Coding Corner

This month's coding corner will introduce readers to R programming and utilizing the iDigBio Search API endpoints for querying specimen data in aggregate.

Before we begin, we need to prepare our work space and load the packages we will need for this project:

```
library(jsonlite)
library(lubridate)
library(seas)
library(ggplot2)
```

The iDigBio API has resources, or "endpoints", for querying data in the aggregate. To facilitate discovery, some of the endpoints provide summary statistics or summary data: https://github.com/iDigBio/idigbio-search-api/wiki#summary

In this coding corner, we will use the "Date Histogram" to begin our data exploration.

To begin, let's tell R what API summary endpoint we would like to use by creating a vector representing the endpoint:

```
apiEndpoint <- "http://search.idigbio.org/v2/summary/datehist/"
```

We need to set up our query to follow the API query format: https://github.com/iDigBio/idigbio-search-api/wiki/Query-Format . This call to the API will take its arguments in JSON, similar to this example: https://github.com/iDigBio/idigbio-search-api/wiki/Query-Format#searching-for-a-value-within-a-field

```
rq <- toJSON(list(scientificname="Cladonia rangiferina"))
```

We can now construct a URL to query the endpoint and assign it to a vector that we will pass to our parsing function

```
queryURL <- URLencode(paste(apiEndpoint,"?rq=",rq,sep=""))
```

And see what response we get back:

```
res <- fromJSON(queryURL)
summary(res)
```

```
##            Length Class  Mode
## dates      166    -none- list
## itemCount    1    -none- numeric
## rangeCount   1    -none- numeric
```

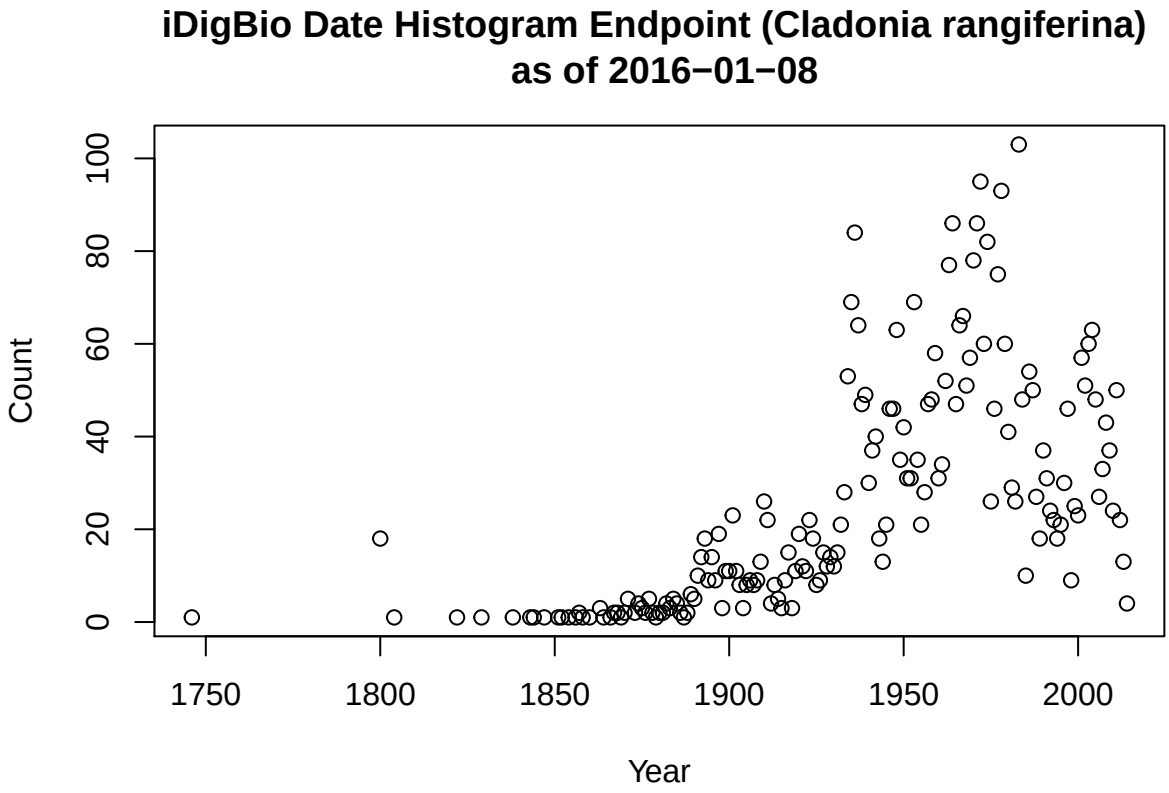
The API has returned a nested list of years and counts. Lets create a tidy data frame from the response so that we can create a plot.

```
df <- data.frame(unlist(res$dates),as.Date(names(res$dates)))
names(df)[1:2] <- c("count","year")
str(df)
```

```
## 'data.frame':    166 obs. of  2 variables:
##  $ count: int  1 18 1 1 1 1 1 1 1 1 ...
##  $ year : Date, format: "1746-01-01" "1800-01-01" ...
```

