Cleaning and exploration of Belgian Coccinellidae GBIF dataset

Gilles San Martin

01 November 2016

Contents

Belgian Ladybirds dataset

The GBIF__data/Coccinellidae__Belgium directory contains several files based on an original dataset publicly available on the GBIF portal (here). All the additional files have been created with the R script `GBIFdata_CreateData.R` and the pdf report with the same name describes the process and the content of the dataset.

`Belgian_Coccinellidae.csv`
Original dataset from GBIF.

`UTM5_Coccinellidae_long.csv`
Cleaned dataset with subset of the columns of the original dataset and data more recent than 1980, identified up to species level and with a geographic coordinates precision <4000m.
The X Y projected coordinates have been added (Belgian Lambert 1972) along with the corresponding 5 km² UTM grid square code (column "MGRS"). A species code (column "spcode") has also been created as the 3 first letters of genus name and 3 first letters of the species name. The correspondance with the real species names are provided in the `taxlib_Coccinellidae.csv` dataset.
The other columns are standard GBIF columns.

`UTM5_Coccinellidae.csv`
This dataset is based on `UTM5_Coccinellidae_long.csv` and provides the MGRS code on the lines and the species codes as columns (MGRS x spcode crosstable).
The numbers are the number of data for each species and each MGRS grid square, i.e. the number of lines in the dataset that corresponds normaly to one species on one date in one location and by one observer.

`UTM5_Coccinellidae_sampling_effort.csv`
This dataset provides additional information about the sampling effort on each 5 km² MGRS grid squares (on lines). The column `min1sp` provides the number of visits (different dates) with at least 1 species observed for each 5 km² MGRS grid squares. The column `min5sp` provides the number of visits (different dates) with at least 5 species observed for each 5 km² MGRS grid squares. etc...

`taxlib_Coccinellidae.csv`
Provides a taxonomic list and the corresponding species codes created.

Import the data and load packages

Define the working directory. "GISfolder" is the place where the spatial data are stored (typically on an extrenal harddrive). `source` allows you to silently execute an R script and put all its objects in the memory (here several useful functions)

```
setwd("/home/gilles/stats/Formation_R_stats/UCL_LBOE2121/GBIF")
GISfolder <- "/home/gilles/stats/Formation_R_stats/UCL_LBOE2121/GBIF/data/Spatial"
source("/home/gilles/stats/mytoolbox.R")

library(sp)
library(rgdal)
library(rgeos)
library(raster)
library(reshape2) # for dcast and melt functions
```

Import data and remove unuseful columns. Note that in the `read.table` function `quote = ""` is necessary to avoid imports problems due to end of line characters in the midlle of a character chain.

```
d <- read.table("data/GBIF_data/Coccinellidae_Belgium/Belgian_Coccinellidae.csv",
                sep = "\t", dec = ".", header = TRUE, encoding = "latin1", quote = "")

# vector of potentially interesting variables names
varnames <- c("family", "genus", "species", "infraspecificepithet",
              "taxonrank", "locality","decimallatitude", "decimallongitude",
```

```
              "coordinateuncertaintyinmeters",
              "day", "month", "year")

d <- d[,varnames] # keep only these variables
dim(d)
```

```
## [1] 72185    12
```

```
summary(d)
```

```
##           family             genus                                  species
##  Coccinellidae:72185   Coccinella:16558   Coccinella septempunctata     :12989
##                        Adalia    :10411   Propylaea quatuordecimpunctata: 7543
##                        Harmonia  : 8498   Harmonia axyridis             : 7400
##                        Propylaea : 7543   Adalia bipunctata             : 7135
##                        Psyllobora: 4171   Psyllobora vigintiduopunctata : 4171
##                        Calvia    : 4165   Adalia decempunctata          : 3268
##                        (Other)   :20839   (Other)                       :29679
##  infraspecificepithet      taxonrank     locality       decimallatitude decimallongitude
##            :72183     GENUS     :  680   Mode:logical   Min.   :49.51   Min.   :2.536
##  apetzoides:    2     SPECIES   :71503   NA's:72185     1st Qu.:50.49   1st Qu.:4.102
##                       SUBSPECIES:    2                  Median :50.80   Median :4.578
##                                                         Mean   :50.74   Mean   :4.571
##                                                         3rd Qu.:51.01   3rd Qu.:5.149
##                                                         Max.   :51.50   Max.   :6.364
##
##  coordinateuncertaintyinmeters      day           month            year
##  Min.   :   5.0                Min.   : 1.00   Min.   : 1.000   Min.   :1811
##  1st Qu.: 707.1                1st Qu.: 8.00   1st Qu.: 5.000   1st Qu.:1991
##  Median : 707.1                Median :15.00   Median : 6.000   Median :2003
##  Mean   :1115.5                Mean   :15.38   Mean   : 6.461   Mean   :1989
##  3rd Qu.: 707.1                3rd Qu.:23.00   3rd Qu.: 8.000   3rd Qu.:2006
##  Max.   :7071.0                Max.   :31.00   Max.   :12.000   Max.   :2011
##                                NA's   :3925    NA's   :3925     NA's   :3924
```

```
head(d)
```

```
##          family      genus                        species infraspecificepithet taxonrank locality
## 1 Coccinellidae   Harmonia              Harmonia axyridis                        SPECIES       NA
## 2 Coccinellidae     Adalia           Adalia decempunctata                        SPECIES       NA
## 3 Coccinellidae Coccinella      Coccinella septempunctata                        SPECIES       NA
## 4 Coccinellidae     Scymnus                                                        GENUS       NA
## 5 Coccinellidae   Harmonia              Harmonia axyridis                        SPECIES       NA
## 6 Coccinellidae  Propylaea Propylaea quatuordecimpunctata                        SPECIES       NA
##   decimallatitude decimallongitude coordinateuncertaintyinmeters day month year
## 1          50.807            4.379                         70.71  11     6 2008
## 2          50.066            4.557                         70.71  12     4 2007
## 3          50.094            4.518                       5000.00  12     8 2004
## 4          50.079            4.636                        100.00  31     3 2011
## 5          50.723            3.839                        999.00   4     9 2010
## 6          49.656            5.678                       1000.00   6     6 2004
```

