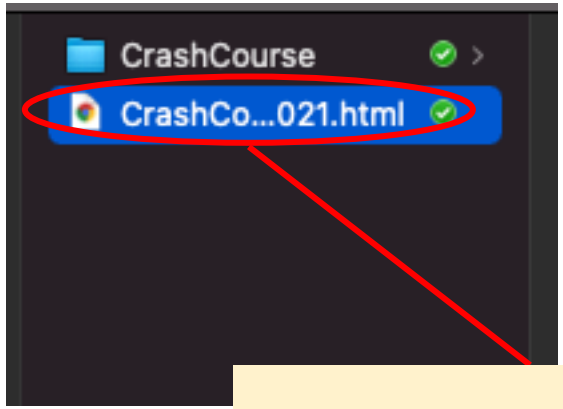
ENM Crash Course

University of Florida

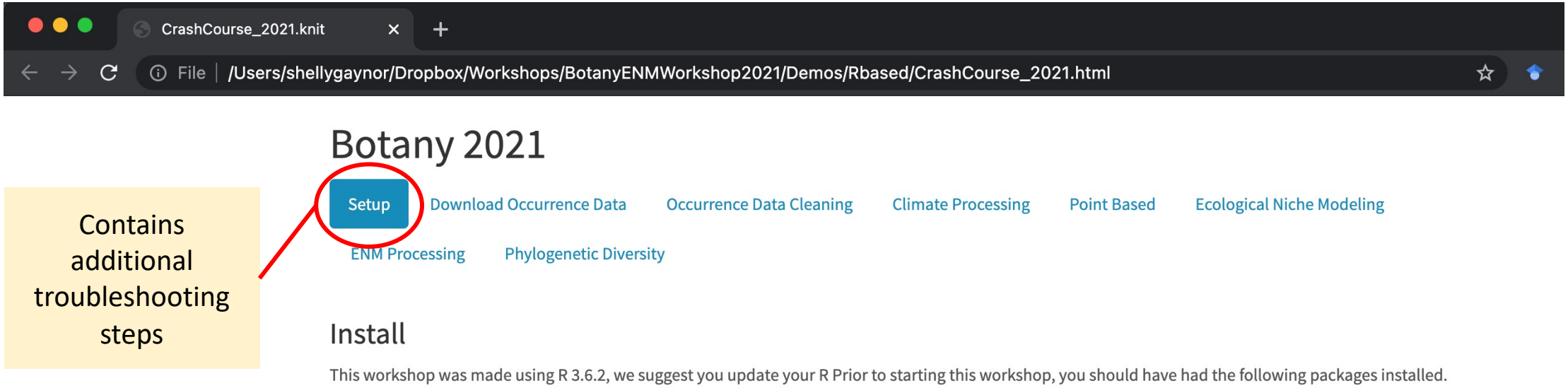
iDigBio
Integrated Digitized Biocollections

