rMyCoPortal vignette

Franz-Sebastian Krah

10/8/2018

The MyCoPortal

The Mycology Collections data Portal (MyCoPortal) is a database of fungal diversity with records mainly from North America. For more - and detailled - information, please visit http://mycoportal.org/portal/index.php.

The rMyCoPortal R package is an interface to the content stored on the MyCoPoral website. It allows to download records from the database readily in R for further analysis. It further provides some basic plotting functions. Below I will show the basic usability and some further possibilites of using the data.

Install rMyCoPortal

```
install.packages("devtools")
devtools::install_github("FranzKrah/rMyCoPortal")
```

Docker

Before we start using rMyCoPortal, we need to install docker (https://docs.docker.com/install/). Docker performs virtualization, also known as "containerization". rMyCoPortal interally uses the R package RSelenium to create a Selenium Server from which the MyCoPortal website is addressed.

Download records for *Amanita muscaria*, the fly agaric

```
## Load library
library("rMyCoPortal")

## Download records

am.rec <- records(taxon = "Amanita muscaria") # please run again if server doesn't respond immediatelly

## [1] "docker: Error response from daemon: driver failed programming external connectivity on endpoint
## Server is running
## Open website
## Send user query to website:
##             taxon
## "Amanita muscaria"
## Downloading 6570 records
## Make sure you have a stable internet connection!
## next page ( 1 )... done
## download... done
## next page ( 2 )... done
## download... done
## next page ( 3 )... done
```

```
## download... done
## next page ( 4 )... done
## download... done
## next page ( 5 )... done
## download... done
## next page ( 6 )... done
## download... done
## next page ( 7 )... done
## download[1] "FALSE"

## Retrying at 2018-10-08 14:01:06

## ... done
## download[1] "FALSE"

## Retrying at 2018-10-08 14:01:08

## ... done
## download[1] "FALSE"

## Retrying at 2018-10-08 14:01:10

## ... done
## download[1] "FALSE"

## Retrying at 2018-10-08 14:01:12

## ... done
## download[1] "FALSE"

## Retrying at 2018-10-08 14:01:14

## Close website and quit server
```

```
am.rec
```

```
## Fungal distribution table with 6570 records
## The retrieved object stores a distribution table with 6570 records.
```

```
head(am.rec@records)
```

```
##   Symbiota.ID Collection Catalog.Number      Family
## 1     2136920         MU      000087190 Amanitaceae
## 2     2133368         MU      000139002 Amanitaceae
## 3     2136881         MU      000139283 Amanitaceae
## 4     2136886         MU      000139284 Amanitaceae
## 5     2136890         MU      000139285 Amanitaceae
## 6     2136892         MU      000139286 Amanitaceae
##                        Scientific.Name Country State.Province     County
## 1     Amanita muscaria  (L.: Fr.) Lam.     USA      New York    Chemung
## 2     Amanita muscaria  (L.: Fr.) Lam.     USA        Oregon    Klamath
## 3     Amanita muscaria  (L.: Fr.) Lam.     USA          Ohio     Butler
## 4 Amanita muscaria var. formosa  Pers.     USA          Ohio     Preble
## 5 Amanita muscaria var. formosa  Pers.     USA          Ohio Columbiana
## 6     Amanita muscaria  (L.: Fr.) Lam.     USA          Ohio      Perry
##                                                                    Locality
## 1 Found between 114 Daffodil Drive, Horseheads and Gardner Road Elementary School,...
## 2                 Upper Munson Meadow, Crater Lake National Park, 42.9405 -122.13384
## 3       Miami University Bachelor Wildlife Preserve, near Oxford, 39.53832 -84.72901
## 4 Swan-Beatty Rd, Pleasant Vineyard Retreat Center, Somers Township, sect. 20, 39....
```

```
## 5      Sprucevale area, Beaver Creek State Park, St. Clair Township, 40.72578 -80.61316
## 6                                                 Perry County, 39.73715 -82.23614
##    Habitat Host Elevation       Event.Date    Collector Number
## 1                     301    23 June 2009 M. A. Vincent    002
## 2                    1981   22 August 1962   W. B. Cooke  33584
## 3                     249 07 November 1992    J. Studer      23
## 4                     298      29 July 1992 M. A. Vincent   5627
## 5                     265  02 October 1993 M. A. Vincent   6413
## 6                     299  07 October 1992    T. Lanese     028
##    Individual.Count Life.Stage  Sex            coord      lat      lon
## 1                NA <NA>            <NA>        NA       NA
## 2                NA <NA> 42.9405 -122.13384 42.94050 -122.13384
## 3                NA <NA> 39.53832 -84.72901 39.53832  -84.72901
## 4                NA <NA>            <NA>        NA       NA
## 5                NA <NA> 40.72578 -80.61316 40.72578  -80.61316
## 6                NA <NA> 39.73715 -82.23614 39.73715  -82.23614
##             spec
## 1 Amanita muscaria
## 2 Amanita muscaria
## 3 Amanita muscaria
## 4 Amanita muscaria
## 5 Amanita muscaria
## 6 Amanita muscaria
```

