CoastGIS 2021

Presentation theme and type:
Data Analyses / Analytical approaches;
Climate change; Biodiversity.

IOI International Ocean Institute

'Advancing Sustainable Coastal Planning through Effective Use of Open-Access Biodiversity Information Systems'

Where getting names right always matters!

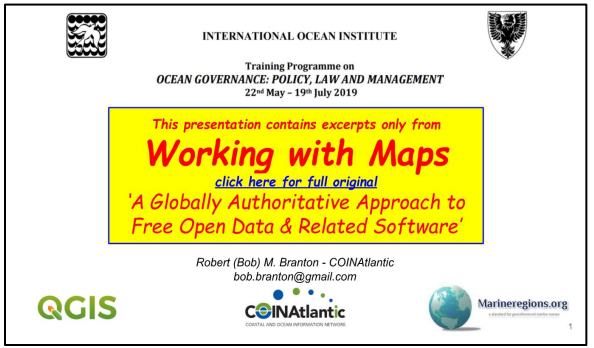
Robert M. Branton
% International Ocean Institute - Canada, Halifax, Nova Scotia, Canada
bob.branton@gmail.com

Outline

- International Ocean Institute training
  - public domain legacy data
  - training challenges
- Ecosystems' Services
  - Invasive species
  - What's in a name
- R-Spocc & ggplot
  - selection & enrichment
  - sample outputs
- R-Shiny - dashboard
- Discussion
  - features and benefits
  - limitations
  - software components
- Appendices
  - synthetic data objects
  - list of Links

Brief history of biodiversity information ...

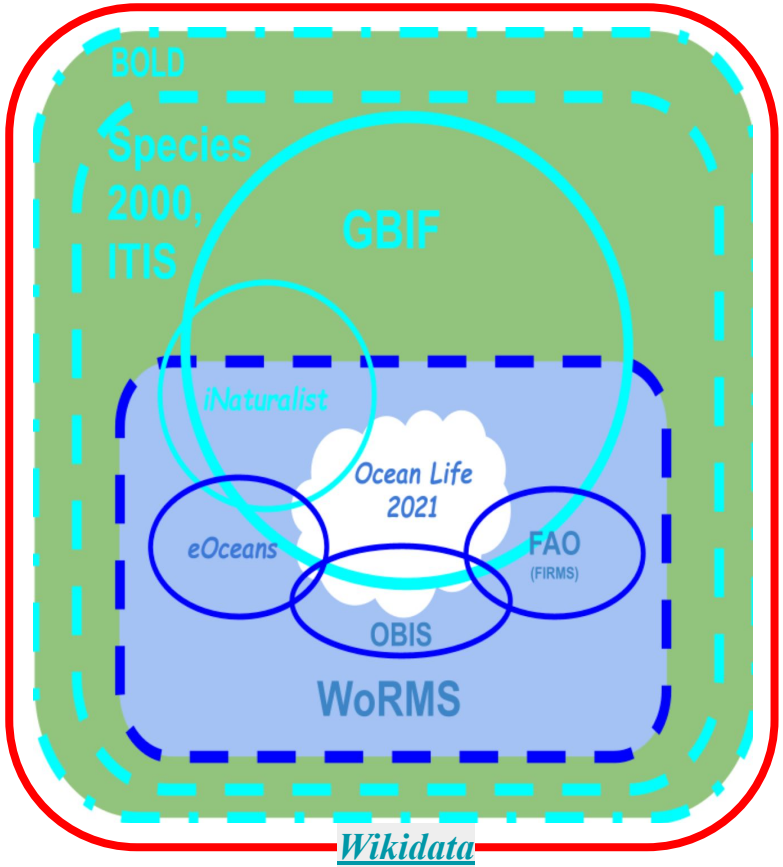
... US Ex 1838-42 and Challenger 1872-76 global ocean expeditions, Convention on Biological Diversity 1992, Census of Marine Life 2001-10, IODE Steering Group for OBIS 2011 ...