Plot time! We'll make a scatter plot, using R's base plot package, of the data frame we just created with dates on the x axis and counts on the y axis:

```
plot(df$year,df$count,
     main=paste("iDigBio Date Histogram Endpoint (Cladonia rangiferina) \n as of ",Sys.Date(),sep=""),
     xlab="Year",
     ylab="Count")
```



Now that we have an idea of the distribution of collection dates in the data, let's take a further look into how these collection events are distributed by locality, using the "ridigbio" package

```
library(ridigbio)
```

Let's query the iDigBio API for a response that contains locality information, along with our collection dates, and restrict it to our species of interest:

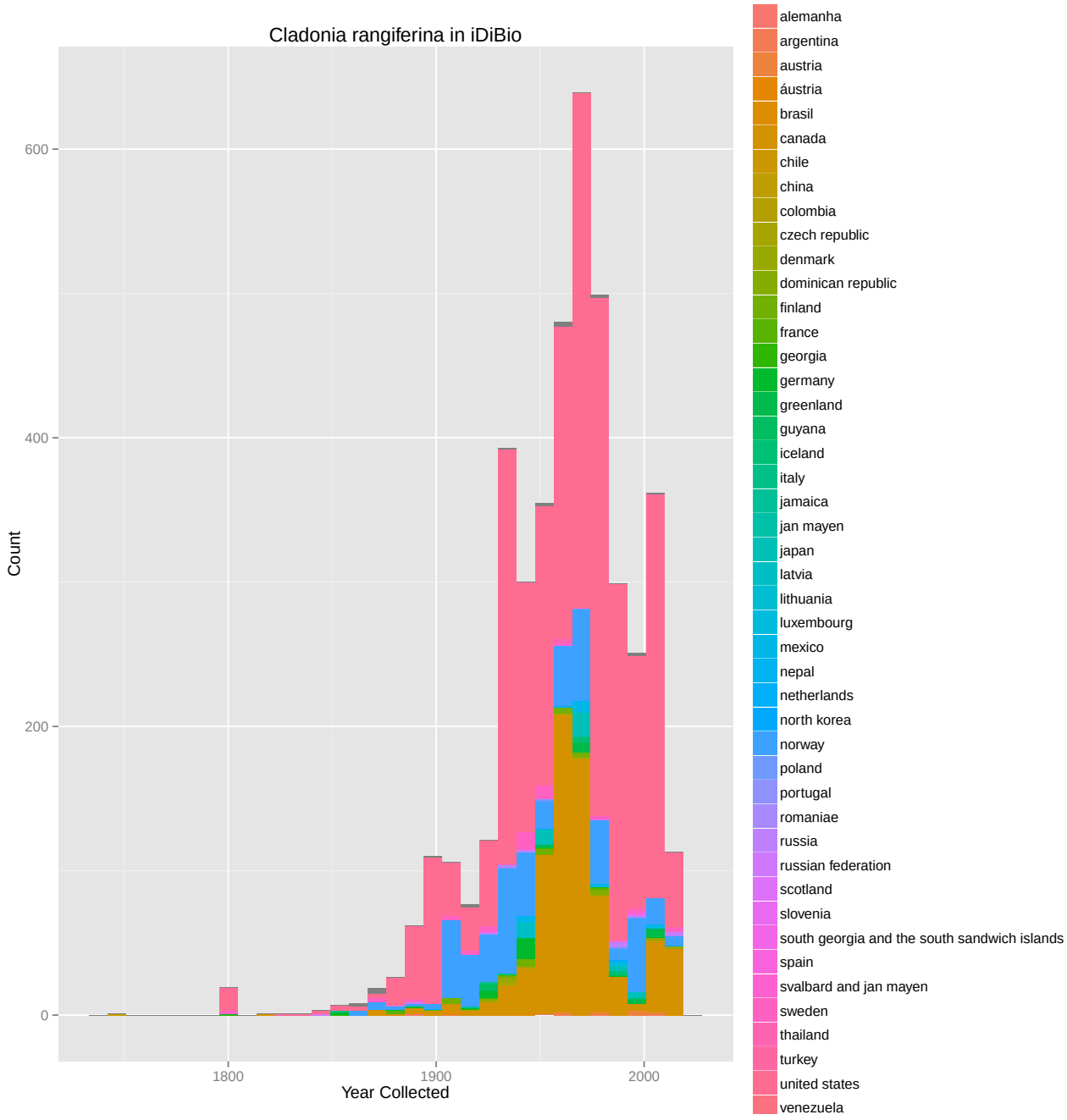
```
lichenData <- idig_search_records(rq = list(scientificname="Cladonia rangiferina"), fields = c("datecoll
```

