelse` returns one value if a TRUE and another if FALSE. It could be useful depending on your approach. Hint hint hint: `grep1` will return or TRUE or FALSE depending on whether it finds the pattern. Be careful with periods in patterns because alone they are a wild card character.

```
oba$plantResolution <- ifelse(
  str_detect(oba$Associated.plant, "aceae$"), "family",
  ifelse(
    str_detect(oba$Associated.plant, " sp\\\\.$"), "genus",
    ifelse(oba$plantTaxaWordCount == 1, "genus",
           "species")
  )
)
head(oba[, c("Associated.plant", "plantResolution")], 50)
```

```
##          Associated.plant plantResolution
## 13          Salix sp.          genus
## 14          Salix sp.          genus
## 15          Salix sp.          genus
## 16          Salix sp.          genus
## 17          Salix sp.          genus
## 18 Arctostaphylos densiflora    species
## 19      Lithodora diffusa      species
## 20      Ceanothus gloriosus    species
## 21      Ceanothus gloriosus    species
## 22      Ceanothus gloriosus    species
```

```

## 23      Ceanothus gloriosus      species
## 24      Ceanothus gloriosus      species
## 25      Ceanothus gloriosus      species
## 26              Salix sp.        genus
## 27              Salix sp.        genus
## 28              Salix sp.        genus
## 29      Ceanothus sp.            genus
## 30      Ceanothus sp.            genus
## 31      Ceanothus sp.            genus
## 32      Ceanothus sp.            genus
## 33      Ceanothus sp.            genus
## 34      Ceanothus sp.            genus
## 35      Ceanothus sp.            genus
## 36      Ceanothus sp.            genus
## 37      Ceanothus sp.            genus
## 38      Ceanothus sp.            genus
## 39      Ceanothus sp.            genus
## 40      Ceanothus sp.            genus
## 41      Ceanothus sp.            genus
## 42      Ceanothus sp.            genus
## 43      Ceanothus sp.            genus
## 44      Ceanothus sp.            genus
## 45      Ceanothus sp.            genus
## 58      Berberidaceae sp.        genus
## 59      Berberidaceae sp.        genus
## 60      Berberidaceae sp.        genus
## 61      Berberidaceae sp.        genus
## 62      Berberidaceae sp.        genus
## 63      Berberidaceae sp.        genus
## 64      Berberidaceae sp.        genus
## 65      Berberidaceae sp.        genus
## 66      Ceanothus cuneatus        species
## 67      Ceanothus cuneatus        species
## 68      Ceanothus cuneatus        species
## 69      Ceanothus cuneatus        species
## 70      Ceanothus cuneatus        species
## 71      Ceanothus cuneatus        species
## 72      Ceanothus cuneatus        species
## 73      Ceanothus cuneatus        species
## 74              Prunus sp.        genus

```

- c. For those that are identified to genus but are lacking an sp., add that now so that they will not be treated as separate plant species (i.e., Rosa vs Rosa sp.). You can do this with a regular expression and using 'gsub' or 'string_replace_all' or by counting up the number of words in Associated.plant.

```

oba$Associated.plant <- ifelse(
  str_count(oba$Associated.plant, "\\S+") == 1 & !str_detect(oba$Associated.plant, " sp\\.\\.$"),
  paste0(oba$Associated.plant, " sp."),
  oba$Associated.plant
)

## To check that it worked
unique(oba$Associated.plant[oba$plantResolution == "genus"])

