

hw_05.Rmd

Vinay K L

2023-11-01

Q1 Create a new column named `genus` and print out random 100 rows

Answer

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(neonDivData)
```

```
#attaching the data frame  
neonDivData::data_plant
```

```
## # A tibble: 1,034,093 x 26  
##   location_id siteID plotID unique_sample_id subplotID subplot_id subsubplot_id  
##   <chr>      <chr> <chr> <chr>          <chr>      <chr>      <chr>  
## 1 BART_006.b~ BART  BART_~ BART_006.basePl~ 31.4.1    31         4  
## 2 BART_006.b~ BART  BART_~ BART_006.basePl~ 31.4.1    31         4  
## 3 BART_006.b~ BART  BART_~ BART_006.basePl~ 41.1.1    41         1  
## 4 BART_006.b~ BART  BART_~ BART_006.basePl~ 41.4.1    41         4  
## 5 BART_006.b~ BART  BART_~ BART_006.basePl~ 41.4.1    41         4  
## 6 BART_006.b~ BART  BART_~ BART_006.basePl~ 41.4.1    41         4  
## 7 BART_006.b~ BART  BART_~ BART_006.basePl~ 32.2.1    32         2  
## 8 BART_006.b~ BART  BART_~ BART_006.basePl~ 41.1.1    41         1  
## 9 BART_006.b~ BART  BART_~ BART_006.basePl~ 32.4.1    32         4  
## 10 BART_006.b~ BART  BART_~ BART_006.basePl~ 32.4.1    32         4  
## # i 1,034,083 more rows
```

```
## # i 19 more variables: observation_datetime <dtm>, taxon_id <chr>,
## #   taxon_name <chr>, taxon_rank <chr>, variable_name <chr>, value <dbl>,
## #   unit <chr>, presence_absence <dbl>, boutNumber <chr>,
## #   nativeStatusCode <chr>, heightPlantOver300cm <chr>,
## #   heightPlantSpecies <chr>, release <chr>, sample_area_m2 <chr>,
## #   latitude <dbl>, longitude <dbl>, elevation <dbl>, plotType <chr>, ...
```

```
# saving the results on to a new data frame
```

```
data_plant <- data_plant %>%
```

```
  mutate(genus = sub("^\\s*([A-Za-z]+)\\s.*", "\\1", taxon_name))
```

```
set.seed(777) # setting the seed for reproducibility
```

```
random_100_genus <- sample(unique(data_plant$genus), 100)
```

```
print(random_100_genus)
```

```
##   [1] "Pinus"           "Smilax"           "Noccaea"           "Caperonia"
##   [5] "Lyonia"          "Abronia"          "Bassia"             "Schoenoplectus"
##   [9] "Cynoglossum"     "Lasthenia"        "Galium"             "Anthriscus"
##  [13] "Spermacoe"       "Packera"          "Polygonella"        "Amphiachyris"
##  [17] "Gamochaeta"      "Digitalis"        "Odontosoria"        "Sarcodes"
##  [21] "Pholistoma"      "Sideritis"        "Tripterocalyx"     "Manihot"
##  [25] "Zornia"          "Consolea"         "Sedella"           "Baptisia"
##  [29] "Diospyros"       "Myrsine"          "Streptanthella"    "Aralia"
##  [33] "Clermontia"      "Cupania"          "Diapensia"         "Lasiacis"
##  [37] "Physaria"        "Erythroxylum"   "Pentodon"          "Marrubium"
##  [41] "Callisia"        "Orbexilum"        "Nandina"           "Crumia"
##  [45] "Petiveria"       "Lachnanthes"      "Dichromanthus"     "Castanea"
##  [49] "Gaga"            "Sacciolepis"      "Clusia"            "Gayophytum"
##  [53] "Jamesia"         "Paronychia"       "Monotropa"         "Populus"
##  [57] "Colocasia"       "Lepuropetalon"    "Calylophus"        "Indigofera"
##  [61] "Gratiola"         "Machaonia"        "Trichomanes"       "Calycanthus"
##  [65] "Pseudotsuga"     "Muhlenbergia"     "Nama"              "Zinnia"
##  [69] "Nuttallanthus"    "Campsis"          "Erysimum"          "Garcinia"
##  [73] "Echinocereus"    "Cissus"           "Triticum"          "Ctenitis"
##  [77] "Bartonia"        "Striga"           "Phoenicaulis"      "Galactia"
##  [81] "Triodanis"       "Forsythia"        "Crotalaria"        "Prosartes"
##  [85] "Hybanthus"       "Sinapis"          "Paxistima"         "Pithecellobium"
##  [89] "Ionactis"        "Achlys"           "Dichondra"         "Stipulicida"
##  [93] "Chamaebatia"     "Laportea"         "Habenaria"         "Pomaria"
##  [97] "Oplonia"         "Deiregyne"        "Astelia"           "Onoclea"
```

Q2 Create a new column taxon_name2 and split the taxon_name entries to only first two names

Answer

```
#Since the data frame already exists - not calling again
```

```
data_plant <- data_plant %>%
```

```
  mutate(taxon_name2 = sub("^((\\S+\\s+\\S+).*)", "\\1", taxon_name)) #basic regex to split the taxon_name
```

```
# To print out the random 100 values
set.seed(777) # for reproducibility purpose
random_100_values <- sample(unique(data_plant$taxon_name2), 100)
print(random_100_values)
```