We're going to want to add some dimension to out plots, so lets calculate the "season" the specimen was collected

```
lichenData$seasons <- mkseas(as.Date(lichenData$datecollected),"DJF")
```

Plot the result, starting with a histogram

```
ggplot(lichenData,aes(x=year(datecollected),fill=as.factor(country))) +
        geom_histogram() +
        labs(x="Year Collected",y="Count", title="Cladonia rangiferina in iDiBio")
```

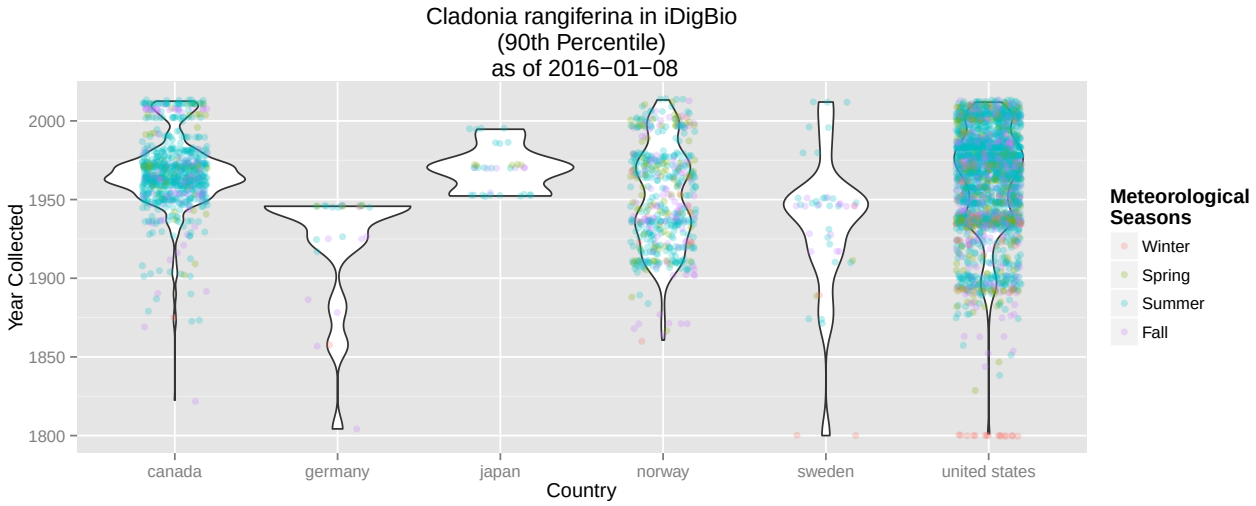


Subset by countries with 90% of records and create a new data frame

```
tt <- as.data.frame(table(lichenData$country))
tt$Pct <- tt$Freq / sum(tt$Freq)
tt <- tt[tt$Pct>quantile(tt$Pct, .9),]
df2 <- lichenData[lichenData$country %in% tt$Var1,]
```

Create fancy visualization from this 90th percentile data

```
ggplot(df2, aes(x=country, y=year(datecollected)))+geom_violin()+
        geom_jitter(alpha=0.25,
                    aes(color=seasons),
                    position = position_jitter(width = .2))+
        labs(x="Country",y="Year Collected", title=paste("Cladonia rangiferina in iDigBio\n(90th Percen
        theme(legend.title = element_text(size=12, face="bold"))+
        scale_color_discrete(name="Meteorological \nSeasons", labels=c("Winter","Spring","Summer","Fall
```



The iDigBio API also returns attribution data with each request. The "ridigbio" package also add this attribution data as an attribute of the data frame it creates using the "idigbio_search_records" function. See if you can work out how we created the following block of attribution text using these methods.