Data exploration

Distribution of the precision of the estimates. Most of the data have originally been encoded as $1km^2$ data (precision 707.1m = sqrt(2 * 500^2)) or $5km^2$ data (precision 3536m = sqrt(2 * 2500^2))

```r
table(d$coordinateuncertaintyinmeters)
```

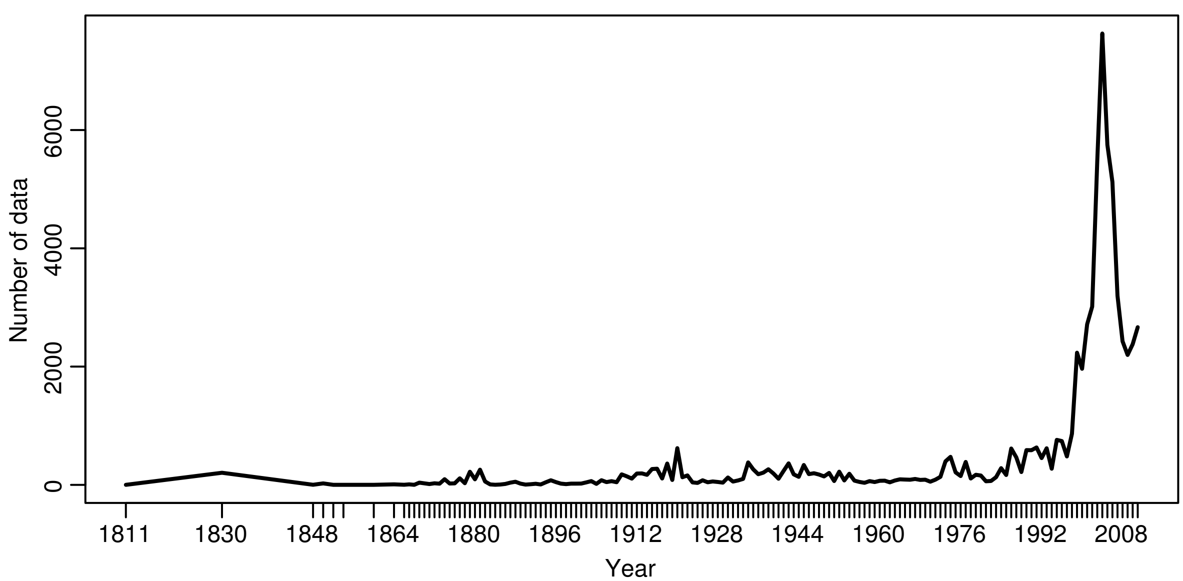
```
##
##     5    10    15    20    25    30    50 70.71   100   150   200   250   300   400   500   700
##     1   785     1   165     1    33    63  6678  1073     5    45    10     2     1     4     1
## 707.1   999  1000  3536  5000  7071
## 48037  1200  2138 11734   171    37
```

The number of data begins to rise at the end of the 90ies

```r
yearcounts <- table(d$year)
yearcounts
```

```
##
## 1811 1830 1848 1850 1852 1854 1860 1864 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877
##    1  205    2   25    1    1    1    9    2    9    1   38   26   14   27   20   96   23   26  111
## 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897
##   28  222   94  257   57    8    2    6   16   38   53   21    5   11   19    7   44   78   47   21
## 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917
##   13   21   21   22   42   64   17   79   47   63   45  178  145  106  192  194  170  268  273  108
## 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937
##  361   83  621  129  160   42   33   78   44   58   49   38  125   56   75  101  380  263  180  207
## 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957
##  264  192  104  236  364  178  136  336  182  197  175  143  202   65  223   75  188   72   50   35
## 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977
##   64   48   69   71   43   74   94   90   87  100   84   88   54   88  136  398  474  212  149  388
## 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
##  108  171  158   61   66  130  283  167  615  465  217  588  586  633  453  619  273  761  741  481
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##  865 2235 1963 2715 3014 5471 7632 5750 5121 3199 2428 2197 2378 2666
```

```r
# dev.new(width = 16/2.54, height = 8/2.54)
par(mar = c(3, 3, 1,1), mgp = c(1.8, 0.6, 0), cex = 0.75)
plot(yearcounts, type = "l", xlab = "Year", ylab = "Number of data", las = 0)
```