Setup



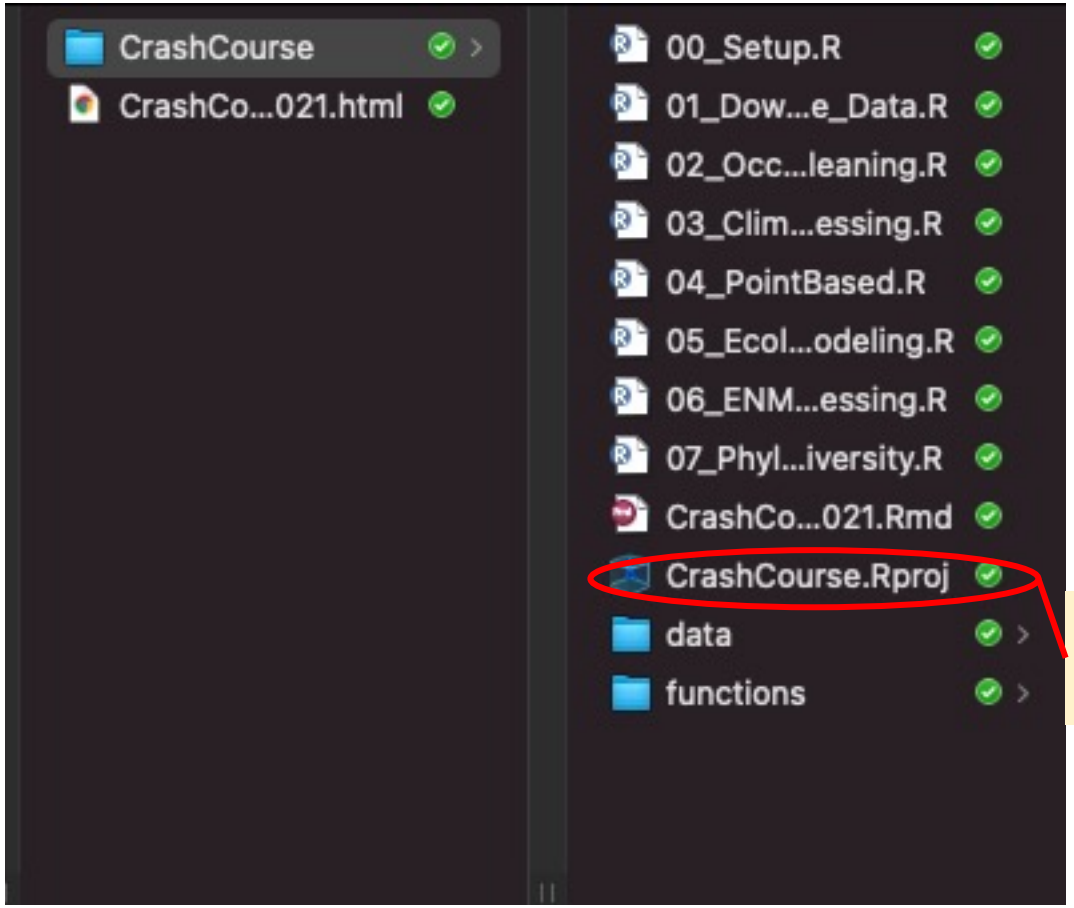
HTML Version of all R scripts

Setup

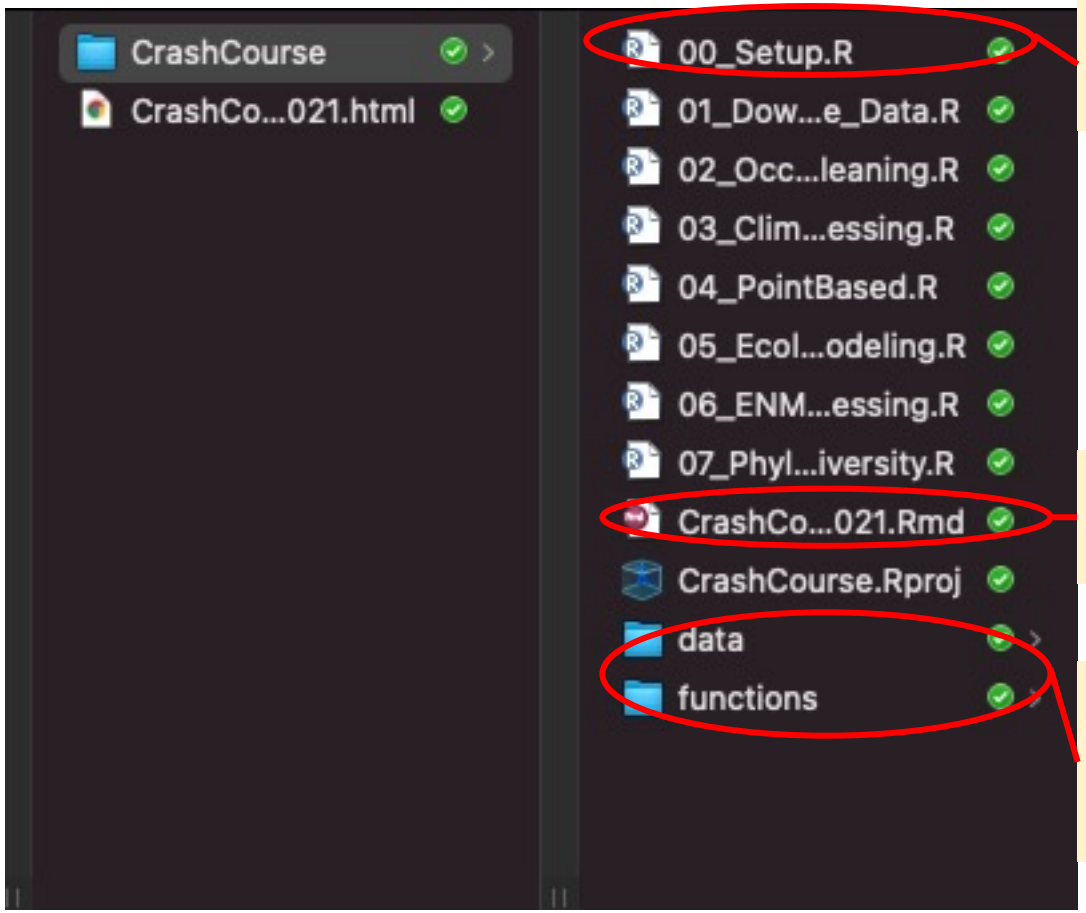


Contains additional troubleshooting steps

Setup


CrashCourse
CrashCo...021.html
00_Setup.R
01_Dow...e_Data.R
02_Occ...leaning.R
03_Clim...essing.R
04_PointBased.R
05_Ecol...odeling.R
06_ENM...essing.R
07_Phyl...iversity.R
CrashCo...021.Rmd
CrashCourse.Rproj
data
functions
Opens the R project

Setup



Run first!

