A practical example showing how to retrieve a dataset from the EcoSIS spectral database and plot the results

Shawn P. Serbin

Overview

This is an R Markdown Notebook to illustrate how to retrieve datasets from the EcoSIS spectral library (https://ecosis.org/). When you click the **Knit** button in Rstudio a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Getting Started

Installation

```
list.of.packages <- c("readr","httr","dplyr","reshape2","ggplot2")
# check for dependencies and install if needed
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
```

Load libraries

```
# load libraries needed for script
library(readr)
library(dplyr)
library(reshape2)
library(ggplot2)
```

Prepare helpers

```
# define function to grab PLSR model from GitHub
#devtools::source_gist("gist.github.com/christophergandrud/4466237")
source_GitHubData <-function(url, sep = ",", header = TRUE) {
  require(httr)
  request <- GET(url)
  stop_for_status(request)
  handle <- textConnection(content(request, as = 'text'))
  on.exit(close(handle))
  read.table(handle, sep = sep, header = header)
}

# not in
`%notin%` <- Negate(`%in%`)
```

```
print(paste0("Output temporary directory: ",outdir))
```

[1] “Output temporary directory: /var/folders/xp/h3k9vf3n2jx181ts786_yjrn9c2gjq/T//RtmpFKPhLE”

```
setwd(outdir) # set working directory
getwd()  # check wd
```

[1] “/private/var/folders/xp/h3k9vf3n2jx181ts786_yjrn9c2gjq/T/RtmpFKPhLE”

Grab data from EcoSIS

URL: https://ecosis.org/package/ngee-arctic-2016-leaf-spectral-reflectance-kougarok-road-seward-peninsula-alaska-2016

```
print("**** Downloading Ecosis data ****")
```

[1] “**** Downloading Ecosis data ****”

```
ecosis_id <- "960dbb0c-144e-4563-8117-9e23d14f4aa9"  # NGEE-Arctic dataset
ecosis_file <- sprintf(
  "https://ecosis.org/api/package/%s/export?metadata=true",
  ecosis_id
)
message("Downloading data...")
dat_raw <- read_csv(ecosis_file)
message("Download complete!")
names(dat_raw)[1:40]
```

[1] “BNL_Barcode” “CN_Ratio”
[3] “C_area_g_m2” “Cmass_g_g”
[5] “Common Name” “Foreoptic Specifications” [7] “Instrument Model” “LMA_g_m2”
[9] “Latin Genus” “Latin Species”
[11] “Latitude” “Longitude”
[13] “Measurement Quantity” “N_area_g_m2”
[15] “Nmass_g_g” “Overlap_Handling”
[17] “Overlap_Matching_Type” “Overlap_Removal”
[19] “Processing Interpolated” “Processing Resampled”
[21] “Reflectance_Type” “Sample Collection Date”
[23] “Sample_ID” “Site”
[25] “Spectra_Name” “Spectra_Type”
[27] “Spectra_Units” “Spectral_Resolution”
[29] “USDA Symbol” “White_Reference_Standard” [31] “350” “351”
[33] “352” “353”
[35] “354” “355”
[37] “356” “357”
[39] “358” “359”

```
head(dat_raw)
```

A tibble: 6 x 2,181

```
BNL_Barcode  CN_Ratio  C_area_g_m2  Cmass_g_g  Common Name  Foreoptic Spec~  <chr>
<dbl>       <dbl>     <dbl> <chr>          <chr>             1 BNL2181          19.3        48.1
47.8 Siberian ald~ Fiber_1_LC_RP_P~ 2 BNL2194          21.7        43.9      50.8 tealeaf
```