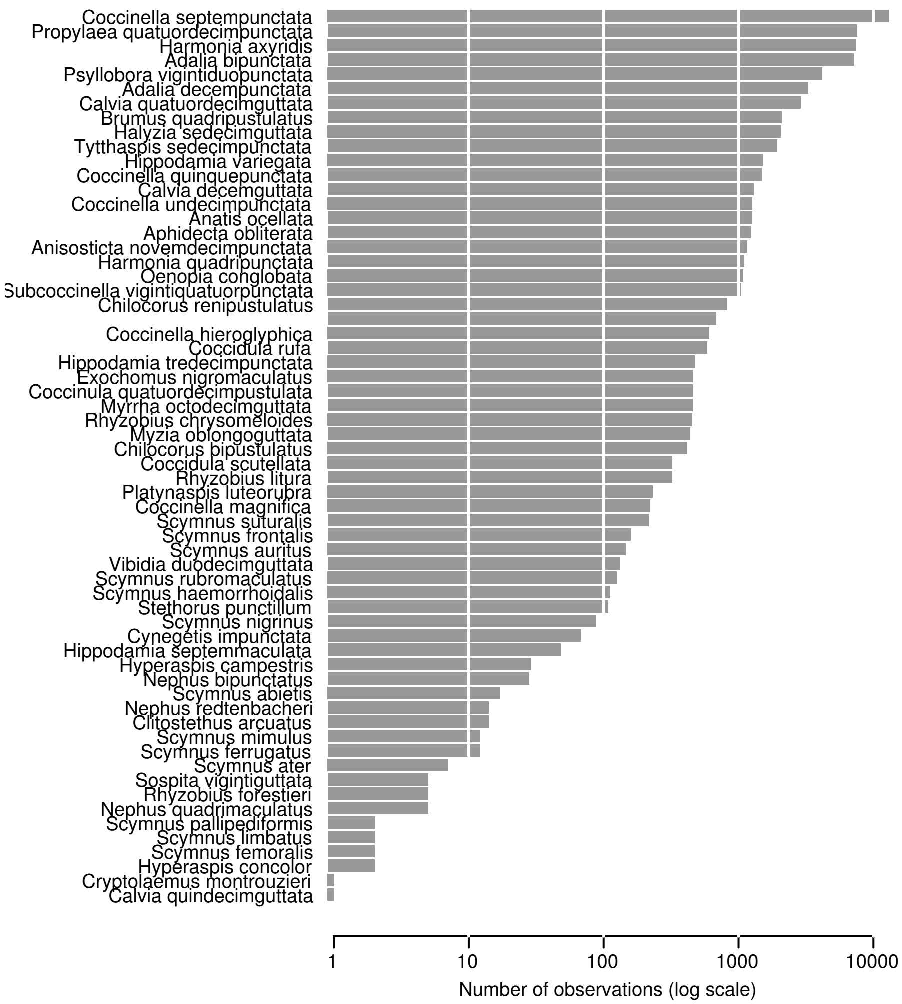
Check the number of data per species.
The species without identification correspond to specimens identified up to genus level.

```r
spcounts <- sort(table(d$species))
spcounts
```

```
## 
##                 Calvia quindecimguttata            Cryptolaemus montrouzieri 
##                                       1                                    1 
##                      Hyperaspis concolor                    Scymnus femoralis 
##                                       2                                    2 
##                        Scymnus limbatus            Scymnus pallipediformis 
##                                       2                                    2 
##                   Nephus quadrimaculatus                 Rhyzobius forestieri 
##                                       5                                    5 
##                   Sospita vigintiguttata                         Scymnus ater 
##                                       5                                    7 
##                      Scymnus ferrugatus                       Scymnus mimulus 
##                                      12                                   12 
##                   Clitostethus arcuatus                 Nephus redtenbacheri 
##                                      14                                   14 
##                         Scymnus abietis                   Nephus bipunctatus 
##                                      17                                   28 
##                   Hyperaspis campestris            Hippodamia septemmaculata 
##                                      29                                   48 
##                    Cynegetis impunctata                     Scymnus nigrinus 
##                                      68                                   87 
##                    Stethorus punctillum            Scymnus haemorrhoidalis 
##                                     108                                  111 
##                  Scymnus rubromaculatus            Vibidia duodecimguttata 
##                                     125                                  131 
##                         Scymnus auritus                    Scymnus frontalis 
##                                     146                                  158 
##                      Scymnus suturalis                  Coccinella magnifica 
##                                     217                                  222 
##                  Platynaspis luteorubra                     Rhyzobius litura 
##                                     231                                  321 
##                    Coccidula scutellata            Chilocorus bipustulatus 
##                                     323                                  417 
##                    Myzia oblongoguttata           Rhyzobius chrysomeloides 
##                                     437                                  454 
##                 Myrrha octodecimguttata      Coccinula quatuordecimpustulata 
##                                     456                                  459 
##                Exochomus nigromaculatus           Hippodamia tredecimpunctata 
##                                     462                                  474 
##                          Coccidula rufa             Coccinella hieroglyphica 
##                                     584                                  607 
##                                                  Chilocorus renipustulatus 
##                                     680                                  820 
## Subcoccinella vigintiquatuorpunctata                     Oenopia conglobata 
##                                    1048                                 1084 
##                 Harmonia quadripunctata      Anisosticta novemdecimpunctata 
##                                    1097                                 1161 
##                    Aphidecta obliterata                      Anatis ocellata 
##                                    1224                                 1259 
##               Coccinella undecimpunctata                  Calvia decemguttata 
##                                    1263                                 1290 
##               Coccinella quinquepunctata                  Hippodamia variegata 
##                                    1477                                 1512 
##               Tytthaspis sedecimpunctata                 Halyzia sedecimguttata
```

```
##                                1926                              2074
##            Brumus quadripustulatus        Calvia quatuordecimguttata
##                                2086                              2874
##               Adalia decempunctata     Psyllobora vigintiduopunctata
##                                3268                              4171
##                  Adalia bipunctata                 Harmonia axyridis
##                                7135                              7400
##      Propylaea quatuordecimpunctata         Coccinella septempunctata
##                                7543                             12989
```

```r
# dev.new(width = 12/2.54, height = 16/2.54)
par(mar = c(4, 14, 0.1,1), mgp = c(1.8, 0.6, 0), cex = 0.6)
barplot(spcounts, horiz = TRUE, las = 1, log = "x", border = NA, col = "Gray60",
        xlab = "Number of observations (log scale)")
abline(v = 10^c(1:5), col = "white", lwd = 1.5)
```



Data cleaning

Remove unusefull data and create new variables

Keep only the data with coordinates precision compatible with 5km² UTM squares

```
d <- d[d$coordinateuncertaintyinmeters < 4000,]
```

Keep only the data more recent than 1980

```
d <- d[d$year >= 1980, ]
```

Add the date and remove data without date (NA)

```
d$date <- as.Date(paste(d$year, d$month, d$day, sep = "-"))
d <- d[!is.na(d$date),]
```

Remove the data without species identification

```
d <- d[d$species != "", ]
```

Change the factor level order for the species to have the species ordered by decreasing number of observations

```
spcounts <- sort(table(d$species))
d$species <- factor(d$species, levels = rev(names(spcounts)))
```

Create shorter species code with the 3 first letters of the genus name and the 3 first letters of the species name. Then order levels in a similar way as for the full species names.

```
d$spcode <- tolower(paste(substring(d$genus, 1, 3),
                          gsub(".* ([a-z]{3}).*", "\\1", d$species), sep = ""))
spcounts <- sort(table(d$spcode))
d$spcode <- factor(d$spcode, levels = rev(names(spcounts)))
```

