Exploratory Analyses II

Today's agenda:

- Discuss Assignment 1
- Discuss potential data sources to work with
- Continue working through 2.6 (if needed)
- Start to apply techniques from chapter 2 to dataset of interest

Assignment 1

Full information online (see Canvas/Github)

Assignment 1

Overview:

- Select a dataset (you want one that can ask multiple questions)
- Do some exploratory analyses (and document them)

Assignment 1

In this assignment, students will need to:

1. Select a dataset, provide a description, note where it came from
2. Read in their dataset
3. Note the class of the dataset itself (e.g., matrix, list, dataframe, etc.)
4. List the variables of interest and their classes (e.g., numeric, character, etc.)
5. Examine the structure of the dataset and correct any formatting mistakes
6. Provide summary statistics for variables of interest
7. Produce at least 3 different figures that convey different information
8. For the figures in 6, briefly describe what the figures show

Assignment 1: what you'll submit

Needs to document what you did and be reproducible

Either:

1) The output of an .Rmd file showing code and outputs
    a) See **example_R_markdown**/ for .Rmd file and potential output types, e.g. html, pdf

2) A .R file that can be run anywhere
    a) See **example_R_commented**/ for .R file with comment examples

Assignment 1: what you'll submit

Assignments can be submitted on canvas.

You can either:

a) Upload your file to Github and submit the link on Canvas

(require for grads, extra credit undergrads)

b) Upload the file on canvas

Due Friday, Sept. 19th before midnight.

Questions so far?

Getting data

- Collect it
  - pro: you know it best, can be tailored to a specific question
  - con: time-consuming, possibly expensive, probably requires cleaning
- Download it
  - pro: quick
  - con: less knowledge of collection, probably still needs cleaning
- Aggregate it
  - pro: data volume and breadth
  - con: even less knowledge of collection, lots of standardization and cleaning required

Where to look for data

Course Github site:

- Links to resources
- Data folder contains some options

Scientific Publications

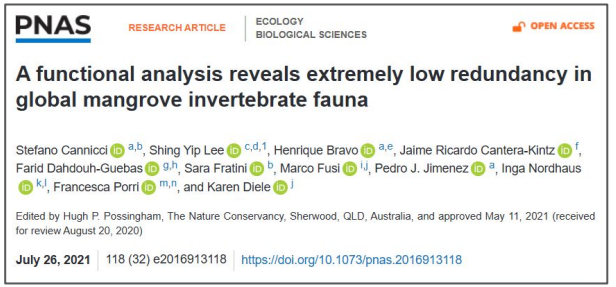
- Supplementary Information
- Data Publications (e.g., https://www.nature.com/sdata/)

Data repositories

- Dryad https://datadryad.org/
- TRY https://www.try-db.org/TryWeb/Home.php

Datasets of interest to this class

- Mangrove invert communities
- Marine microorganisms
- Deep sea megabenthos
- Bird Survey Data
- Coral populations
- ~~Rhino~~ Deer Movement

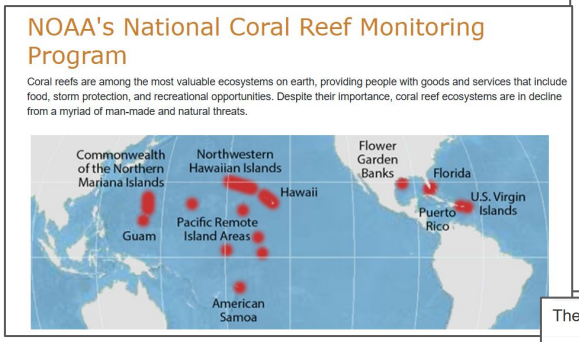

PNAS RESEARCH ARTICLE | ECOLOGY BIOLOGICAL SCIENCES | OPEN ACCESS

A functional analysis reveals extremely low redundancy in global mangrove invertebrate fauna

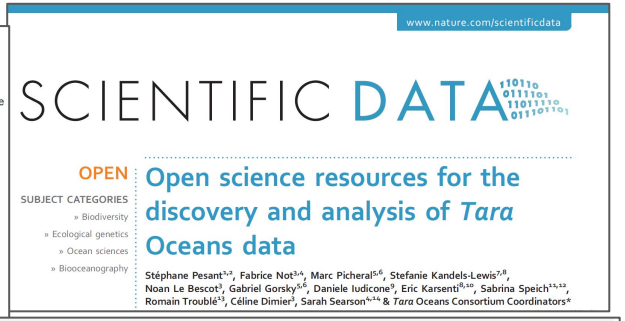
Stefano Cannicci [a,b], Shing Yip Lee [c,d,1], Henrique Bravo [a,e], Jaime Ricardo Cantera-Kintz [f], Farid Dahdouh-Guebas [g,h], Sara Fratini [b], Marco Fusi [i,j], Pedro J. Jimenez [a], Inga Nordhaus [k,l], Francesca Porri [m,n], and Karen Diele [j]