will~ Fiber_1_LC_RP_P~ 3 BNL2195 31.6 48.3 50.5 tealeaf will~ Fiber_1_LC_RP_P~ 4 BNL2196 28.0 49.5 48.4 Siberian ald~ Fiber_1_LC_RP_P~ 5 BNL2197 25.1 44.6 49.0 tealeaf will~ Fiber_1_LC_RP_P~ 6 BNL2198 17.5 51.8 47.3 tealeaf will~ Fiber_1_LC_RP_P~ # ... with 2,175 more variables:Instrument Model<chr>, LMA_g_m2 <dbl>, #Latin Genus<chr>,Latin Species<chr>, Latitude <dbl>, # Longitude <dbl>,Measurement Quantity<chr>, N_area_g_m2 <dbl>, # Nmass_g_g <dbl>, Overlap_Handling <chr>, Overlap_Matching_Type <chr>, # Overlap_Removal <chr>,Processing Interpolated<chr>,Processing # Resampled<chr>, Reflectance_Type <chr>,Sample Collection Date<dbl>, # Sample_ID <dbl>, Site <chr>, Spectra_Name <chr>, Spectra_Type <chr>, # Spectra_Units <chr>, Spectral_Resolution <chr>,USDA Symbol<chr>, # White_Reference_Standard <chr>,350<dbl>,351<dbl>,352<dbl>, #353<dbl>,354<dbl>,355<dbl>,356<dbl>,357<dbl>, #358<dbl>,359<dbl>,360<dbl>,361<dbl>,362<dbl>, #363<dbl>,364<dbl>,365<dbl>,366<dbl>,367<dbl>, #368<dbl>,369<dbl>,370<dbl>,371<dbl>,372<dbl>, #373<dbl>,374<dbl>,375<dbl>,376<dbl>,377<dbl>, #378<dbl>,379<dbl>,380<dbl>,381<dbl>,382<dbl>, #383<dbl>,384<dbl>,385<dbl>,386<dbl>,387<dbl>, #388<dbl>,389<dbl>,390<dbl>,391<dbl>,392<dbl>, #393<dbl>,394<dbl>,395<dbl>,396<dbl>,397<dbl>, #398<dbl>,399<dbl>,400<dbl>,401<dbl>,402<dbl>, #403<dbl>,404<dbl>,405<dbl>,406<dbl>,407<dbl>, #408<dbl>,409<dbl>,410<dbl>,411<dbl>,412<dbl>, #413<dbl>,414<dbl>,415<dbl>,416<dbl>,417<dbl>, #418<dbl>,419<dbl>,420<dbl>,421<dbl>,422<dbl>, #423<dbl>,424<dbl>,425‘ , ...

Prepare spectra data

```
Start.wave <- 500
End.wave <- 2400
wv <- seq(Start.wave,End.wave,1)
spectra <- data.frame(dat_raw[,names(dat_raw) %in% wv])
names(spectra) <- c(paste0("Wave_",wv))
head(spectra)[,1:5]
```

Wave_500 Wave_501 Wave_502 Wave_503 Wave_504 1 2.687619 2.705000 2.699524 2.709048 2.716429 2 2.174524 2.231667 2.283810 2.340952 2.380952 3 2.159762 2.231667 2.322857 2.432381 2.516190 4 3.229524 3.258333 3.285000 3.335000 3.342143 5 2.364286 2.380000 2.413571 2.458810 2.544762 6 2.226429 2.228333 2.255238 2.283810 2.310238

```
sample_info <- dat_raw[,names(dat_raw) %notin% seq(350,2500,1)]
head(sample_info)
```

A tibble: 6 x 30

BNL_Barcode CN_Ratio C_area_g_m2 Cmass_g_g Common Name Foreoptic Spec~ <chr> <dbl> <dbl> <dbl> <chr> <chr> 1 BNL2181 19.3 48.1 47.8 Siberian ald~ Fiber_1_LC_RP_P~ 2 BNL2194 21.7 43.9 50.8 tealeaf will~ Fiber_1_LC_RP_P~ 3 BNL2195 31.6 48.3 50.5 tealeaf will~ Fiber_1_LC_RP_P~ 4 BNL2196 28.0 49.5 48.4 Siberian ald~ Fiber_1_LC_RP_P~ 5 BNL2197 25.1 44.6 49.0 tealeaf will~ Fiber_1_LC_RP_P~ 6 BNL2198 17.5 51.8 47.3 tealeaf will~ Fiber_1_LC_RP_P~ # ... with 24 more variables:Instrument Model<chr>, LMA_g_m2 <dbl>,Latin # Genus<chr>,Latin Species<chr>, Latitude <dbl>, Longitude <dbl>, #Measurement Quantity<chr>, N_area_g_m2 <dbl>, Nmass_g_g <dbl>, # Overlap_Handling <chr>, Overlap_Matching_Type <chr>, Overlap_Removal <chr>, #Processing Interpolated<chr>,Processing Resampled<chr>, # Reflectance_Type <chr>,Sample Collection Date<dbl>, Sample_ID <dbl>, # Site <chr>, Spectra_Name <chr>, Spectra_Type <chr>, Spectra_Units <chr>, # Spectral_Resolution <chr>,USDA Symbol‘ , # White_Reference_Standard