Check that the new species code are unique. The unique spcode should have the same length as the number of rows of the full taxa names.

```
taxlib <- unique(d[, c("family", "genus", "species", "spcode")])
taxlib <- taxlib[order(as.character(taxlib$species)),]
nrow(taxlib)
```

```
## [1] 56
```

```
length(unique(taxlib$spcode))
```

```
## [1] 56
```

```
pander(taxlib) # print the table
```

	family	genus	species	spcode
44	Coccinellidae	Adalia	Adalia bipunctata	adabip
2	Coccinellidae	Adalia	Adalia decempunctata	adadec

	family	genus	species	spcode
372	Coccinellidae	Anatis	Anatis ocellata	anaoce
20	Coccinellidae	Anisosticta	Anisosticta novemdecimpunctata	aninov
94	Coccinellidae	Aphidecta	Aphidecta obliterata	aphobl
137	Coccinellidae	Brumus	Brumus quadripustulatus	bruqua
58	Coccinellidae	Calvia	Calvia decemguttata	caldec
21	Coccinellidae	Calvia	Calvia quatuordecimguttata	calqua
271	Coccinellidae	Chilocorus	Chilocorus bipustulatus	chibip
18	Coccinellidae	Chilocorus	Chilocorus renipustulatus	chiren
732	Coccinellidae	Clitostethus	Clitostethus arcuatus	cliarc
449	Coccinellidae	Coccidula	Coccidula rufa	cocruf
210	Coccinellidae	Coccidula	Coccidula scutellata	cocscu
788	Coccinellidae	Coccinella	Coccinella hieroglyphica	cochie
939	Coccinellidae	Coccinella	Coccinella magnifica	cocmag
68	Coccinellidae	Coccinella	Coccinella quinquepunctata	cocqui
11	Coccinellidae	Coccinella	Coccinella septempunctata	cocsep
91	Coccinellidae	Coccinella	Coccinella undecimpunctata	cocund
525	Coccinellidae	Coccinula	Coccinula quatuordecimpustulata	cocqua
4669	Coccinellidae	Cynegetis	Cynegetis impunctata	cynimp
155	Coccinellidae	Exochomus	Exochomus nigromaculatus	exonig
13	Coccinellidae	Halyzia	Halyzia sedecimguttata	halsed
1	Coccinellidae	Harmonia	Harmonia axyridis	haraxy
29	Coccinellidae	Harmonia	Harmonia quadripunctata	harqua
2802	Coccinellidae	Hippodamia	Hippodamia septemmaculata	hipsep
169	Coccinellidae	Hippodamia	Hippodamia tredecimpunctata	hiptre
59	Coccinellidae	Hippodamia	Hippodamia variegata	hipvar
3155	Coccinellidae	Hyperaspis	Hyperaspis campestris	hypcam
18080	Coccinellidae	Hyperaspis	Hyperaspis concolor	hypcon
31	Coccinellidae	Myrrha	Myrrha octodecimguttata	myroct
256	Coccinellidae	Myzia	Myzia oblongoguttata	myzobl
17994	Coccinellidae	Nephus	Nephus bipunctatus	nepbip
24244	Coccinellidae	Nephus	Nephus quadrimaculatus	nepqua
123	Coccinellidae	Nephus	Nephus redtenbacheri	nepred
36	Coccinellidae	Oenopia	Oenopia conglobata	oencon
569	Coccinellidae	Platynaspis	Platynaspis luteorubra	plalut
6	Coccinellidae	Propylaea	Propylaea quatuordecimpunctata	proqua
16	Coccinellidae	Psyllobora	Psyllobora vigintiduopunctata	psyvig
39	Coccinellidae	Rhyzobius	Rhyzobius chrysomeloides	rhychr
10307	Coccinellidae	Rhyzobius	Rhyzobius forestieri	rhyfor
161	Coccinellidae	Rhyzobius	Rhyzobius litura	rhylit
3020	Coccinellidae	Scymnus	Scymnus abietis	scyabi
8429	Coccinellidae	Scymnus	Scymnus auritus	scyaur
6361	Coccinellidae	Scymnus	Scymnus ferrugatus	scyfer
83	Coccinellidae	Scymnus	Scymnus frontalis	scyfro
1043	Coccinellidae	Scymnus	Scymnus haemorrhoidalis	scyhae
31724	Coccinellidae	Scymnus	Scymnus limbatus	scylim
9278	Coccinellidae	Scymnus	Scymnus mimulus	scymim
9044	Coccinellidae	Scymnus	Scymnus nigrinus	scynig
3139	Coccinellidae	Scymnus	Scymnus pallipediformis	scypal
190	Coccinellidae	Scymnus	Scymnus rubromaculatus	scyrub
604	Coccinellidae	Scymnus	Scymnus suturalis	scysut
1460	Coccinellidae	Stethorus	Stethorus punctillum	stepun
25	Coccinellidae	Subcoccinella	Subcoccinella vigintiquatuorpunctata	subvig
66	Coccinellidae	Tytthaspis	Tytthaspis sedecimpunctata	tytsed
23	Coccinellidae	Vibidia	Vibidia duodecimguttata	vibduo

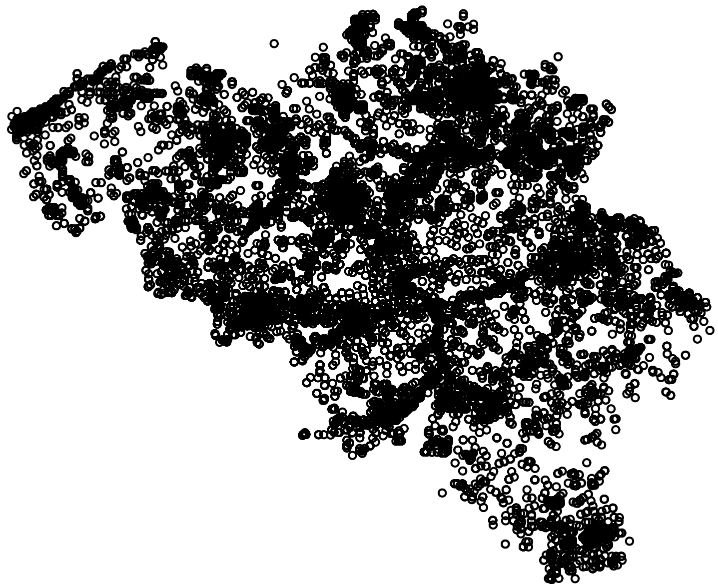
Intersect with UTM 5 grid squares

Project the long lat coordinates in Belgian Lambert 1972 coordinate reference system (in meters)

```
tmp <- d[, c("decimallatitude",  "decimallongitude")] # copy of the coordinates
tmp$ID <- 1:nrow(tmp)
coordinates(tmp) <- ~ decimallongitude + decimallatitude # transform into a spatial object
proj4string(tmp) <- CRS("+proj=longlat +datum=WGS84") # specify the coordinates reference
xy <- spTransform(tmp, CRS("+init=epsg:31370")) # Projection toward Belgian Lambert
```