... IOI Canada training 1981-, 'howto' demos of free open software, sample maps using Global Biodiversity Information Facility & World Data Base of Protected Areas ...

International Ocean Institute training June ‘21

Provide provide ocean governance trainees and citizen scientists with globally relevant training materials demonstrating preparation and use of authoritative species lists as starting points for casual internet browsing and for querying public accessible global biogeographic information systems.

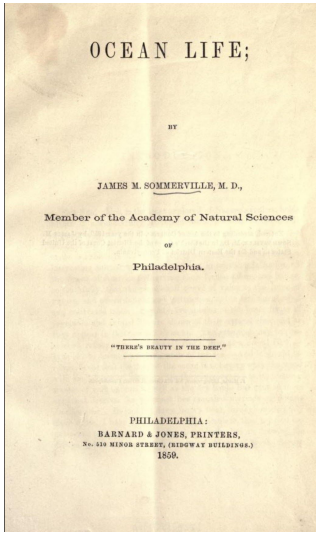


See IOI Training

See Oceanlife-2021/domains...

Public domain legacy data

in the form of PDF files may be downloaded, shared, and remixed without restriction. *Ocean Life - 1859 by J.M. Sommerville,* appearing on the Internet Archive Website, contains a list of 73 scientific species names used in IOI training ...



OCEAN LIFE;

BY

JAMES M. SOMMERVILLE, M. D.,

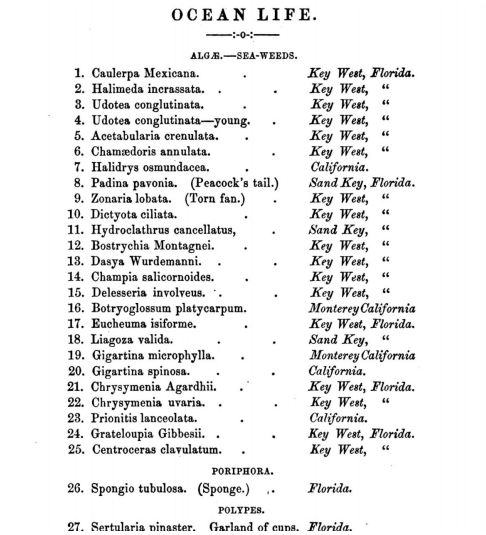
Member of the Academy of Natural Sciences

OF

Philadelphia.

"THERE'S BEAUTY IN THE DEEP."

PHILADELPHIA:
BARNARD & JONES, PRINTERS,
No. 510 MINOR STREET, (RIDGWAY BUILDING.)
1859.



OCEAN LIFE.

—:-o-:—

ALGÆ.—SEA-WEEDS.

1. Caulerpa Mexicana. . *Key West, Florida.*
2. Halimeda incrassata. . . *Key West,* "
3. Udotea conglutinata. . *Key West,* "
4. Udotea conglutinata—young. . *Key West,* "
5. Acetabularia crenulata. . *Key West,* "
6. Chamædoris annulata. . *Key West,* "
7. Halidrys osmundacea. . *California.*
8. Padina pavonia. (Peacock's tail.) *Sand Key, Florida.*
9. Zonaria lobata. (Torn fan.) . *Key West,* "
10. Dictyota ciliata. . *Key West,* "
11. Hydroclathrus cancellatus, . *Sand Key,* "
12. Bostrychia Montagnei. . *Key West,* "
13. Dasya Wurdemanni. . . *Key West,* "
14. Champia salicornoides. . *Key West,* "
15. Delesseria involveus. . . *Key West,* "
16. Botryoglossum platycarpum. *Monterey California*
17. Eucheuma isiforme. . *Key West, Florida.*
18. Liagoza valida. . . *Sand Key,* "
19. Gigartina microphylla. . *Monterey California*
20. Gigartina spinosa. . . *California.*
21. Chrysymenia Agardhii. . *Key West, Florida.*
22. Chrysymenia uvaria. . . *Key West,* "
23. Prionitis lanceolata. . *California.*
24. Grateloupia Gibbesii. . . *Key West, Florida.*
25. Centroceras clavulatum. . *Key West,* "

PORIPHORA.