Visualization

We can now use several plotting methods to visualize the data.

```
x <- am.rec
```

```
## plot_distmap can be used to plot interactive and static distribution maps
plot_distmap(x = x, mapdatabase = "world") # the default is interactive
```

```
plot_distmap(x = x, mapdatabase = "world", interactive = FALSE) # the default is interactive
```
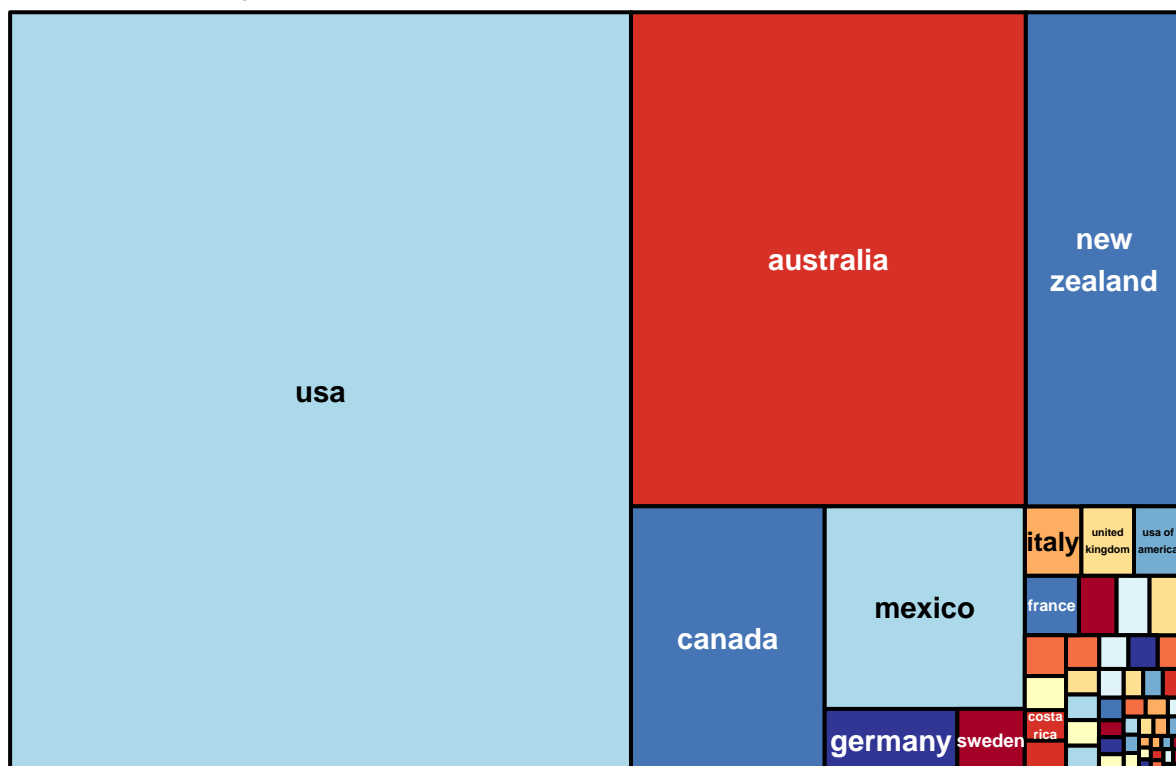
```
## Rendering map...plotting 4870 points. Not all records have coordinates.
```