grahical check

```
par(mar = c(0,0,0,0), mgp = c(1.8, 0.6, 0), cex = 0.5)
plot(xy, pch = 1)
```



Read UTM geodata and intersect the points with the UTM (~1.5 min)

```
utm <- readOGR(GISfolder, "UTMBEL05_Poly", p4s = "+init=epsg:31370", verbose = FALSE)
xy <- intersect(xy, utm)
gc() # clean the memory
```

```
##           used (Mb) gc trigger   (Mb)  max used   (Mb)
## Ncells 1247022 66.6    2164898  115.7   2164898  115.7
## Vcells 2286383 17.5  312907972 2387.3 302318669 2306.6
```

```
xy@data$x <- coordinates(xy)[,1]
xy@data$y <- coordinates(xy)[,2]
xy <- xy@data
colnames(xy)[2] <- "MGRS"

d$ID <- 1:nrow(d)
d <- merge(d, xy, by = "ID", all.x = TRUE)
summary(d)
```

```
##         ID                    family              genus                                      species
```

```
##  Min.   :    1   Coccinellidae:54246   Coccinella:13776   Coccinella septempunctata     :11530
##  1st Qu.:13562                         Harmonia  : 8079   Harmonia axyridis             : 7238
##  Median :27124                         Adalia    : 7120   Propylaea quatuordecimpunctata: 5538
##  Mean   :27124                         Propylaea : 5538   Adalia bipunctata             : 5423
##  3rd Qu.:40685                         Calvia    : 3682   Psyllobora vigintiduopunctata : 3496
##  Max.   :54246                         Psyllobora: 3496   Calvia quatuordecimguttata    : 2447
##                                        (Other)   :12555   (Other)                       :18574
##  infraspecificepithet      taxonrank     locality       decimallatitude decimallongitude
##            :54244     GENUS     :    0   Mode:logical   Min.   :49.51   Min.   :2.536
##  apetzoides:    2     SPECIES   :54244   NA's:54246     1st Qu.:50.49   1st Qu.:4.058
##                       SUBSPECIES:    2                  Median :50.81   Median :4.639
##                                                         Mean   :50.73   Mean   :4.551
##                                                         3rd Qu.:51.01   3rd Qu.:5.136
##                                                         Max.   :51.50   Max.   :6.364
##
##  coordinateuncertaintyinmeters      day            month             year           date
##  Min.   :   5.0                Min.   : 1.00   Min.   : 1.000   Min.   :1980   Min.   :1980-03-10
##  1st Qu.: 707.1                1st Qu.: 8.00   1st Qu.: 5.000   1st Qu.:2001   1st Qu.:2001-07-04
##  Median : 707.1                Median :15.00   Median : 7.000   Median :2004   Median :2004-07-22
##  Mean   : 680.8                Mean   :15.46   Mean   : 6.498   Mean   :2003   Mean   :2003-05-07
##  3rd Qu.: 707.1                3rd Qu.:23.00   3rd Qu.: 8.000   3rd Qu.:2006   3rd Qu.:2006-08-30
##  Max.   :3536.0                Max.   :31.00   Max.   :12.000   Max.   :2011   Max.   :2011-11-10
##
##     spcode              MGRS             x                y
##  cocsep :11530   31UFS225025: 1655   Min.   : 21547   Min.   : 22622
##  haraxy : 7238   31UES775475: 1305   1st Qu.:128112   1st Qu.:131110
##  proqua : 5538   31UES075475:  986   Median :169112   Median :167283
##  adabip : 5423   31UES275325:  878   Mean   :162999   Mean   :158509
##  psyvig : 3496   31UFS275725:  719   3rd Qu.:204178   3rd Qu.:189035
##  calqua : 2447   (Other)    :48692   Max.   :291940   Max.   :243027
##  (Other):18574   NA's       :   11   NA's   :11       NA's   :11
```

Create a crosstable with one UTM square on each line and the ladybirds species code as columns. The values are the numer of "data" (date x species x location)

```
UTMladybirds <- dcast(d, MGRS ~ spcode, fun = length)

# 10 first lines and 12 first columns
UTMladybirds[1:10, 1:12]
```

```
##           MGRS cocsep haraxy proqua adabip psyvig calqua halsed bruqua adadec tytsed caldec
## 1  31UDS675525      2      4      0      2      0      0      0      0      0      0      0
## 2  31UDS675575     17      2      0      1      4      0      1      0      1      0      0
## 3  31UDS675625     24      3      3      5     10      2      0      3      1      1      1
## 4  31UDS725425      2      1      0      1      0      0      0      0      0      0      0
## 5  31UDS725475      1      1      0      0      0      0      0      0      0      0      0
## 6  31UDS725525      1      0      0      1      0      0      0      1      0      0      0
## 7  31UDS725575    251    114      6     17      8      1     74      1      7      3      1
## 8  31UDS725625    116     60     19     26     29      5     33      3     14      4      0
## 9  31UDS775325      1      2      1      0      0      1      0      1      0      0      0
## 10 31UDS775375      5      5     13      6      1      2      2      1      1      0      0
```