Attribution text for the figures above:

```
## http://www.idigbio.org/portal 2016,
## 6910 records, accesed on 2016-01-08 13:29:45,
## Contributed by 65 Recordsets, Recordset identifiers:
##  http://www.idigbio.org/portal/recordsets/3f508496-c860-4701-93e4-84e940c8395e (1073) records
##  http://www.idigbio.org/portal/recordsets/6f82f182-39b4-4b3f-9087-91f6afafc04e (1050) records
##  http://www.idigbio.org/portal/recordsets/58402fe3-37c1-4d15-9e07-0ff1c4c9fb11 (1010) records
##  http://www.idigbio.org/portal/recordsets/29d217e3-754b-4a72-9e57-5cd05312e7c0 (697) records
##  http://www.idigbio.org/portal/recordsets/5ea005e8-626f-47de-afee-972e976cc3a7 (521) records
##  http://www.idigbio.org/portal/recordsets/ef04e127-bb7d-4bf0-82d3-767d43108f81 (289) records
##  http://www.idigbio.org/portal/recordsets/c481fbc6-4bd7-4c50-8537-ba1993d4eb88 (275) records
##  http://www.idigbio.org/portal/recordsets/d29b9265-07e6-4e73-8f72-fc42d3d83fb1 (205) records
##  http://www.idigbio.org/portal/recordsets/35879d2c-063f-4046-9ac6-eda6410e21a9 (192) records
##  http://www.idigbio.org/portal/recordsets/1bb33d2d-0714-4fc9-968e-b66bab1cf3d3 (144) records
##  http://www.idigbio.org/portal/recordsets/d2c71720-e156-4943-8182-0a7bbe477a37 (99) records
##  http://www.idigbio.org/portal/recordsets/7110b8ba-0ead-4666-8279-e30f53e343d0 (94) records
##  http://www.idigbio.org/portal/recordsets/a6743a43-b86a-4265-9521-fad3a24461a6 (88) records
##  http://www.idigbio.org/portal/recordsets/bdf65f9c-a730-4083-bd8d-a2def3037637 (85) records
##  http://www.idigbio.org/portal/recordsets/540e18dc-09aa-4790-8b47-8d18ae86fabc (83) records
##  http://www.idigbio.org/portal/recordsets/0fcbf959-b714-4ba2-8152-0c1440e31323 (69) records
##  http://www.idigbio.org/portal/recordsets/2823b0c8-dd5f-487b-a0d0-7411005a4eaa (67) records
##  http://www.idigbio.org/portal/recordsets/1ad40bde-8a2a-46bb-9252-0cdc53df5683 (66) records
##  http://www.idigbio.org/portal/recordsets/6b565194-9707-42da-8052-9f9cf5f9aa60 (63) records
##  http://www.idigbio.org/portal/recordsets/dfd53a42-8f63-4040-93a5-3f1347ce7686 (56) records
```

```
##  http://www.idigbio.org/portal/recordsets/9756b9a4-c070-4359-8a07-2383b09d0d04 (55) records
##  http://www.idigbio.org/portal/recordsets/7c927849-94ed-4034-90e9-af34ac0cb47c (40) records
##  http://www.idigbio.org/portal/recordsets/f4bec217-9676-4fc0-be90-856b4b89d4d1 (39) records
##  http://www.idigbio.org/portal/recordsets/df22987f-d20d-41db-b8eb-8b5f5fca6df0 (37) records
##  http://www.idigbio.org/portal/recordsets/4b92de1f-866d-4b82-af69-37d46753f289 (36) records
##  http://www.idigbio.org/portal/recordsets/821c1855-6817-40ee-8732-7f472d238513 (31) records
##  http://www.idigbio.org/portal/recordsets/063825dc-b8c3-4962-aea4-9994bcc09bc8 (29) records
##  http://www.idigbio.org/portal/recordsets/0e0e9bbc-1dea-4de4-95ae-aecc90844bbf (29) records
##  http://www.idigbio.org/portal/recordsets/9368e302-f8e7-4714-aed4-db2faa861e5c (28) records
##  http://www.idigbio.org/portal/recordsets/2eb8ff2f-4826-4fc3-be68-22d805bcae88 (26) records
##  http://www.idigbio.org/portal/recordsets/a748a0fe-a6ae-4ce7-b88f-4e4ec1dc080c (26) records
##  