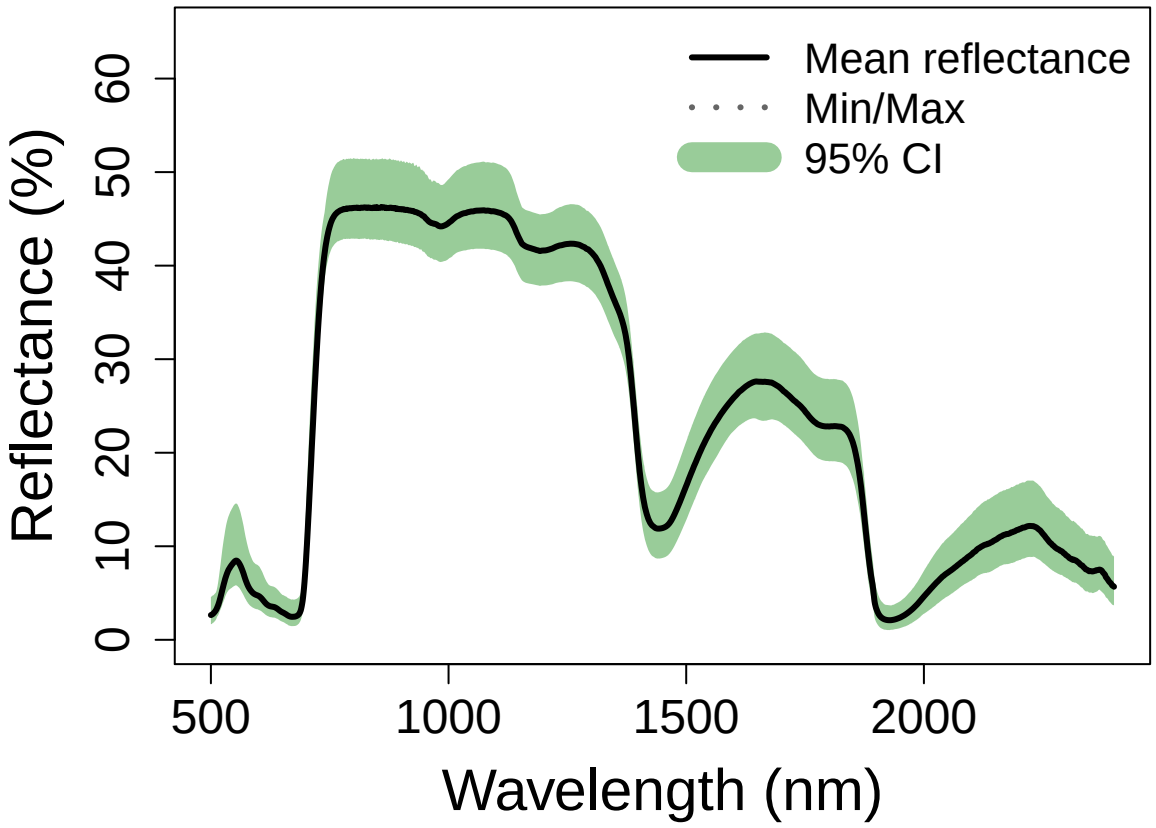
```
cexaxis <- 1.5
cexlab <- 1.8
ylim <- 65
ylim2 <- 65

# calculate some stats
mean_spec <- colMeans(spectra[,which(names(spectra) %in% paste0("Wave_",wv))])
spectra_quantiles <- apply(spectra[,which(names(spectra) %in% paste0("Wave_",wv))],
                           2,quantile,na.rm=T,probs=c(0,0.025,0.05,0.5,0.95,0.975,1))

print("**** Plotting Ecosis specrtal data. Writing to scratch space ****")
```