Save the datasets on the disc

```
write.csv2(UTMladybirds, "data/GBIF_data/Coccinellidae_Belgium/UTM5_Coccinellidae.csv",
           row.names = FALSE)
```

```r
write.csv2(d, "data/GBIF_data/Coccinellidae_Belgium/UTM5_Coccinellidae_long.csv",
           row.names = FALSE)
write.csv2(taxlib, "data/GBIF_data/Coccinellidae_Belgium/taxlib_Coccinellidae.csv",
           row.names = FALSE)
```

Evaluate sampling effort

```
# Unique list of species observed at a given date on a given UTM square
tmp <- unique(d[, c("MGRS", "spcode", "date")])
# Count the number of species observed at each date
tmp <- aggregate(list(nbsp = tmp$spcode), tmp[,c("MGRS", "date")], FUN = length)
# Make a cross table with UTM squares as lines and as colums the number of
# different dates ("visits") with exactly 1 species observed, 2 species observed, etc...
tmp <- dcast(tmp, MGRS ~ nbsp, fun = length)
colnames(tmp)[-1] <- paste0("min",colnames(tmp)[-1], "sp")

# Make the cumulative sum to obtain the number of vists with at least 1 species,
# number of visits with at least 2 species,...
tmp <- data.frame(
    MGRS = tmp$MGRS,
    t(apply(tmp[,-1], 1, function(x) rev(cumsum(rev(x)))))
    )
samplingEffort <- tmp
```

The first line of this table looks as follows. So the 31UDS675575 square (second line) has been visited at 17 different dates. At 6 dates at least 2 species were observed and only once, 4 species were observed.

```
head(tmp)
```

```
##          MGRS min1sp min2sp min3sp min4sp min5sp min6sp min7sp min8sp min9sp min10sp min11sp
## 1 31UDS675525      2      2      1      0      0      0      0      0      0       0       0
## 2 31UDS675575     17      6      1      1      0      0      0      0      0       0       0
## 3 31UDS675625     28      9      6      4      3      3      3      2      2       1       1
## 4 31UDS725425      2      1      0      0      0      0      0      0      0       0       0
## 5 31UDS725475      1      1      0      0      0      0      0      0      0       0       0
## 6 31UDS725525      3      0      0      0      0      0      0      0      0       0       0
##   min12sp min13sp min14sp min15sp min16sp min17sp min18sp min19sp min20sp min21sp min22sp min26sp
## 1       0       0       0       0       0       0       0       0       0       0       0       0
## 2       0       0       0       0       0       0       0       0       0       0       0       0
## 3       1       1       1       0       0       0       0       0       0       0       0       0
## 4       0       0       0       0       0       0       0       0       0       0       0       0
## 5       0       0       0       0       0       0       0       0       0       0       0       0
## 6       0       0       0       0       0       0       0       0       0       0       0       0
```

We cans see how many squares are remaining if you remove the squares with the lowest sampling effort with different thresholds.

There are for example 554 squares with at lest 1 visit on which at least 5 species were observed :

```
apply(tmp[,-1],2, function(x) sum(x>=1))
```

```
##  min1sp  min2sp  min3sp  min4sp  min5sp  min6sp  min7sp  min8sp  min9sp min10sp min11sp min12sp
##    1230    1030     836     666     554     431     330     256     196     144     105      61
## min13sp min14sp min15sp min16sp min17sp min18sp min19sp min20sp min21sp min22sp min26sp
##      40      25      20      12       7       6       4       3       2       2       1
```

But there are for only 218 squares with at least 3 visits on which at least 5 species were observed :

```
apply(tmp[,-1],2, function(x) sum(x>=3))
```

```
##  min1sp  min2sp  min3sp  min4sp  min5sp  min6sp  min7sp  min8sp  min9sp min10sp min11sp min12sp
##     984     630     421     300     218     142      91      65      35      18      11       6
## min13sp min14sp min15sp min16sp min17sp min18sp min19sp min20sp min21sp min22sp min26sp
##       3       2       2       2       2       2       2       1       0       0       0
```

```
write.csv2(samplingEffort, "data/GBIF_data/Coccinellidae_Belgium/UTM5_Coccinellidae_sampling_effort.csv",
           row.names = FALSE)
```

Playing with the dataset

It is easy to transform the UTM x species table in a presence/absence table

```
UTMladybirds_pa <- UTMladybirds # copy of the data
UTMladybirds_pa[,-1] <- ifelse(UTMladybirds_pa[,-1]>0 , 1, 0)

# Visualize the first lines and columns
UTMladybirds_pa[1:10, 1:12]
```

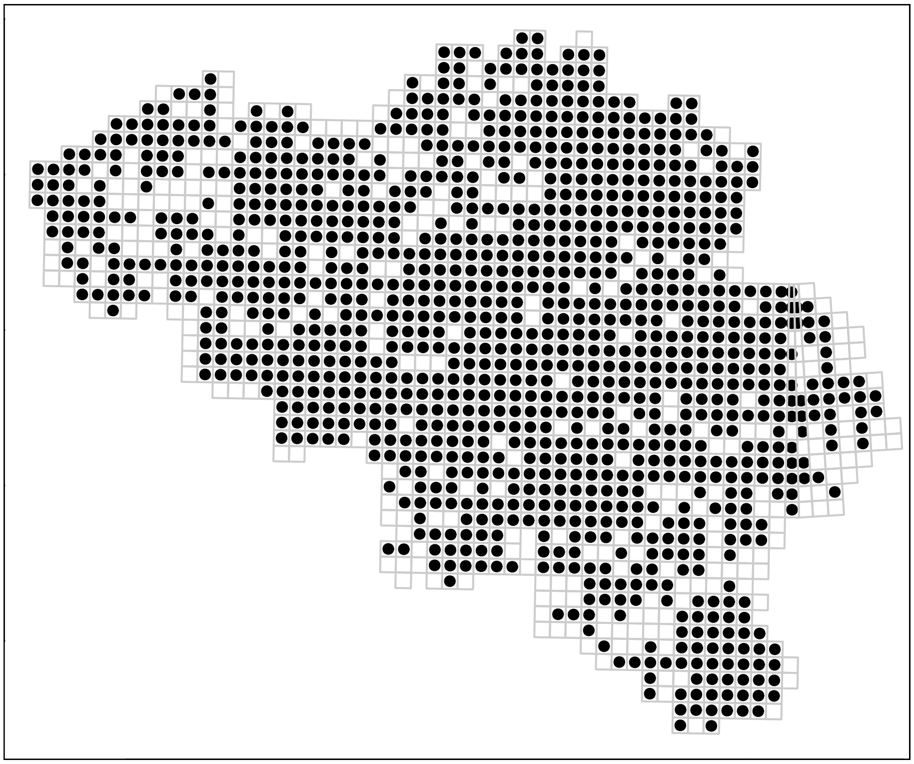
```
##           MGRS cocsep haraxy proqua adabip psyvig calqua halsed bruqua adadec tytsed caldec
## 1  31UDS675525      1      1      0      1      0      0      0      0      0      0      0
## 2  31UDS675575      1      1      0      1      1      0      1      0      1      0      0
## 3  31UDS675625      1      1      1      1      1      1      0      1      1      1      1
## 4  31UDS725425      1      1      0      1      0      0      0      0      0      0      0
## 5  31UDS725475      1      1      0      0      0      0      0      0      0      0      0
## 6  31UDS725525      1      0      0      1      0      0      0      1      0      0      0
## 7  31UDS725575      1      1      1      1      1      1      1      1      1      1      1
## 8  31UDS725625      1      1      1      1      1      1      1      1      1      1      0
## 9  31UDS775325      1      1      1      0      0      1      0      1      0      0      0
## 10 31UDS775375      1      1      1      1      1      1      1      1      1      0      0
```

