

We're going to need a little help installing iDigBio's R library, so lets start there:

```
install.packages("devtools")
library(devtools)
```

The console may output some message from code example #1, although it will not return an output. As long as we don't encounter any errors, now we can use the function "install_github" to install iDigBio's R library and then load it into memory:

```
install_github("idigbio/ridigbio")
library(ridigbio)
```

Let's take a look at this month's biodiversity spotlight, the fathead minnow, "Pimephales promelas", in iDigBio by creating a dataframe using the function "idig_search_records":

```
fatheadData <- idig_search_records(rq=list(scientificname="pimephales promelas"))
str(fatheadData)
```

```
## 'data.frame': 4463 obs. of 12 variables:
## $ uuid : chr "0008d42b-d35d-495a-8f1a-85da16a56f95" "001c0813-9193-4587-845c-b1fa36c3a3a4" ...
## $ occurrenceid : chr "dca6d8ae-1ed3-11e3-bfac-90b11c41863e" "urn:lsid:biosci.ohio-state.edu:osuc_..." ...
## $ catalognumber : chr "34424" "osum 7449" "7223" "832" ...
## $ family : chr "cyprinidae" "cyprinidae" "cyprinidae" "cyprinidae" ...
## $ genus : chr "pimephales" "pimephales" "pimephales" "pimephales" ...
## $ scientificname: chr "pimephales promelas" "pimephales promelas" "pimephales promelas" "pimephales..." ...
## $ country : chr "united states" "united states" "united states" "united states" ...
## $ stateprovince : chr "kansas" "ohio" "colorado" "illinois" ...
## $ geopoint.lat : chr "39.7361111" "41.5064" NA "41.48569" ...
## $ geopoint.lon : chr "-101.8788889" "-83.0577" NA "-87.60199" ...
## $ datecollected : chr "2003-06-03T00:00:00+00:00" NA NA "1969-11-07T00:00:00+00:00" ...
## $ collector : chr "kansas department of wildlife and parks" "daiber, f. c. (frank c.) & ryther"
```

In the last block of code, we asked R to provide a glimpse of the structure of the dataframe. We can see that the iDigBio search API returned columns of data, and we can check their names with:

```
colnames(fatheadData)
```

```
## [1] "uuid" "occurrenceid" "catalognumber" "family"
## [5] "genus" "scientificname" "country" "stateprovince"
## [9] "geopoint.lat" "geopoint.lon" "datecollected" "collector"
```

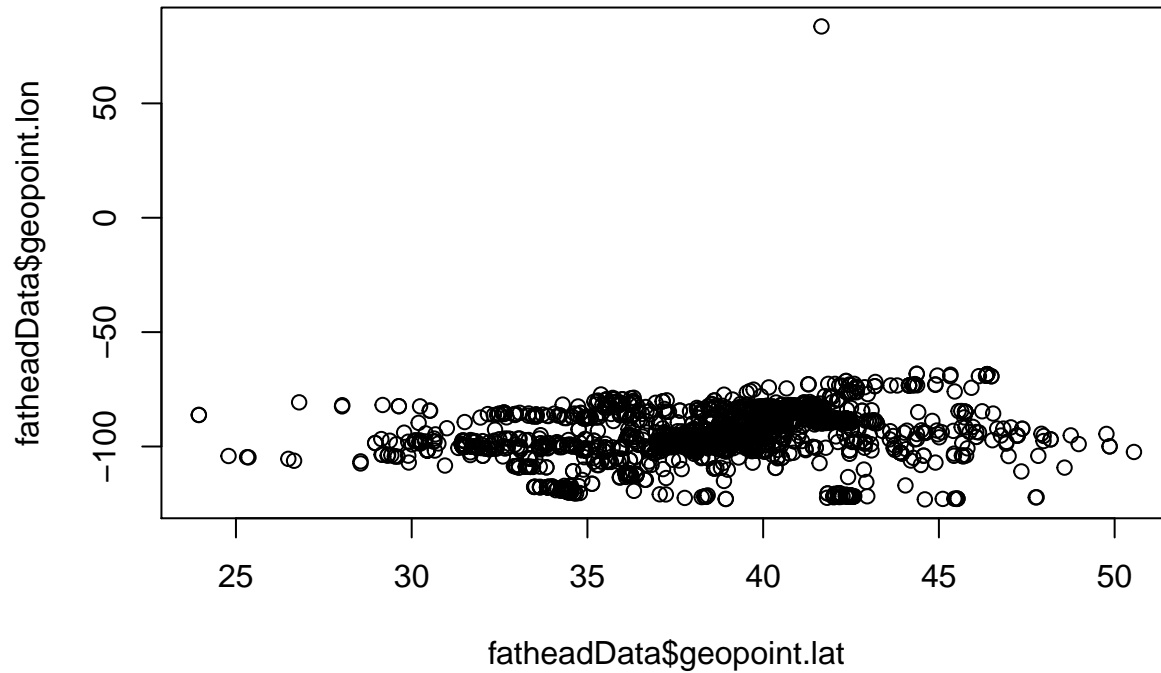
Our goal for this coding corner is to make a basic map. For some sanity checking of the data, let's see what data are present in the "country" column:

```
unique(fatheadData$country)
```

```
## [1] "united states" "mexico" "canada" "nicaragua"
## [5] NA
```

And we can create a plot of the latitude and longitude of these records:

```
plot(fatheadData$geopoint.lat, fatheadData$geopoint.lon)
```



Since we know all of our values must land within the US, Mexico, Nicaragua, and Canada, it appears like we may have an outlier. We can take a closer look:

```
outlier <- fatheadData[which(fatheadData$geopoint.lon >0),]  
outlier$country
```

```
## [1] "united states"
```

Since we know that this outlier is an error in the data, we will create a new dataframe with this outlier omitted

```
cleanedData <- fatheadData[which(fatheadData$suuid != outlier$suuid),]
```

Now we can output this data to a CSV file so we can work with it in other applications:

```
write.csv(cleanedData, file="idigbio-fathead-data.csv", row.names=FALSE)
```

As an exercise for the reader, can you recreate the following map with the concepts presented above?

Pimephales promelas in iDigBio

