Projet stat pour données environnementales

Louis Delignac, Baptiste Gerbouin, Théo Lavandier, Alexandre Leys, Hamad Tria

28 février, 2024

Contents

```
oiseaux <- read.csv("data/Oiseaux_up_to_2023.csv", header = TRUE, sep = "\t")
LUP <- read.csv("data/LandUsePer_BM_2023_cartoISea.csv", header = TRUE)
```

```
# CREATE A NEW COLUMN IN THE OISEAUX DATAFRAME THAT CONTAINS THE LATIN NAME OF THE BIRD
# THE LATIN NAME IS THE FIRST NAME IN THE "Nom_Taxon_Cite" COLUMN
# IF THE NAME CONTAINS A "|", THE LATIN NAME IS THE FIRST NAME BEFORE THE "|"

# Split the names
my_split <- function(array, str = " \\| ") {
  out <- rep(NA, length(array))
  for (i in 1:length(array)) {
    out[i] <- unlist(strsplit(array[i], str))[1]
  }
  return (out)
}

# Test the function
only_latin <- my_split(as.vector(oiseaux$Nom_Taxon_Cite))
length(unique(only_latin)) == length(unique(oiseaux$Code_Ref))
```

```
## [1] TRUE
```

```
# Add the new column to the dataframe
oiseaux$latin <- only_latin
```

```
# THE MOST FREQUENT BIRD SPECIES OBSERVED IN THE DATASET
as.data.frame(sort(table(oiseaux$latin), decreasing = TRUE)[1:10])
```

Var1	Freq
Sylvia atricapilla	308
Parus major	300
Turdus merula	299
Columba palumbus	291
Phylloscopus collybita	251
Erithacus rubecula	230
Sturnus vulgaris	221
Troglodytes troglodytes	211
Fringilla coelebs	198
Certhia brachydactyla	169

```r
# FREQUENCY OF BIRD SPECIES OBSERVED IN THE DATASET BY YEAR
Annee <- my_split(as.vector(oiseaux$Date), str = "-")
oiseaux$Annee <- as.factor(Annee)
as.data.frame.matrix(table(oiseaux$latin, oiseaux$Annee))
```

	2018	2019	2022	2023
Acrocephalus schoenobaenus	1	2	3	4
Acrocephalus scirpaceus	3	5	1	2
Actitis hypoleucos	1	0	0	1
Aegithalos caudatus	26	38	39	39
Alcedo atthis	4	7	9	6
Anas platyrhynchos	9	10	5	19
Anthus pratensis	4	1	2	5
Anthus trivialis	6	2	2	3
Ardea cinerea	4	4	1	2
Bubulcus ibis	0	1	0	0
Buteo buteo	3	0	0	0
Caprimulgus europaeus	0	0	0	1
Carduelis cannabina	0	3	2	3
Carduelis carduelis	20	13	17	15
Carduelis chloris	18	17	18	15
Carduelis spinus	1	0	0	0
Casmerodius albus	0	1	0	0
Certhia brachydactyla	46	42	41	40
Cettia cetti	42	36	25	27
Ciconia ciconia	0	1	1	1
Circus pygargus	0	1	0	0
Cisticola juncidis	18	16	16	20
Columba livia f. domestica	0	3	20	25
Columba palumbus	69	64	85	73
Corvus corone	36	29	36	48
Cuculus canorus	11	10	0	1
Cyanistes caeruleus	54	27	43	39
Cygnus olor	1	2	0	1
Dendrocopos major	19	25	22	18
Dendrocopos minor	3	1	5	4
Dryocopus martius	0	0	1	0
Egretta garzetta	1	2	1	1
Emberiza calandra	0	0	0	1
Emberiza cirlus	6	7	4	0
Emberiza schoeniclus	1	0	0	0
Erithacus rubecula	53	52	64	61
Falco tinnunculus	1	4	0	0
Ficedula hypoleuca	0	1	0	3
Fringilla coelebs	62	63	38	35
Fulica atra	0	5	4	2
Gallinula chloropus	14	12	12	13
Garrulus glandarius	23	29	20	22
Himantopus himantopus	0	0	0	1
Hippolais polyglotta	16	31	18	13
Hirundo rustica	3	0	0	0
Jynx torquilla	2	0	0	1
Lanius collurio	1	0	0	1