Then it is easy to merge these data with the environmental data describing the UTM grid squares

```
# load the environmental dataset
env <- read.table("UTM5data.csv", sep = ";", dec = ",", header = TRUE, encoding = "utf8")
# summary(env)
# Merge the two datasets
tmp <- merge(UTMladybirds_pa, env[,1:3], by = "MGRS", all.x = TRUE)
```

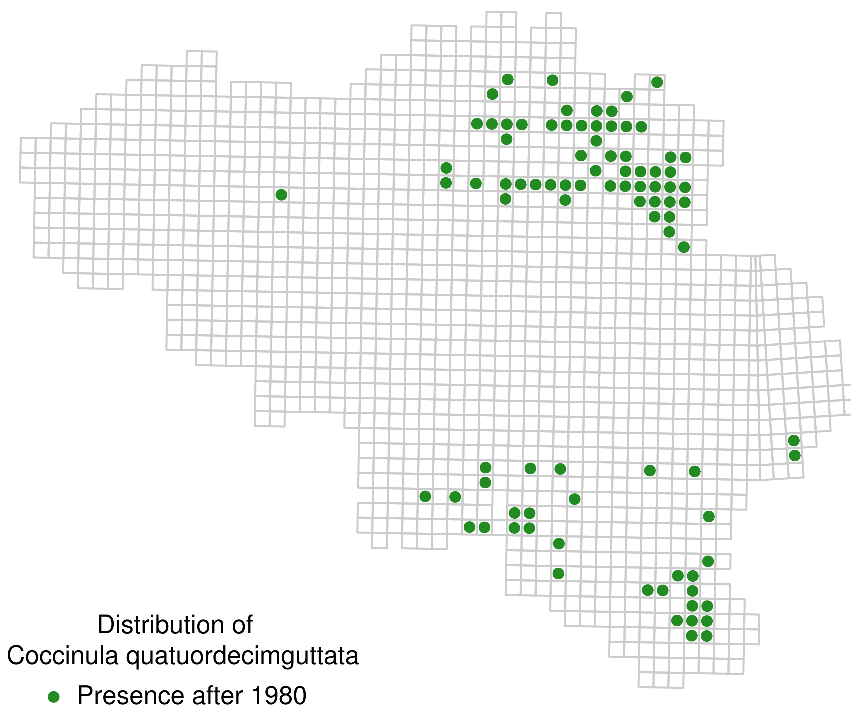
Visual check : map of the most common species (Coccinella septempunctata)

```
# dev.new(width = 12/2.54, height = 10/2.54)
par(mar = c(0,0,0,0))
plot(y~x, data = tmp[tmp$cocsep == 1,], asp = 1, pch = 20)
plot(utm, add = TRUE, border = "gray80")
```



Slightly improved map with an other species

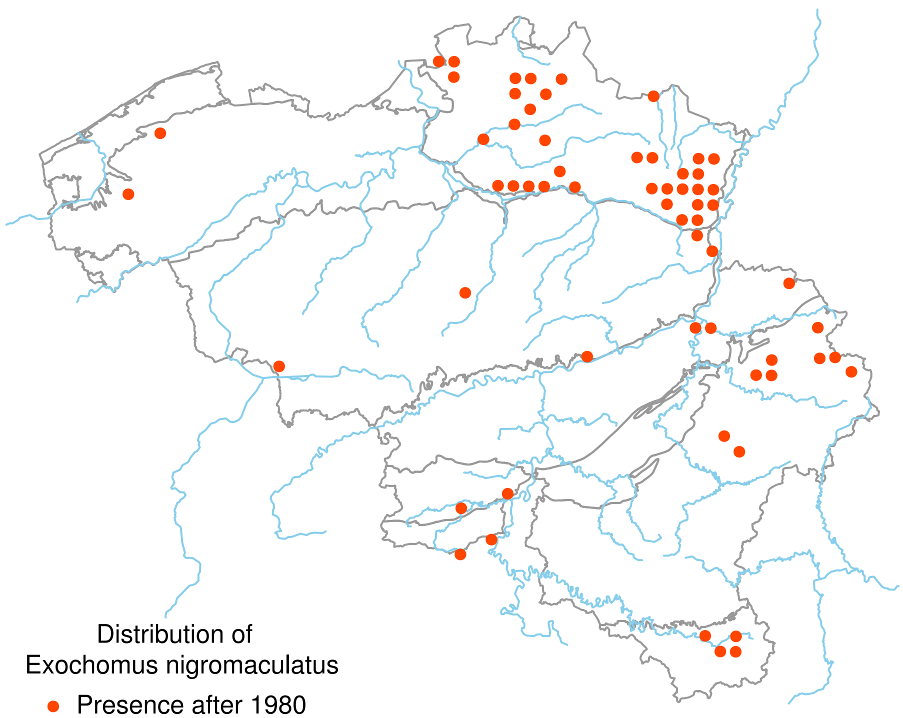
```
# dev.new(width = 12/2.54, height = 10/2.54)
par(mar = c(0,0,0,0))
plot(utm,  border = "gray80")
points(y~x, data = tmp[tmp$cocqua == 1,],
     asp = 1, pch = 20, cex = 1, col = "forestgreen")
legend("bottomleft", legend = "Presence after 1980",
       title = "Distribution of \n Coccinula quatuordecimguttata",
       pch = 20, pt.cex = 1, col = "forestgreen", bty = "n", cex = 0.8)
```