[1] "**** Plotting Ecosis specrtal data. Writing to scratch space ****"

```
par(mfrow=c(1,1), mar=c(4.5,5.7,0.3,0.4), oma=c(0.3,0.9,0.3,0.1)) # B, L, T, R
plot(wv,mean_spec,ylim=c(0,ylim),cex=0.00001, col="white",xlab="Wavelength (nm)",
     ylab="Reflectance (%)",cex.axis=cexaxis, cex.lab=cexlab)
polygon(c(wv ,rev(wv)),c(spectra_quantiles[7,], rev(spectra_quantiles[3,])),
        col="#99CC99",border=NA)
lines(wv,mean_spec,lwd=3, lty=1, col="black")
lines(wv,spectra_quantiles[1,]*100,lwd=1.85, lty=3, col="grey40")
lines(wv,spectra_quantiles[7,]*100,lwd=1.85, lty=3, col="grey40")
legend("topright",legend=c("Mean reflectance","Min/Max", "95% CI"),lty=c(1,3,1),
       lwd=c(3,3,15),col=c("black","grey40","#99CC99"),bty="n", cex=1.3)
```




```
# save to scratch space
png(file=file.path(outdir,'NGEE-Arctic_2016_Kougarok_leaf_spectra_summary_plot.png'),height=3000,
    width=3900, res=340)
```

```
par(mfrow=c(1,1), mar=c(4.5,5.7,0.3,0.4), oma=c(0.3,0.9,0.3,0.1)) # B, L, T, R
plot(wv,mean_spec,ylim=c(0,ylim),cex=0.00001, col="white",xlab="Wavelength (nm)",
     ylab="Reflectance (%)",cex.axis=cexaxis, cex.lab=cexlab)
polygon(c(wv ,rev(wv)),c(spectra_quantiles[7,], rev(spectra_quantiles[3,])),
        col="#99CC99",border=NA)
lines(wv,mean_spec,lwd=3, lty=1, col="black")
lines(wv,spectra_quantiles[1,]*100,lwd=1.85, lty=3, col="grey40")
lines(wv,spectra_quantiles[7,]*100,lwd=1.85, lty=3, col="grey40")
legend("topright",legend=c("Mean reflectance","Min/Max", "95% CI"),lty=c(1,3,1),
       lwd=c(3,3,15),col=c("black","grey40","#99CC99"),bty="n", cex=1.7)
box(lwd=2.2)
dev.off()
```

pdf 2

Plot associated leaf functional trait data

```
print("**** Plotting Ecosis trait data. Writing to scratch space ****")
```

[1] "**** Plotting Ecosis trait data. Writing to scratch space ****"

```
# Organize leaf trait data
names(sample_info)
```

[1] "BNL_Barcode" "CN_Ratio"
[3] "C_area_g_m2" "Cmass_g_g"
[5] "Common Name" "Foreoptic Specifications" [7] "Instrument Model" "LMA_g_m2"
[9] "Latin Genus" "Latin Species"
[11] "Latitude" "Longitude"
[13] "Measurement Quantity" "N_area_g_m2"
[15] "Nmass_g_g" "Overlap_Handling"
[17] "Overlap_Matching_Type" "Overlap_Removal"
[19] "Processing Interpolated" "Processing Resampled"
[21] "Reflectance_Type" "Sample Collection Date"
[23] "Sample_ID" "Site"
[25] "Spectra_Name" "Spectra_Type"
[27] "Spectra_Units" "Spectral_Resolution"
[29] "USDA Symbol" "White_Reference_Standard"

```
trait_data <- sample_info %>%
  select(Site,Sample_ID,USDA_Species_Code=`USDA Symbol`,Common_Name=`Common Name`,LMA_gDW_m2=LMA_g_m2,
         Nmass_g_g,N_area_g_m2,Cmass_g_g,C_area_g_m2,CN_Ratio)
head(trait_data)
```