26. Spongio tubulosa. (Sponge.) . *Florida.*

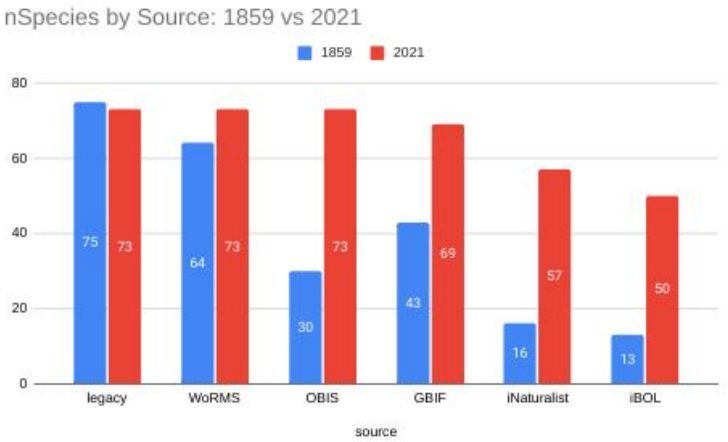
POLYPES.

27. Sertularia pinaster. Garland of cups. *Florida.*

Comma separated text.csv file were derived from images using Google Workspace software tools ...

```
group,key,scientificName,commonName,locality,regi
on,etc
ALGAE-SEA-WEEDS
,1,Caulerpa Mexicana,,Key West,Florida
,2,Halimeda incrassata,,Key West,Florida
,3,Udotea conglutinata,,Key West,Florida  ...
```

World Registry of Marine Species (WORMS) accepted names were shown to greatly improve query results ...



See Oceanlife-2021/legacy...

See Oceanlife-2021/worms

training challenges

Data are not all in one place ..

- 63 of 78m OBIS obs (~80%) and 13 of 40m iNat Research Grade obs (~33%) not on GBIF

Scientific names change over time

- 55 of 73 Ocean Life 1859 scientificNames required updating …

WoRMS/OBIS.authority <> GBIF.authority

- 26 of 68 GBIF.authority contained abbreviations, 14 of 68 GBIF.authority did not include year and so rejected WoRMS/OBIS.scientificName that did contain dates

Too easy to **not** get comparable data from GBIF & OBIS

- GBIF.occurenceStatus = Everything | Present | Absent i.e. default is Everything whereas OBIS.occurenceStatus=Present is the only option i.e. no choice given

iNaturalist taxa sometimes not found in GBIF and so required using WoRMS scientificNames (i.e. 5 of 57)

Inconsistent approach between portals regarding data downloads & citations

Ocean Life scale OBIS and GBIF downloads too large for Google Sheets

- iNaturalist is only portal offering column download selection
- can't exclude collections/networks (i.e. iNaturalist, OBIS) from GBIF downloads
- not easy to detect and remove duplicate occurrences

Conclusion: try focusing on shorter species list using R ...

See Oceanlife-2021/challenges...

Ecosystems' Services

Understanding the roles of individual species in ecosystems' provision of services to humans is a vital part of cooperative international decision making for sustainable planning.

For example: **Invasive Species**

"Removing green crab proves effective in controlling invasive species."

HALIFAX, 2016 … trap invented by a local fisherman was used to catch more than two million green crab [*Carcinus maenas* Linnaeus, 1758] from one estuary over a few years. Since then, eelgrass [*Zostera L.*] in that estuary has recovered by about 34 per cent and soft-shell clam [*Mya arenaria* Linnaeus, 1758] populations are on the rise ...

Atlantic CTV News

What's in a name? Wikidata, World Register of Marine Species#History, Taxonomy (biology)#Taxonomic descriptions, Taxonomy mnemonic

R - Spocc, ggplot - data selection & enrichment ...