How the R project was made

All data and functions needed for these scripts

Data Download

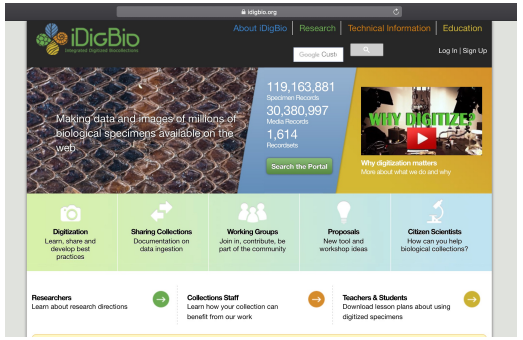
University of Florida

iDigBio
Integrated Digitized Biocollections

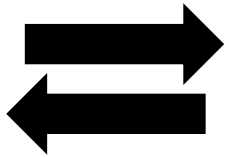
API = Application Programming Interface

- Allows users to interact with a system

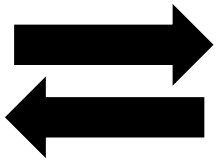
Web portal



Database



API



Interface

{JSON}

python

idigbio.org

iDigBio
Integrated Digitized Biocollections

About iDigBio | Research | Technical Information | Education

ENHANCED BY

Log In | Sign Up

Making data and images of millions of biological specimens available on the web

121,428,342
Specimen Records

31,871,262
Media Records

1,621
Recordsets

Search the Portal

Why digitization matters
More about what we do and why

Digitization
Learn, share and develop best practices

Sharing Collections
Documentation on data ingestion

Working Groups
Join in, contribute, be part of the community

Proposals
New tool and workshop ideas

Citizen Scientists
How can you help biological collections?

Researchers
Learn about research directions

Collections Staff
Learn how your collection can benefit from our work

Teachers & Students
Download lesson plans about using digitized specimens

iDigBio API

- Multiple ways to access the API:

API Name	Info
Search API	ridigbio R package <100,000 records
Download API	>100,000 records
Record API	Single record
Media API	Single record

biodiversity-specimen-data/**specimen-data-use-case**

R based

"../CrashCourse/CrashCourse.Rproj"

- Navigate to 01_Download_Occurence_Data.R

Load Packages

```
library(dplyr)
library(tidyr)
library(plyr)
library(spocc)
library(ridigbio)
library(tibble)
library(rbison)
```

Load functions

This is a function I created with Natalie Patten. It will be part of her R package gatoRs (Geographic And Taxonomic Occurrence R-based Scrubbing).

```
source("functions/DownloadingDataMore.R")
```

eBird VertNet Atlas of Living Australia ala.org.au OBIS OCEAN BIODIVERSITY INFORMATION SYSTEM Berkeley Ecoinformatics Engine

An open API serving UC Berkeley's Natural History Data

Downloading data using ridigbio

- First, we are searching for the species Galax urceolata
- Next, download occurrence records for the family Diapensiaceae

Search for the species Galax urceolata.

```
iDigBio_GU <- idig_search_records(rq=list(scientificname="Galax
 urceolata"))
```

Search for the family Diapensiaceae.

```
iDigBio_GU_family <- idig_search_records(rq=list(family="Diapen
siaceae"), limit=1000)
```

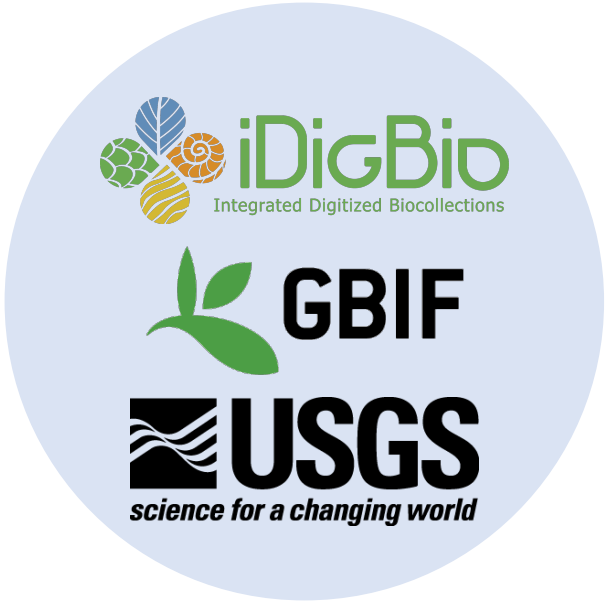
Records only in North America

```
rq_input <- list("scientificname"=list("type"="exists"),
                 "family"="Diapensiaceae",
                 geopoint=list(
                   type="geo_bounding_box",
                   top_left=list(lon = -98.16, lat = 48.92),
                   bottom_right=list(lon = -64.02, lat = 23.06)
                   )
                 )
```