A tibble: 6 x 10

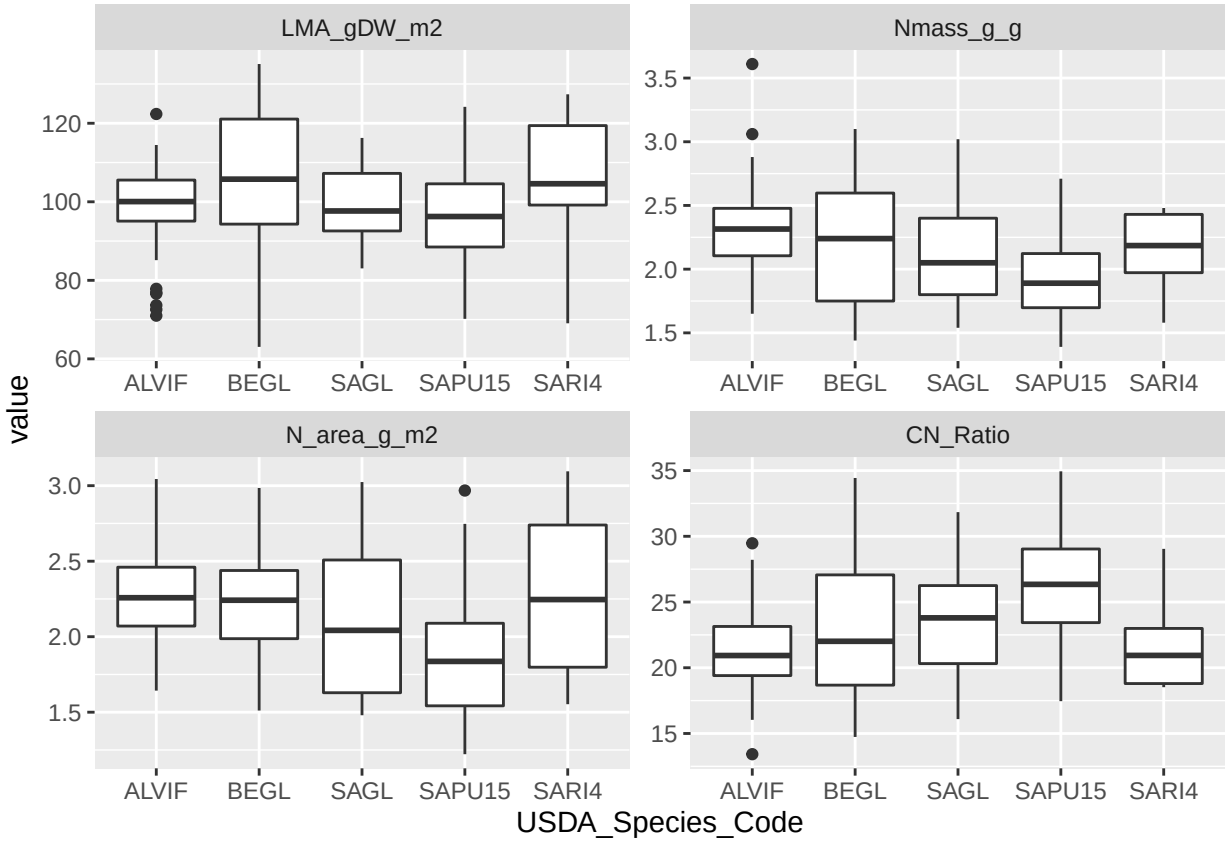
Site Sample_ID USDA_Species_Co~ Common_Name LMA_gDW_m2 Nmass_g_g N_area_g_m2 1 Koug~ 2181 ALVIF Siberian a~ 101. 2.47 2.48 2 Koug~ 2194 SAPU15 tealeaf wi~ 86.4 2.34 2.02 3 Koug~ 2195 SAPU15 tealeaf wi~ 95.7 1.6 1.53 4 Koug~ 2196 ALVIF Siberian a~ 102. 1.73 1.77 5 Koug~ 2197 SAPU15 tealeaf wi~ 91.0 1.95 1.78 6 Koug~ 2198 SAPU15 tealeaf wi~ 110. 2.71 2.97 # ... with 3 more variables: Cmass_g_g , C_area_g_m2 , CN_Ratio

```
# Prepare data for ggplot
trait_data <- melt(data = trait_data, id.vars = "USDA_Species_Code", measure.vars = c("LMA_gDW_m2",
                                                                                      "Nmass_g_g",
                                                                                      "N_area_g_m2",
                                                                                      "CN_Ratio"))

head(trait_data)
```

USDA_Species_Code variable value 1 ALVIF LMA_gDW_m2 100.63 2 SAPU15 LMA_gDW_m2 86.39 3 SAPU15 LMA_gDW_m2 95.67 4 ALVIF LMA_gDW_m2 102.33 5 SAPU15 LMA_gDW_m2 91.02 6 SAPU15 LMA_gDW_m2 109.53

```
# Graph the trait data and save a file to the scratch space
p2 <- ggplot(trait_data, aes(x=USDA_Species_Code, y=value)) +
  geom_boxplot() +
  facet_wrap(~variable, scale="free")
p2 # plot the results
```




```
# save the results to a file in the scratch space
ggsave(filename = file.path(outdir,"NGEE-Arctic_2016_Kougarok_Trait_data.png"), plot = p2,
       width = 40, height = 20, units = "cm")
```

Disclaimer

This vignette provided for demonstration purposes only