```

## [1]	"Salix sp."	"Ceanothus sp."	"Berberidaceae sp."
## [4]	"Prunus sp."	"Vicia sp."	"Lupinus sp."
## [7]	"Salvia sp."	"Lamium sp."	"Monarda sp."
## [10]	"Teucrium sp."	"Hyacinthus sp."	"Geranium sp."
## [13]	"Alyssum sp."	"Aster sp."	"Sedum sp."
## [16]	"Caryopteris sp."	"Rudbeckia sp."	"Aquilegia sp."
## [19]	"Eriophyllum sp."	"Physocarpus sp."	"Sidalcea sp."
## [22]	"Escallonia sp."	"Taraxacum sp."	"Camassia sp."
## [25]	"Plagiobothrys sp."	"Pseudoveronica sp."	"Hebe sp."
## [28]	"Stachys sp."	"Mahonia sp."	"Pyrus sp."
## [31]	"Vaccinium sp."	"Acer sp."	"Dahlia sp."
## [34]	"Penstemon sp."	"Helenium sp."	"Lavandula sp."
## [37]	"Helianthus sp."	"Leucanthemum sp."	"Brassica sp."
## [40]	"Castanea sp."	"Heuchera sp."	"Rubus sp."
## [43]	"Apiaceae sp."	"Rosaceae sp."	"Caprifoliaceae sp."
## [46]	"Collinsia sp."	"Hydrophyllum sp."	"Syringa sp."
## [49]	"Fragaria sp."	"Thelypodium sp."	"Trifolium sp."
## [52]	"Delphinium sp."	"Wyethia sp."	"Symphoricarpos sp."
## [55]	"Senecio sp."	"Asteraceae sp."	"Potentilla sp."
## [58]	"Phacelia sp."	"Cirsium sp."	"Agastache sp."
## [61]	"Chaenactis sp."	"Melilotus sp."	"Apocynum sp."
## [64]	"Erysimum sp."	"Solidago sp."	"Lamiaceae sp."
## [67]	"Gaillardia sp."	"Malus sp."	"Rosa sp."
## [70]	"Plectritis sp."	"Phlox sp."	"Ericaceae sp."
## [73]	"Marah sp."	"Claytonia sp."	"Forsythia sp."
## [76]	"Salicaceae sp."	"Bellis sp."	"Rhododendron sp."
## [79]	"Hieracium sp."	"Lotus sp."	"Arctostaphylos sp."
## [82]	"Lomatium sp."	"Berberis sp."	"Composite sp."
## [85]	"Photinia sp."	"Eriogonum sp."	"Iris sp."
## [88]	"Lonicera sp."	"Arenaria sp."	"Ranunculus sp."
## [91]	"Clarkia sp."	"Verbena sp."	"Brodiaea sp."
## [94]	"Nepeta sp."	"Monardella sp."	"Hesperis sp."
## [97]	"Origanum sp."	"Foeniculum sp."	"Mentha sp."
## [100]	"Erigeron sp."	"Coreopsis sp."	"Fagopyrum sp."
## [103]	"Chrysothamnus sp."	"Crepis sp."	"Magnoliopsida sp."
## [106]	"Spiraea sp."	"Cistus sp."	"Myosotis sp."
## [109]	"Mimulus sp."	"Wisteria sp."	"Antirrhinum sp."
## [112]	"Cryptantha sp."	"Allium sp."	"Poales sp."
## [115]	"Chamaenerion sp."	"Atriplex sp."	"Cucurbita sp."
## [118]	"Cosmos sp."	"Raphanus sp."	"Grossulariaceae sp."
## [121]	"Narcissus sp."	"Erythronium sp."	"Hyacinthoides sp."
## [124]	"Boraginaceae sp."	"Doronicum sp."	"Crataegus sp."
## [127]	"Calendula sp."	"Weigela sp."	"Cichorioideae sp."
## [130]	"Verbenaceae sp."	"Dianthus sp."	"Achillea sp."
## [133]	"Tradescantia sp."	"Boraginales sp."	"Convolvulus sp."
## [136]	"Anthemideae sp."	"Cichorieae sp."	"Bromus sp."
## [139]	"Hypericum sp."	"Euphorbia sp."	"Betula sp."
## [142]	"Echinacea sp."	"Catalpa sp."	"Artemisia sp."
## [145]	"Sorbus sp."	"Ilex sp."	"Lunaria sp."
## [148]	"Balsamorhiza sp."	"Echinops sp."	"Scabiosa sp."
## [151]	"Chamaemelum sp."	"Yarrow sp."	"Arabis sp."
## [154]	"Nemophila sp."	"Crocidium sp."	"Gilia sp."
## [157]	"Magnoliophyta sp."	"Umbelliferae sp."	"Solanum sp."
## [160]	"Centaurea sp."	"Malva sp."	"Calochortus sp."


```
## [163] "Viola sp."          "Epilobium sp."      "Veronica sp."
## [166] "Symphytum sp."      "Cornus sp."         "Zinnia sp."
## [169] "Sphaeralcea sp."    "Tithonia sp."       "Crocus sp."
## [172] "Aguilegia sp."      "Abelia sp."         "Ericameria sp."
## [175] "Thymus sp."         "Net sp."            "Stephanomeria sp."
## [178] "Eriogonum sp."      "Tarweed sp."
```