* https://docs.ropensci.org/spocc/

Providers

ala - Atlas of Living Australia*
bison - Biodiversity Information Serving Our Nation [USA]*
bold - Barcode of Life Data System
gbif - Global Biodiversity Information System*
idigbio - Integrated Digitized Biocollections*
inat - iNaturalist*
obis - Ocean Biodiversity Information System*

Data Scrubbing ...

Dealing with duplicates ...

*ocean.si.edu/ocean-life/5-invasive-species-you-should-know

Species

Carcinus maenas [European Green Crab]*
Caulerpa taxifolia [Killer Algae]*
Codium fragile [Dead Man's Fingers]
Dreissena polymorpha [Zebra Mussel]*
Mnemiopsis leidyi [Sea Walnut]*
Pterois volitans [Lion Fish]
Rapana venosa [Veined Rapa Whelk]*

Provided in a csv file ...:
qNames - Worms accepted sName
cNames - common name,
sColors - color
wikidata - taxon code reference
fImages - link to Wikidata image

-Results saved as *invasives*.rda file ...
-Easy to make and use other lists …
OceanLife, MarinePlants

https://www.marineregions.org/

FAO

Boundaries of the FAO Fishing Areas

See:https://www.marineregions.org/sources.php#fao

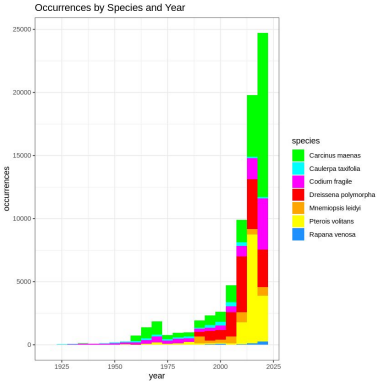
EEZ

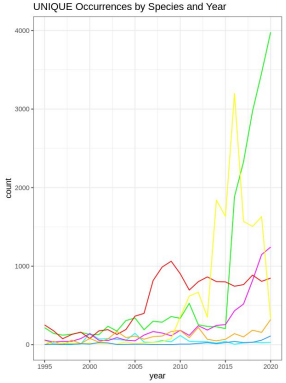
Marineregions: the intersect of the Exclusive Economic Zones and IHO areas

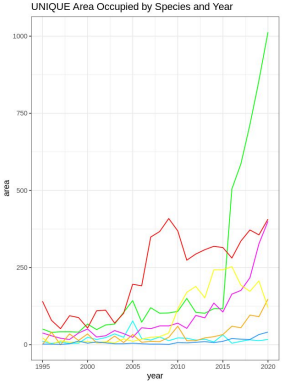
See:https://www.marineregions.org/sources.php#ihoeez

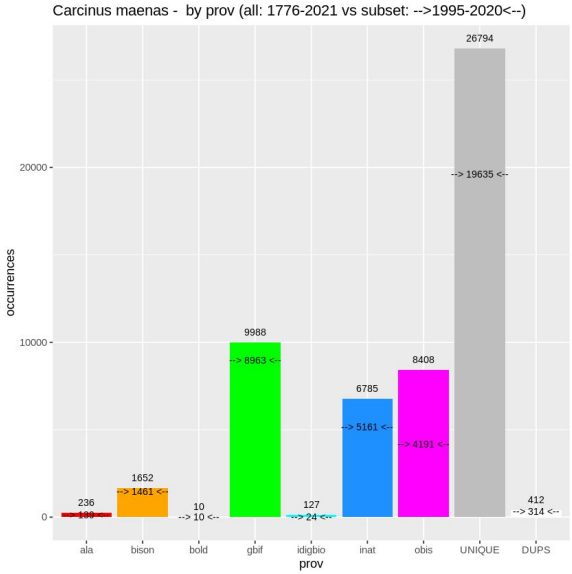
Sample outputs (total occurrences by species & year) ....