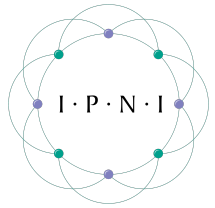
Search using the input you just made

```
iDigBio_GU_family_USA <- idig_search_records(rq_input, limit=10
00)
```

Data download using spocc_combine



- To pull data from GBIF, BISON, and iDigBio for a set of synonyms
- Identifying synonyms:
  - Taxonomic Name Resolution Service
    - https://tnrs.biendata.org/
    - Used in soltislab/BotanyENMWorkshops 2020
  - R package taxize
    - 20 sources for synonyms
    - https://docs.ropensci.org/taxize/



Data download using spocc_combine

Make synonym lists

Object

List of strings

```
Shortia_galacifolia <- c("Shortia galacifolia", "Sherwoodia gal
acifolia")
Galax_urceolata <- c("Galax urceolata", "Galax aphylla")
Pyxidanthera_barbulata <- c("Pyxidanthera barbulata","Pyxidanth
era barbulata var. barbulata")
Pyxidanthera_brevifolia <- c("Pyxidanthera brevifolia", "Pyxida
nthera barbulata var. brevifolia")
```

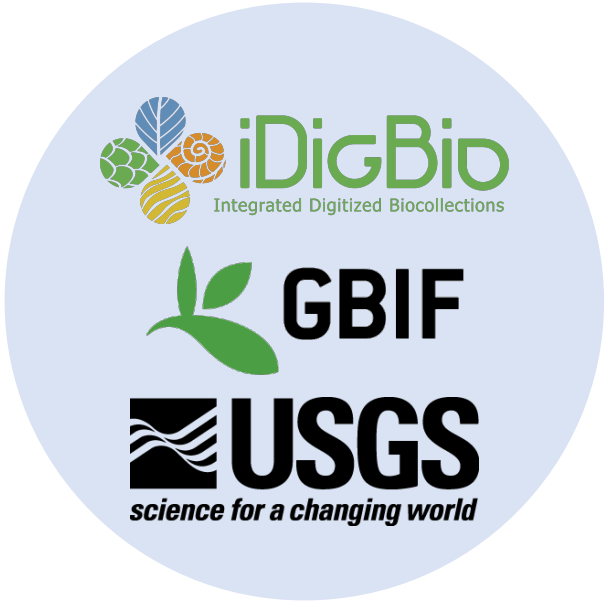
Data download using spocc_combine

spocc_combine(Synonym list , Path to save csv file)

```
spocc_combine(Shortia_galacifolia, "data/download/raw/Shortia_galacifolia_raw_20210614.csv")
spocc_combine(Galax_urceolata, "data/download/raw/Galax_urceolata_raw_20210614.csv")
spocc_combine(Pyxidanthera_barbulata, "data/download/raw/Pyxidanthera_barbulata_raw_20210614.csv")
spocc_combine(Pyxidanthera_brevifolia, "data/download/raw/Pyxidanthera_brevifolia_raw_20210614.csv")
```

Column	Description
name	scientific name, http://rs.tdwg.org/dwc/terms/scientificName
basis	basis of record, http://rs.tdwg.org/dwc/terms/basisOfRecord
date	event data, http://rs.tdwg.org/dwc/terms/eventDate
institutionID	institution ID, http://rs.tdwg.org/dwc/terms/institutionID
collectionCode	collection code, http://rs.tdwg.org/dwc/terms/collectionCode
collectionID	collection ID, http://rs.tdwg.org/dwc/terms/collectionID
country	country, http://rs.tdwg.org/dwc/terms/country
county	county, http://rs.tdwg.org/dwc/terms/county
state	stateprovince, http://rs.tdwg.org/dwc/terms/stateProvince
locality	http://rs.tdwg.org/dwc/terms/locality or http://rs.tdwg.org/dwc/terms/verbatimLocality
Latitude	http://rs.tdwg.org/dwc/terms/decimalLatitude
Longitude	http://rs.tdwg.org/dwc/terms/decimalLongitude
ID	idigbio = uuid, gbif = key, bison = occurrenceID
coordinateUncertaintyInMeters	http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters
habitat	http://rs.tdwg.org/dwc/iri/habitat
prov	indicates who provided the data: gbif, bison, or idigbio

Other considerations:



- Download from multiple repositories
- Identify flagged data
- Identify data for georeferencing

Other considerations:

- Download from multiple repositories
- Identify flagged data
- Identify data for georeferencing

mgaynor1/CURE-FL-Plants