- d. Create a new column called plantGenus that is the genus if the associated taxa was resolved to species or genus, and NA if it was resolved to family.

```
oba$plantGenus <- ifelse(oba$plantResolution == "family", NA, word(oba$Associated.plant, 1))
```

```
## To check finish with
table(oba$plantGenus)
```

```
##
##      Abelia      Abronia      Acer      Achillea      Aclepias
##          5          2         40         106          10
##      Aesculus    Agastache    Aguilegia      Alcea      Allium
##         13         50          2         18         67
##      Alyssum     Amelanchier    Amsinckia    Anaphalis    Anemone
##          1         41          2          4          1
##      Anethum      Angelica    Antennaria    Anthemideae    Antirrhinum
##          1          4          9          11          6
##      Apiaceae     Apocynum     Aquilegia      Arabis      Arbutus
##          2         32         20          2          1
##      Arctostaphylos    Arenaria      Arnica      Artemisia    Asclepias
##        98          9         51         16         69
##      Asparagus      Aster     Asteraceae    Astragalus    Astragalus
##          7         60         29         22          1
##      Atriplex      Aurinia    Balsamorhiza    Barbarea    Bellardia
##          1          1         54          8          2
##      Bellis     Berberidaceae    Berberis      Beta      Betula
##        214          8        159          4          2
##      Bidens      Bistorta    Blepharipappus    Boechera    Boraginaceae
##          4         15          9          7          6
##      Boraginales      Borago      Brassica    Brodiaea      Bromus
##          6          2        178         12          1
##      Calendula    Calochortus    Calystegia    Camassia    Caprifoliaceae
##          9         44          3        102          5
##      Cardaria    Carpenteria    Caryopteris    Castanea    Castilleja
##          8          1          9          2          1
##      Catalpa      Ceanothus    Centaurea    Centauria    Cerastium
##          1        388         25          1         19
##      Chaenactis    Chaenomeles    Chamaemelum    Chamaenerion    Chorisporea
##          6          2          1        51         20
##      Chrysopsis    Chrysothamnus    Chrysanthus    Cichorieae    Cichorioideae
##          3        120          5          5         20
##      Cirsium      Cistus      Clarkia    Claytonia    Cleome
##         69         12         29         64         13
##      Collinsia    Composite    Convolvulus    Coreopsis    Coriandrum
##         12          3          7         17          1
##      Cornus      Cosmos      Cotinus    Crataegus    Crepis
```