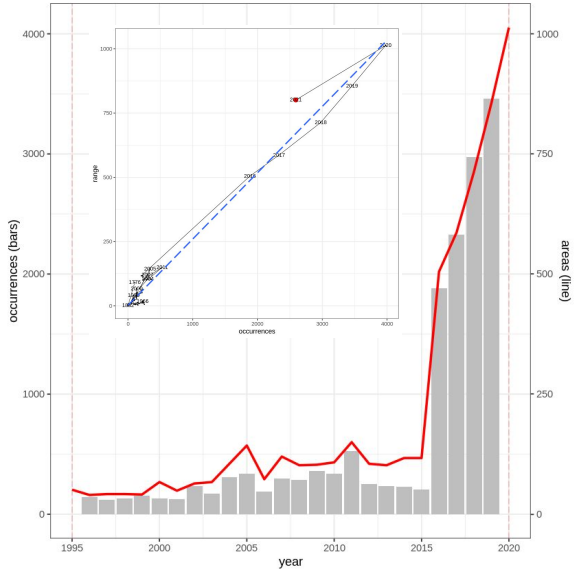
	ala	bison	bold	gbif	idigbio	inat	obis	UNIQUE	DUPS	total
Carcinus maenas	236	1652	10	9988	127	6785	8408	26794	412	54412
Caulerpa taxifolia	501	248	0	1355	403	55	637	2432	767	6398
Codium fragile	454	1268	0	6006	895	2341	759	10345	1378	23446
Dreissena polymorpha	0	4717	20	9979	514	1284	769	15185	2098	34566
Mnemiopsis leidyi	0	87	0	1166	12	262	2314	3479	362	7682
Pterois volitans	515	1213	0	9999	282	1523	2602	15175	959	32268
Rapana venosa	0	0	1	272	10	191	178	627	25	1304
total	1706	9185	31	38765	2243	12441	15667	74037	6001	160076

