

LA Tree data:
'treedata'

Packages

tidyverse (tidyr, dplyr, stringr, read/write_csv) – general data twerking
stringdist – fuzzy text matching
rnrf::osgb_parse() – converting BNG-format locations
rgdal, sf – spatial data handling incl. conversions
leaflet – displaying data to confirm correct locations

Identify categories
within relevant data
fields (base::unique)

Map existing categories
to new 'standard'
categories
(base::gsub, in a for loop)

Output:
category
mapping

Identify scientific name field, set to all caps,
drop punctuation, drop common 'add-ons'
(Sp, Spp, cv) base::gsub
stringr::str_squish to clean up whitespace

Reference
species list:
'specieslist'

Match (exact) species names in treedata to
specieslist. Separate matched from unmatched
records.

Output:
unmatched
records

Generate list of 'species' in unmatched records.
Fuzzy match to specieslist
stringdist::stringdistmatrix

Output:
First fuzzy
matches

Run Find-Replace on unmatched records

Copy to 'Find-
Replace file and
manually edit

Match to species list. Reject unmatched records

Output:
unmatched
records

Join all matched records. Convert co-ords to Lat
Long if reqd. Map records to check position

Create empty data
frame with data
standard fields

Map matched treedata to
data standard fields

Output: matched,
standardised
records