Distribution map for Amanita muscaria
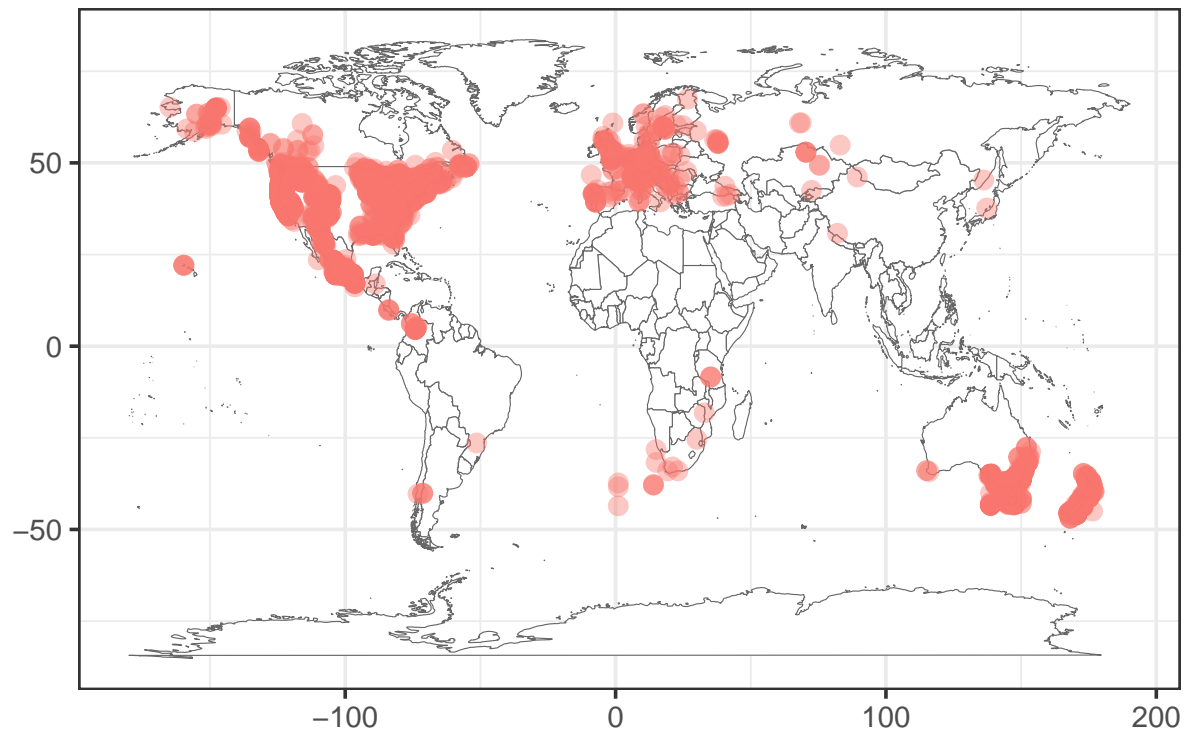

```
## plot_recordstreemap can be used to visualize relative importance of aspects of the data
plot_recordstreemap(x = x, groupvar = "country", log = FALSE) # e.g., the country distribution
```

## Log[10] Number of records for Amanita muscaria

```
## plot_datamap can be used to get a quick overview of which countries are most records rich
plot_datamap(x = x, mapdatabase = "world")
```

Log[10] Number of records for Amanita muscaria
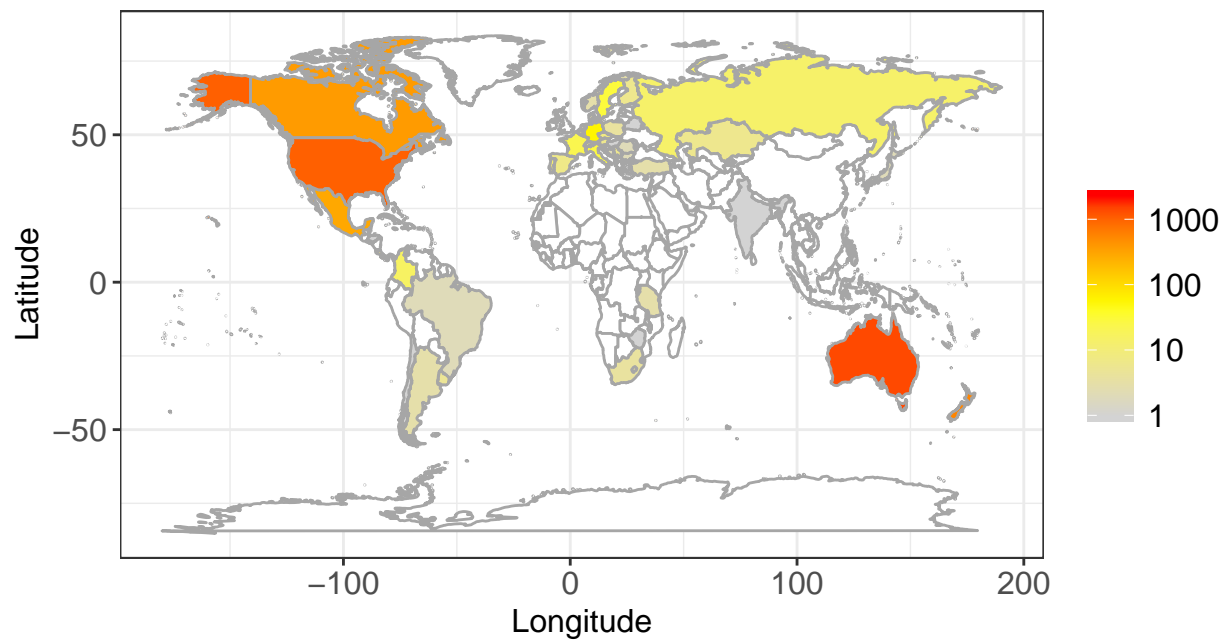

```
## the same but cropped to Europe
plot_datamap(x = x, mapdatabase = "world",
             area = list(min_long = -10, max_long = 30, min_lat = 30, max_lat = 70))
```