	2018	2019	2022	2023
Lanius senator	0	0	1	0
Locustella luscinioides	0	1	1	2
Locustella naevia	1	0	1	0
Lophophanes cristatus	3	12	5	9
Luscinia megarhynchos	39	48	36	23
Luscinia svecica	2	0	0	0
Milvus migrans	19	18	1	0
Motacilla alba	12	8	9	11
Motacilla cinerea	1	2	2	4
Motacilla flava	2	2	3	4
Muscicapa striata	1	4	1	2
Oenanthe oenanthe	0	1	1	0
Oriolus oriolus	9	6	8	10
Pandion haliaetus	0	1	0	0
Parus major	81	77	69	73
Passer domesticus	19	23	42	38
Pernis apivorus	1	0	0	0
Phalacrocorax carbo	3	5	1	3
Phasianus colchicus	2	1	0	2
Phoenicurus ochruros	8	7	11	15
Phoenicurus phoenicurus	3	5	6	6
Phylloscopus bonelli	8	12	11	8
Phylloscopus collybita	70	67	63	51
Phylloscopus trochilus	2	6	4	8
Pica pica	28	36	48	52
Picus viridis	30	33	39	34
Podiceps cristatus	0	2	1	0
Poecile palustris	1	0	3	2
Prunella modularis	37	25	25	23
Pyrrhula pyrrhula	0	1	1	2
Regulus ignicapilla	25	28	35	35
Saxicola rubicola	8	10	7	8
Serinus serinus	24	25	19	17
Sitta europaea	34	21	33	42
Streptopelia decaocto	19	31	31	35
Streptopelia turtur	2	4	4	1
Sturnus vulgaris	55	42	64	60
Sylvia atricapilla	71	81	82	74
Sylvia communis	3	8	5	7
Sylvia undata	2	5	1	2
Tringa ochropus	0	0	2	1
Troglodytes troglodytes	70	45	51	45
Turdus merula	73	67	84	75
Turdus philomelos	31	37	36	28
Turdus viscivorus	0	1	1	2
Upupa epops	2	2	1	2

```
# CREATE A NEW COLUMN IN THE OISEAUX DATAFRAME THAT CONTAINS THE MOS11 VALUE OF THE POINT
filter <- LUP$BufferSize == 500
LUP_500_MOS11 <- LUP[filter,c("Geometry", "ID", "X", "Y", "BufferSize", "MOS11")]
rownames(LUP_500_MOS11) <- 1:nrow(LUP_500_MOS11)
LUP_500_MOS11[, c("ID","MOS11")]
```

ID	MOS11
E4190N64415	0.0076825
E4210N64410	0.1774112
E4215N64390	0.0088472
E4225N64375	0.0178249
E4045N64355	0.0133477
E4050N64355	0.0475924
E4030N64350	0.0131716
E4050N64350	0.0036325
E4250N64350	0.0426183
E4170N64345	0.0423233
E4225N64340	0.1158542
E4255N64340	0.0422937
E4140N64335	0.1338980
E4145N64335	0.2441183
E4265N64330	0.1742767
E4135N64325	0.1739114
E4160N64325	0.0039921
E4115N64320	0.0501003
E4000N64315	0.0017596
E3990N64305	0.0236148
E4015N64305	0.0140282
E3985N64300	0.0191608
E4000N64290	0.0236132
E4120N64290	0.0934285
E4140N64285	0.0533292
E4035N64270	0.0427638
E4230N64265	0.3302321
E4235N64240	0.2926001
E4025N64235	0.1292651
E4150N64235	0.4972268
E4185N64225	0.5799538
E4130N64215	0.4054738
E4210N64210	0.3637253
E4085N64200	0.3119369
E4090N64185	0.1033095
E4145N64185	0.3618082
E4170N64185	0.6401005
E4215N64185	0.1955214
E4235N64185	0.1084737
E4240N64185	0.0568741
E4200N64180	0.4274846
E4065N64165	0.1709362
E4155N64165	0.2765180
E4185N64160	0.2068616
E4160N64150	0.4098935
E4145N64140	0.2678870
E4155N64140	0.3150761
E4205N64140	0.2221061
E4120N64135	0.2265947
E4135N64135	0.2193205

ID	MOS11
E4180N64130	0.4304017
E4180N64125	0.1991207
E4175N64430	0.0062400

```
MOS11 <- rep(NA, nrow(oiseaux))
for (i in 1:nrow(oiseaux)) {
  MOS11[i] <- which(oiseaux$Code_Maille[i] == LUP_500_MOS11$ID)
}

# Add the new column to the dataframe
oiseaux$MOS11 <- LUP_500_MOS11$MOS11[MOS11]
```

```
# THE NUMBER OF BIRDS OBSERVED IN EACH MOS11 AREA BY YEAR
data <- oiseaux %>%
  group_by(Code_Maille, Annee, MOS11) %>%
  select(Denombrement_min) %>%
  summarise(sum = sum(Denombrement_min, na.rm = TRUE))
```

```
## Adding missing grouping variables: `Code_Maille`, `Annee`, `MOS11`
## `summarise()` has grouped output by 'Code_Maille', 'Annee'. You can override
## using the `.groups` argument.
```

```
ggplot(data, aes(x=MOS11, y=sum, color=Annee)) +
      geom_point(size=2)
```