You can easily load other maps for your plots

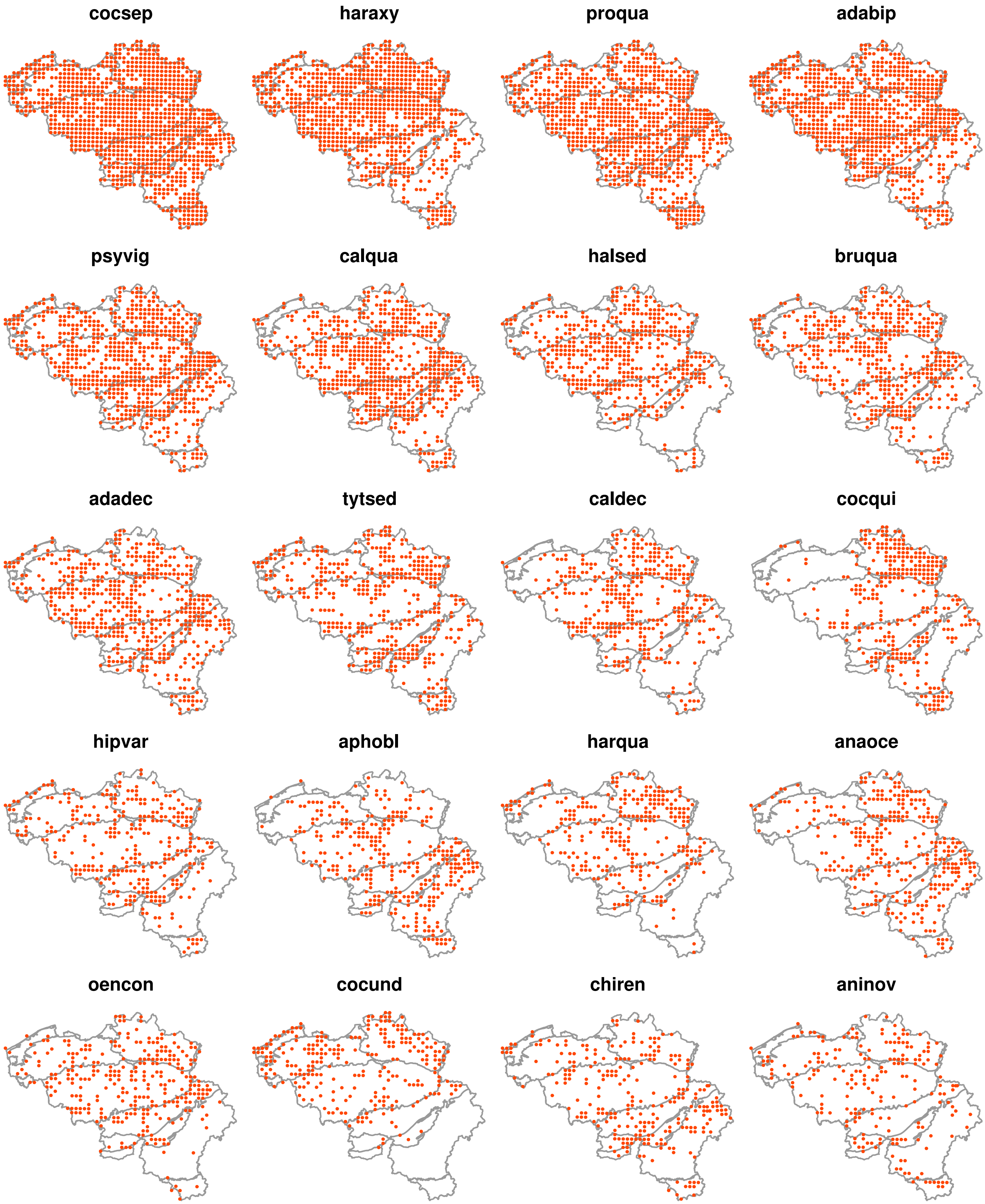
```
nr <- readOGR(GISfolder, "Regions_Naturelles", p4s = "+init=epsg:31370", verbose = FALSE)
rivers <- readOGR(GISfolder, "Rivieres", p4s = "+init=epsg:31370", verbose = FALSE)
```

```
# dev.new(width = 12/2.54, height = 10/2.54)
par(mar = c(0,0,0,0))
plot(nr,  border = "gray60")
plot(rivers, add = TRUE, col = "skyblue")
points(y~x, data = tmp[tmp$exonig == 1,],
       pch = 20, cex = 1, col = "orangered")
legend("bottomleft", legend = "Presence after 1980",
       title = "Distribution of \n Exochomus nigromaculatus",
       pch = 20, pt.cex = 1, col = "orangered", bty = "n", cex = 0.8)
```



Maps for the 20 first species in a small loop

```
# dev.new(width = 18/2.54, height = 22/2.54)
par(mfrow = c(5,4), mar = c(0,0,1.5,0))
for(i in 2:21) {
    plot(nr,  border = "gray60")
    points(y~x, data = tmp[tmp[,i] == 1,],
           asp = 1, pch = 20, cex = 0.1, col = "orangered")
    title(colnames(tmp)[i])
}
```


cocsep
haraxy
proqua
adabip
psyvig
calqua
halsed
bruqua
adadec
tytsed
caldec
cocqui
hipvar
aphobl
harqua
anaoce
oencon
cocund
chiren
aninov