##	40	24	9	42	54
##	Crocidium	Crocus	Cryptantha	Cucurbita	Cuscuta
##	1	2	3	3	2
##	Cynoglossum	Cytisus	Dahlia	Damasonium	Dasiphora
##	2	13	7	1	23
##	Daucus	Delphinium	Descurainia	Deutzia	Dianthus
##	35	36	38	3	7
##	Dicentra	Dichelostemma	Digitalis	Dipsacus	Doronicum
##	5	76	17	13	20
##	Downingia	Drymocallis	Echinacea	Echinops	Elaeagnus
##	2	6	23	3	7
##	Epilobium	Ericaceae	Ericameria	Erigeron	Eriodictyon
##	51	6	524	51	18
##	Eriogonum	Eriogonum	Eriophyllum	Erodium	Eruca
##	1	119	212	12	9
##	Erysimum	Erythronium	Escallonia	Escholtzia	Eschscholzia
##	7	6	1	3	279
##	Euonymus	Euphorbia	Fagopyrum	Foeniculum	Forsythia
##	8	25	4	19	12
##	Fragaria	Fraxinus	Gaillardia	Gentiana	Geranium
##	76	3	26	1	115
##	Geum	Gilia	Gnaphalium	Grindelia	Grossulariaceae
##	6	39	1	53	8
##	Hackelia	Hastata	Hebe	Helenium	Helianthus
##	37	5	5	32	120
##	Heliopsis	Hemizonella	Heracleum	Hesperis	Heuchera
##	9	8	131	5	26
##	Hieracium	Hirschfeldia	Holodiscus	Horkelia	Humulus
##	9	1	30	31	1
##	Hyacinthoides	Hyacinthus	Hydrangea	Hydrophyllum	Hypericum
##	21	2	12	23	6
##	Hypochaeris	Hyssopus	Ilex	Iliamna	Iris
##	52	4	63	11	9
##	Isatis	Jacobaea	Jaumea	Kalmia	Lamiaceae
##	2	13	1	3	3
##	Lamium	Larkspur	Lasthenia	Lathyrus	Lavandula
##	9	3	5	12	20
##	Leontodon	Lepechinia	Lepidium	Leucanthemum	Lewisia
##	5	1	34	103	2
##	Limnanthes	Linaria	Linum	Lithodora	Lithophragma
##	10	2	6	1	6
##	Lithospermum	Lobularia	Lomatium	Lonicera	Lotus
##	28	1	64	66	34
##	Lunaria	Lupinus	Lychnis	Madia	Magnoliophyta
##	1	285	1	21	7
##	Magnoliopsida	Mahonia	Malus	Malva	Marah
##	7	16	32	1	12
##	Matricaria	Medicago	Melelotus	Melilotus	Melissa
##	4	10	10	41	2
##	Mentha	Mentzelia	Microseris	Mimulus	Monarda
##	13	4	19	29	24
##	Monardella	Myosotis	Narcissus	Nemophila	Nepeta
##	32	21	6	19	68
##	Net	Oenante	Oenothera	Origanum	Oxalis

##	13	1	4	30	14
##	Packera	Pastinaca	Penstemon	Perideridia	Perovskia
##	1	1	235	12	37
##	Petasites	Phacelia	Philadelphus	Phlox	Photinia
##	1	300	20	2	3
##	Physocarpus	Pieris	Plagiobothrys	Plantago	Plectritis
##	125	6	41	1	100
##	Poales	Poison	Polygonum	Potentilla	Prosartes
##	13	9	4	193	6
##	Prunella	Prunus	Pseudotsuga	Pseudoveronica	Purshia
##	9	149	15	1	31
##	Pyrus	Quercus	Ranunculus	Raphanus	Rheum
##	24	8	54	194	9
##	Rhododendron	Rhus	Ribes	Robinia	Romneya
##	88	65	300	9	1
##	Rosa	Rosaceae	Rosmarinus	Rubus	Rudbeckia
##	141	8	35	473	14
##	Salicaceae	Salix	Salvia	Sanicula	Sarcobatus
##	6	194	24	6	3
##	Scabiosa	Scandix	Scilla	Scutellaria	Sedum
##	2	61	10	4	18
##	Senecio	Sidalcea	Silene	Silphium	Sinapis
##	75	113	3	1	17
##	Sisymbrium	Sisyrinchium	Small	Solanum	Solidago
##	6	1	1	3	209
##	Sonchus	Sorbus	Sphaeralcea	Spiraea	Stachys
##	19	20	4	133	7
##	Stephanomeria	Styrax	Symphoricarpos	Symphyotrichum	Symphytum
##	1	21	87	59	8
##	Syringa	Tanacetum	Taraxacum	Taraxia	Tarweed
##	17	1	244	4	4
##	Tellima	Teucrium	Thelypodium	Thermopsis	Thymus
##	14	3	135	17	1
##	Tithonia	Tonella	Toxicoscordion	Tradescantia	Tragopogon
##	2	14	6	7	15
##	Trifolium	Triteleia	Umbelliferae	Vaccinium	Veratrum
##	111	9	2	58	2
##	Verbascum	Verbena	Verbenaceae	Veronica	Viburnum
##	3	33	6	42	13
##	Vicia	Viola	Weedy	Weigela	Whipplea
##	218	4	3	6	51
##	Wisteria	Wyethia	Yarrow	Zinnia	
##	21	88	2	1	

Now you have nice clean plant data to make networks out of, or more easily count up the number of plant species in an area.