http://www.idigbio.org/portal/recordsets/b5e5c781-765f-4981-af2a-c19c250e2cf0 (26) records
##  http://www.idigbio.org/portal/recordsets/df516dc6-6ef0-426d-94e3-8a2bbb0439a5 (25) records
##  http://www.idigbio.org/portal/recordsets/fb97dfb4-72be-4dc1-9f5a-2faea75341b4 (25) records
##  http://www.idigbio.org/portal/recordsets/15a1cc29-b66c-4633-ad9c-c2c094b19902 (23) records
##  http://www.idigbio.org/portal/recordsets/1a8eea37-7c72-4032-a38a-254154449ad1 (22) records
##  http://www.idigbio.org/portal/recordsets/33fd0737-6207-42cc-bc64-cc637266b476 (18) records
##  http://www.idigbio.org/portal/recordsets/a5fdee09-34c4-48bc-99ff-a503c93a9d7e (17) records
##  http://www.idigbio.org/portal/recordsets/995cc7f1-69c3-4317-ab77-28fd48f1e535 (15) records
##  http://www.idigbio.org/portal/recordsets/9d2a4189-6048-46e9-bac4-e5ef566334bb (14) records
##  http://www.idigbio.org/portal/recordsets/fd14095c-3658-4e00-8cec-729a89459e92 (14) records
##  http://www.idigbio.org/portal/recordsets/d81c6ad6-fb8f-4c31-bba3-f2b65f780893 (12) records
##  http://www.idigbio.org/portal/recordsets/237bd113-32f3-4091-9710-4a1b074fe26d (10) records
##  http://www.idigbio.org/portal/recordsets/40987883-03cf-494a-a5cf-7c77c7aadb79 (9) records
##  http://www.idigbio.org/portal/recordsets/a4b888a2-94bf-4680-b912-84964a236c82 (9) records
##  http://www.idigbio.org/portal/recordsets/2e185eda-1790-45e3-88d6-261304c37ed4 (8) records
##  http://www.idigbio.org/portal/recordsets/538bdd33-b616-4825-8542-7033cc8a185f (6) records
##  http://www.idigbio.org/portal/recordsets/62f951c1-b6a8-430c-8652-5691f079152c (6) records
##  http://www.idigbio.org/portal/recordsets/cd5bc13d-ee5c-4b68-a550-37edb3e7899d (6) records
##  http://www.idigbio.org/portal/recordsets/3f1100f3-c4be-4c94-af61-bd0d2d011b8a (5) records
##  http://www.idigbio.org/portal/recordsets/5f6fcfc2-598c-42e8-abb3-50ca9c2446e2 (5) records
##  http://www.idigbio.org/portal/recordsets/1c8ec291-8067-4b48-848b-410c2c768420 (4) records
##  http://www.idigbio.org/portal/recordsets/331b6d1b-842e-4c63-aa23-75ef275d8a9f (4) records
##  http://www.idigbio.org/portal/recordsets/4cdf5c2f-1a44-4fd5-bdd8-de08c8a660e2 (4) records
##  http://www.idigbio.org/portal/recordsets/b3d53973-5bac-432a-90d3-7956baa09c5d (4) records
##  http://www.idigbio.org/portal/recordsets/17cea35c-721f-4d9b-b67f-d29250064d25 (3) records
##  http://www.idigbio.org/portal/recordsets/5f513dff-ccd8-4578-ad0b-5e6cf035e4d1 (3) records
##  http://www.idigbio.org/portal/recordsets/fc40fabd-0a70-48fa-b142-79990cd259a5 (3) records
##  http://www.idigbio.org/portal/recordsets/ced8c9bc-e8b5-49e7-860a-289fc913860c (2) records
##  http://www.idigbio.org/portal/recordsets/0bada388-4adf-4b8c-b733-0a1bfc7c233c (1) records
##  http://www.idigbio.org/portal/recordsets/253f90be-3b94-469c-820c-cb727b85bdd4 (1) records
##  http://www.idigbio.org/portal/recordsets/6cab4420-11e4-4b55-85ac-6ecfdda70184 (1) records
##  http://www.idigbio.org/portal/recordsets/a6e02b78-6fc6-4cb6-bb87-8d5a443f2c2a (1) records
##  http://www.idigbio.org/portal/recordsets/b6d0f953-29b4-41da-a255-2ed07c83edf1 (1) records
##  http://www.idigbio.org/portal/recordsets/bf1fee2d-f760-4068-b8e6-d1db63ce434c (1) records
```