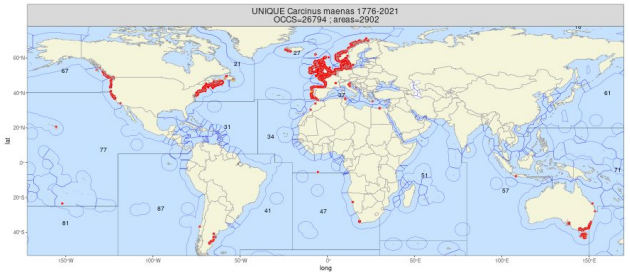


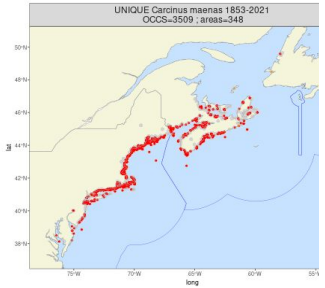








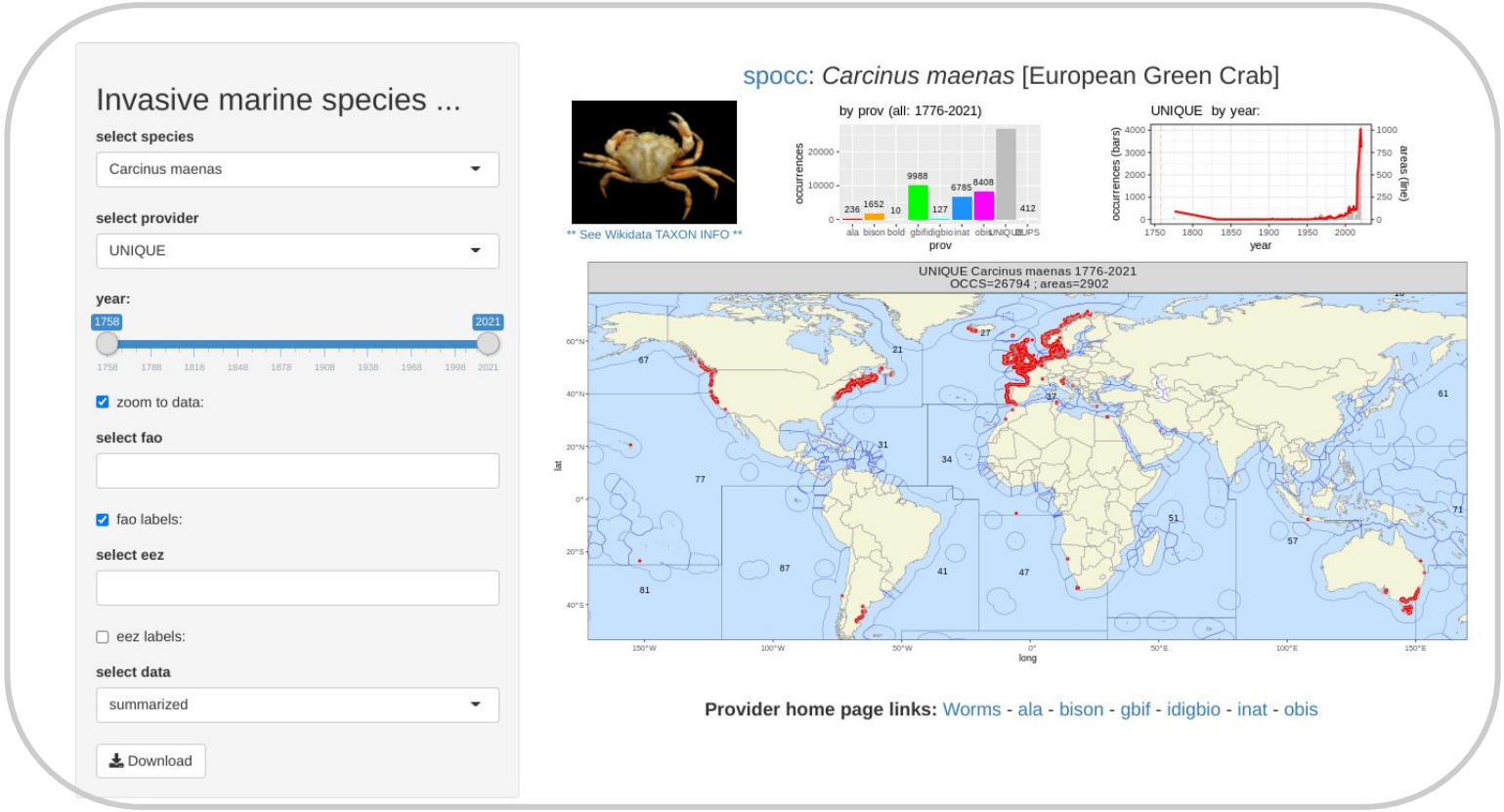




[See: ggplot2]

R-Shiny: ControlPanel / Dashboard

Live Access: https://rmbranto.shinyapps.io/invasives/



See: ShinyApps

Discussion

Features and benefits:

- Queries species occurrence data from available providers using list of Worms accepted names
- Generates synthetic UNIQUE / DUPlicate data objects from available provider data
- Filter and zoom maps and graphs using list of Providers, date range, FAO fishing Area and EEZs
- Deep links to live provider portals and Wikidata taxon info for selected species
- ...

Limitations

- Queries are limited to 9999 occurrence records per species and provider
- bison and inigbio providers no not support deep linking by species ...

Software Components:

- R - highly programmable computer environment, offering a wide variety of statistical and graphical techniques, as Free Software under GNU General Public License;
- Jupyter: combines software code, computational output and explanatory text in a single document;
- Spocc - (SPecies OCCurrence) query and collect species occurrence data from many sources, as well as creating unified outputs;
- ggplot - a general scheme for data visualization;
- Shiny: build interactive web apps straight from R.