Log[10] Number of records for Amanita muscaria
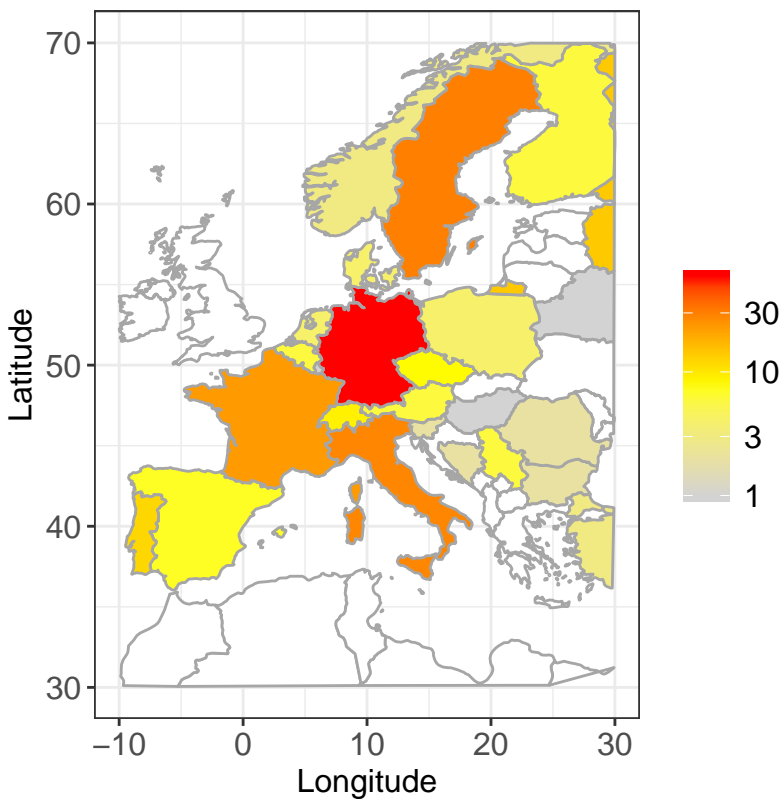
We could now use the data to look at the range of suitable climatic conditions for A. muscaria. Let's use mean annual temperature and mean annual precipitation for now.

```r
library(sf)
```

```
## Linking to GEOS 3.6.1, GDAL 2.1.3, proj.4 4.9.3
```

```r
library(raster)
```

```
## Loading required package: sp
```

```r
rec <- am.rec@records
rec <- rec[!(is.na(rec$lat) | is.na(rec$lon)), ]

my.sf.point <- st_as_sf(x = rec,
                        coords = c("lon", "lat"),
                        crs = "+proj=longlat +datum=WGS84")

## crop to USA
area = list(min_long = -130, max_long = -60, min_lat = 25, max_lat = 52)
my.sf.point <- st_crop(my.sf.point, xmin = area$min_long, ymin = area$min_lat, xmax = area$max_long, yma
```

```
## although coordinates are longitude/latitude, st_intersection assumes that they are planar
```

```
## Warning: attribute variables are assumed to be spatially constant
## throughout all geometries
```

```r
my.sf.point <- SpatialPointsDataFrame(coords = st_coordinates(my.sf.point), data = as.data.frame(my.sf.
```

```
## Retrieve WorldClim data
```

```r
clim <- raster::getData(name = "worldclim", res = "2.5", var = "bio")
clim <- crop(clim, extent(-130, -60, 25, 52))
clim <- stack(clim)

climdat <- extract(clim, my.sf.point)
climdat[,1] <- climdat[,1]/10
dat <- data.frame(as.data.frame(my.sf.point), climdat)

library(ggplot2)
p.mat <- ggplot(dat, aes(x = bio1)) +
  geom_histogram() +
  labs(x ="Mean annual temperature", y = "Count") +
  theme_bw() +
  geom_vline(aes(xintercept = mean(bio1, na.rm = TRUE)), col='red',size=2)

p.map <- ggplot(dat, aes(x = bio12)) +
  geom_histogram() +
  labs(x ="Mean annual precipitation sums", y = "Count") +
  theme_bw() +
  geom_vline(aes(xintercept = mean(bio12, na.rm = TRUE)), col='red',size=2)


library(cowplot )
```

```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggplot2':
##
##     ggsave
```

```r
plot_grid(p.mat, p.map, ncol = 2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 18 rows containing non-finite values (stat_bin).

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 18 rows containing non-finite values (stat_bin).
```