```
## [1] "Vulpia myuros" "Artemisia tilesii"
## [3] "Elephantopus elatus" "Lupinus argenteus"
## [5] "Tradescantia hirsutiflora" "Symphyotrichum pilosum"
## [7] "Castilleja pilosa" "Polygonum amphibium"
## [9] "Polygala cruciata" "Zea mays"
## [11] "Burmanna capitata" "Vaccinium tenellum"
## [13] "Geum spp." "Prenanthes alba"
## [15] "Evolvulus sericeus" "Ayenia sp."
## [17] "Salix sericea" "Oxycaryum sp."
## [19] "Brodiaea sp." "Oenothera acutissima"
## [21] "Chenopodium rubrum" "Brassica tournefortii"
## [23] "Opuntia dillenii" "Pinus jeffreyi"
## [25] "Freesia laxa" "Sideroxylon sp."
## [27] "Carex prairea" "Rubus nivalis"
## [29] "Abutilon grandifolium" "Pinus clausa"
## [31] "Lactuca biennis" "Bifora americana"
## [33] "Eugenia sp." "Trepocarpus aethusae"
## [35] "Calycocarpum lyonii" "Sporobolus heterolepis"
## [37] "Trillium grandiflorum" "Pedicularis labradorica"
## [39] "Allotropa virgata" "Opuntia polyacantha"
## [41] "Dicerandra linearifolia" "Asclepias feayi"
## [43] "Mentzelia albicaulis" "Andropogon sp."
## [45] "Ageratina jucunda" "Amsinckia menziesii"
## [47] "Smilax tamnoides" "Lepidium perfoliatum"
## [49] "Desmodium viridiflorum" "Drepanocladus crassicoatus"
## [51] "Cirsium undulatum" "Hymenopappus flavescens"
## [53] "Cinna latifolia" "Trifolium hybridum"
## [55] "Salix pseudomonticola" "Nymphaea odorata"
## [57] "Andropogon virginicus" "Sedum pulchellum"
## [59] "Nasturtium sp." "Symphyotrichum novae-angliae"
## [61] "Rhexia sp." "Ophioglossum nudicaule"
## [63] "Silphium laciniatum" "Polygonum hirsutum"
## [65] "Circaea xintermedia" "Rubus rosifolius"
## [67] "Rhodiola integrifolia" "Sphaeralcea incana"
## [69] "Vancouveria hexandra" "Sinapis alba"
## [71] "Amaranthus spinosus" "Salix fuscescens"
## [73] "Carphephorus odoratissimus" "Prunella vulgaris"
## [75] "Silphium sp." "Verbena halei"
## [77] "Rhynchospora chinensis" "Pseudognaphalium obtusifolium"
## [79] "Kyllinga odorata" "Abutilon theophrasti"
## [81] "Cunila origanoides" "Logfia sp."
## [83] "Anemone cylindrica" "Polygonum alpinum"
## [85] "Crepis acuminata" "Primula angustifolia"
## [87] "Chasmanthium sp." "Hibiscus sp."
## [89] "Rosa spp." "Lycium exsertum"
## [91] "Betula pumila" "Lupinus villosus"
## [93] "Oryzopsis asperifolia/Schizachne" "Cymopterus purpurascens"
## [95] "Gilia flavocincta" "Sida neomexicana"
```

```
## [97] "Alternanthera pungens"          "Hieracium scabrum"
## [99] "Odontosoria chinensis"          "Dichanthelium dichotomum"
```

Q3 Calculating species richness in three different plot size and then using `left_join` to combine the outputs.

```
# Splitting down the plots in to three groups named n_1, n_10, n_100

# Richness for 1m^2 plots
n_1 <- data_plant %>%
  filter(sample_area_m2 == "1") %>%
  group_by(siteID) %>%
  summarize(richness_1m2 = n_distinct(taxon_name2))

# Richness for 10m^2 plots
n_10 <- data_plant %>%
  filter(sample_area_m2 %in% c("1", "10")) %>%
  group_by(siteID) %>%
  summarize(richness_10m2 = n_distinct(taxon_name2))

# Richness for 100m^2 plots
n_100 <- data_plant %>%
  filter(sample_area_m2 %in% c("1", "10", "100")) %>%
  group_by(siteID) %>%
  summarize(richness_100m2 = n_distinct(taxon_name2))

# Joining all the `n_*` data frames

n_all <- n_1 %>%
  left_join(n_10) %>%
  left_join(n_100)
```

```
## Joining with 'by = join_by(siteID)'
## Joining with 'by = join_by(siteID)'
```

Q4 Transform `n_all` to long data format using `pivot_longer`

Answer

```
library(tidyr) # loading tidyr
```

```
n_all_long <- n_all %>% # appeneding to new dataframe
  pivot_longer(cols = starts_with("richness_"), # since richness has multiple names keeping just "richn
    names_to = "spatial_scale",
    values_to = "richness") %>%
  select(siteID, spatial_scale, richness) # transforming into three columns.
```

Q5 Creating plt using ggplot2

Answer

```
library(ggplot2) #loading the library
```

```
# Question has spatial scale v/s richness graph. So selecting those column to plot
ggplot(n_all_long, aes(x = spatial_scale, y = richness, group = siteID)) +
  geom_line(aes()) +
  geom_point(aes(), size = 3) +
  labs(
    x = "Spatial Scales",
    y = "Richness"
  )
```