Appendices: synthetic data objects

```
species.style
```

A data.frame: 7 × 5

Names	cNames	sColors	wikidata	flmages
<chr>	<chr>	<chr>	<chr>	<chr>
Carcinus maenas	European Green Crab	green	Q27779	https://upload.wikimedia.org/wikipedia/commons/1/17/Carcinus_maenas.jpg
Caulerpa taxifolia	Killer Algae	cyan	Q310961	https://upload.wikimedia.org/wikipedia/commons/e/e7/CaulerpaTaxifolia.jpg
Codium fragile	Dead Mans Fingers	magenta	Q2712208	https://upload.wikimedia.org/wikipedia/commons/e/ed/Codiumfragile.jpg
Dreissena polymorpha	Zebra Mussel	red	Q752130	https://upload.wikimedia.org/wikipedia/commons/thumb/a/a9/Dreissena_polymorpha3.jpg/220px-Dreissena_polymorpha3.jpg
Mnemiopsis leidyi	Sea Walnut	orange	Q133630	https://upload.wikimedia.org/wikipedia/commons/thumb/1/1e/Sea_walnut%2C_Boston_Aquarium.jpg/220px-Sea_walnut%2C_Boston_Aquarium.jpg
Pterois volitans	Lion Fish	yellow	Q824672	https://upload.wikimedia.org/wikipedia/commons/thumb/b/bf/Pterois_volitans_Manado-e_edit.jpg/220px-Pterois_volitans_Manado-e_edit.jpg
Rapana venosa	Veined Rapa Whelk	dodgerblue	Q139053	https://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Rapana_Black_Sea_2008_G1.jpg/220px-Rapana_Black_Sea_2008_G1.jpg

```
head(prov.keys)
```

A data.frame: 6 × 7

	species	ala	bison	gbif	idigbio	inat	obis
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	Carcinus maenas	284294	98734	5178595	5178595	52523	107381
2	Caulerpa taxifolia	83052	6974	2643172	2643172	50919	144476
3	Codium fragile	84082	6897	5272096	5272096	67555	370562
4	Dreissena polymorpha	NA	81339	2287072	2287072	116340	181566
5	Mnemiopsis leidyi	NA	53917	2501248	2501248	180788	106401
6	Pterois volitans	189453	166883	2334438	2334438	47280	159559

```
head(df.prov[df.prov$DUPS>2&df.prov$species=='Carcinus maenas',])
```

A data.frame: 6 × 10

	prov	species	year	date	OCCS	DUPS	longitude	latitude	eez	fao
	<chr>	<chr>	<int>	<date>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<chr>
8687	UNIQUE	Carcinus maenas	2018	2018-09-28	1	3	-122.30957	37.86517	USA	99
11319	UNIQUE	Carcinus maenas	2017	2017-07-25	2	4	-127.76165	51.48631	CAN	67
11320	DUPS	Carcinus maenas	2017	2017-07-25	6	3	-127.76165	51.48631	CAN	67
23906	UNIQUE	Carcinus maenas	2016	2016-08-18	1	3	-60.33113	46.80721	CAN	99
27236	UNIQUE	Carcinus maenas	2017	2017-10-13	1	3	-63.93700	44.51300	CAN	21
30240	UNIQUE	Carcinus maenas	2017	2017-05-31	1	3	-66.89060	45.03820	CAN	21

```
head(df.exp)
```

A data.frame: 6 × 6

	species	prov	year	eez	fao	area
	<chr>	<chr>	<int>	<chr>	<int>	<chr>
1	Carcinus maenas	ala	1916	AUS	99	-38 145.1....
2	Carcinus maenas	ala	1916	AUS	99	-38 145.1....
3	Carcinus maenas	ala	1955	AUS	99	-37.6 149.8..
4	Carcinus maenas	ala	1957	GBR	99	50.3 -4.1....
5	Carcinus maenas	ala	1958	AUS	57	-38.4 144.9..
6	Carcinus maenas	ala	1958	AUS	57	-38.7 145.7..

```
prov.style
```

A data.frame: 9 × 4

prov	id	color	order
<chr>	<chr>	<chr>	<dbl>
UNIQUE	UNI	grey	8
DUPS	DUP	white	9
ala	ala	red	1
bison	bis	orange	2
bold	bol	yellow	3
gbif	gbi	green	4
idigbio	idi	cyan	5
inat	ina	dodgerblue	6
obis	obi	magenta	7

What I’m working on ...

- Brief Shiny animation
- Link to R source code
- List of Links used in this presentation