Edited by Hugh P. Possingham, The Nature Conservancy, Sherwood, QLD, Australia, and approved May 11, 2021 (received for review August 20, 2020)

July 26, 2021 | 118 (32) e2016913118 | https://doi.org/10.1073/pnas.2016913118


NOAA's National Coral Reef Monitoring Program

Coral reefs are among the most valuable ecosystems on earth, providing people with goods and services that include food, storm protection, and recreational opportunities. Despite their importance, coral reef ecosystems are in decline from a myriad of man-made and natural threats.


www.nature.com/scientificdata

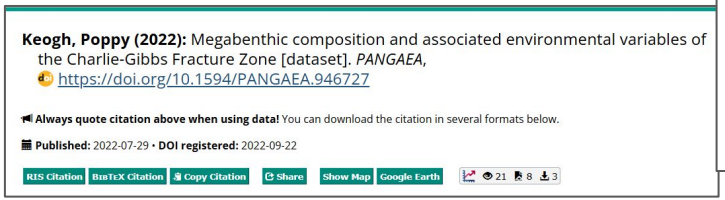
SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Biodiversity
» Ecological genetics
» Ocean sciences
» Biooceanography

Open science resources for the discovery and analysis of *Tara* Oceans data

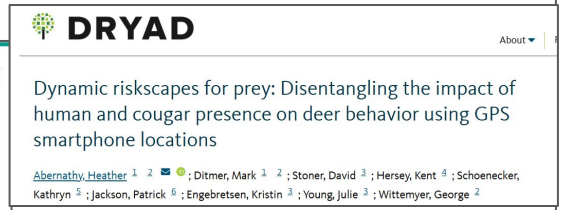
Stéphane Pesant[1,2], Fabrice Not[3,4], Marc Picheral[5,6], Stefanie Kandels-Lewis[7,8], Noan Le Bescot[3], Gabriel Gorsky[5,6], Daniele Iudicone[9], Eric Karsenti[8,10], Sabrina Speich[11,12], Romain Troublé[13], Céline Dimier[3], Sarah Searson[4,14] & *Tara* Oceans Consortium Coordinators*


Keogh, Poppy (2022): Megabenthic composition and associated environmental variables of the Charlie-Gibbs Fracture Zone [dataset]. *PANGAEA*, https://doi.org/10.1594/PANGAEA.946727

Always quote citation above when using data! You can download the citation in several formats below.

Published: 2022-07-29 • **DOI registered:** 2022-09-22

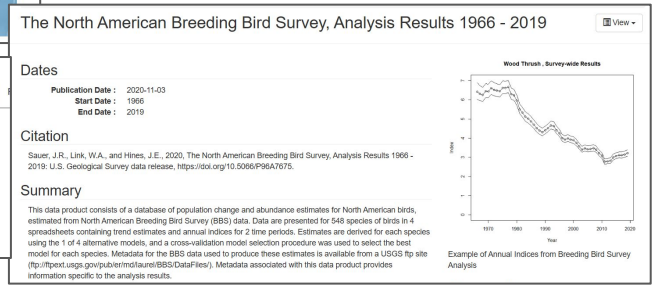
RIS Citation | BibTeX Citation | Copy Citation | Share | Show Map | Google Earth | 21 | 8 | 3


DRYAD

About

Dynamic riskscapes for prey: Disentangling the impact of human and cougar presence on deer behavior using GPS smartphone locations

Abernathy, Heather [1,2]; Ditmer, Mark [1,2]; Stoner, David [3]; Hersey, Kent [4]; Schoenecker, Kathryn [5]; Jackson, Patrick [6]; Engebretsen, Kristin [3]; Young, Julie [3]; Wittemyer, George [2]


The North American Breeding Bird Survey, Analysis Results 1966 - 2019

View

Dates

Publication Date : 2020-11-03
Start Date : 1966
End Date : 2019

Citation

Sauer, J.R., Link, W.A., and Hines, J.E., 2020, The North American Breeding Bird Survey, Analysis Results 1966 - 2019: U.S. Geological Survey data release, https://doi.org/10.5066/P96A7675.

Summary

This data product consists of a database of population change and abundance estimates for North American birds, estimated from North American Breeding Bird Survey (BBS) data. Data are presented for 548 species of birds in 4 spreadsheets containing trend estimates and annual indices for 2 time periods. Estimates are derived for each species using the 1 of 4 alternative models, and a cross-validation model selection procedure was used to select the best model for each species. Metadata for the BBS data used to produce these estimates is available from a USGS ftp site (ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/). Metadata associated with this data product provides information specific to the analysis results.

Example of Annual Indices from Breeding Bird Survey Analysis

Datasets of interest to this class

- General places to look for data noted on the course Github homepage
- Data folders on course Github
  - The “README” files in each folder contain useful information
  - May need to look up the data sources to understand all the column names.

Remainder of class:

- Continue working on 2.6
- Start working on Assignment 1

Next week:

- I’ll be out of town
- Do chapter 3
- Complete Assignment 1