Cayman Islands Species Abundance Analysis

Adam Bruce

09/04/2022

Abstract

Introduction

Since 1998, Lawrence University Professor Bart De Stasio has led a Marine Term trip to Grand Cayman Island in the Western Caribbean. The objective of this trip has been to document abundance and diversity of fish species, provide evidence towards the health of coral reefs surrounding the island, and to understand the ecological relationships between the fish and coral reef environments. To accomplish this, students on the trip used Reef Environmental Education Foundation (REEF) specification to collect data on 246 fish species at 12 different coral reefs. On each dive, data collection lasted approximately 20 to 50 minutes depending on diver air consumption and non-decompression limits (Timpe, 2018). One of the key variables related to the abundance and diversity of fish was Density Index (Den).

Den was recorded as a measure of how many of a certain species was present at each site during each dive. Ultimately, the scale of Den ranged from 0 to 4, and each whole number represented a REEF category of density. These four categories were Single (S), Few (F), Many (M), and Abundant (A). A Den score of 1 corresponded to S, 2 corresponded to Few, 3 represented Many, and the highest score of 4 was Abundant. Since there were often several dives at a single site, the Den scores for an individual species were averaged to calculate an overall Den for that site each year. REEF provides an outline for performing this calculation and for understanding the specification for each of the four categories. This outline can be seen in the image below.

REEF Density Index Categorization and Average Calculation

Density Index (Den)

Density Index (Den) - This is a measure of how many individuals of a species are observed based on a scale of 1-4. It is representative of the abundance category (1-4) which was most frequently recorded for the species when it was observed. Abundance category weights are **S**ingle=1, **F**ew=2, **M**any=3, and **A**bundant=4.

This weighted density average is calculated as:

$$\text{Den} = \frac{(S * 1) + (F * 2) + (M * 3) + (A * 4)}{(\text{Number of surveys in which species was observed})}$$

This number indicates which abundance category the species was most often recorded in when it was recorded. For example, Den=2.2 would be reflective of a species that was most often recorded in category 2 (Few) but since the density index is greater than 2, there were some abundances recorded for this species in the other, larger abundance categories (either category 3 or 4). The density index should be used as a density guide. In this survey method, area is not rigorously controlled. It should also be kept in mind that the density (Den) parameter is reflective of sighting distributions in the four

The data in this study was recorded every other year for twenty years. So, this information can be analyzed as a longitudinal study. As I proceed, I will use linear mixed effect models to investigate both how species density has changed within each dive site and how species density has changed between the dive sites. This information will help explain the ecology behind fish species on Grand Cayman Island. Finally, I will build a machine learning model to best predict a fish species based on OOBE/RMSE comparison metrics. This will allow future researchers to understand what species the might expect to find under certain conditions on the island.

Initial Data Wrangling

Initially, data consisted of 246 rows and 95 columns. However, the data set was not in tidy form as information on one variable, Density Index, was spread across many columns. These columns were also labeled by year and location, which meant they needed to be separated into individual variables. To tidy the data, I used

the concepts of pivoting and separating. First, I made the Density Index variable (DENSITY_INDEX) using the `pivot_longer` function. This function condensed the columns into one hence making the data set "longer". Next, I used the `seperate` function to make individual variable columns for the location and year each observation was collected. Finally, YEAR was coded as a character variable and I changed it to numeric. After this, the data set consisted of 22,632 rows and 5 columns. However, some of the Density Index values followed an outdated version of the REEF scale. So, values greater than the 4.0 threshold were removed.

```
## # A tibble: 6 x 5
##   SPECIES_NAME   FISH_ID  YEAR LOCATION DENSITY_INDEX
##   <chr>            <dbl> <dbl> <chr>            <dbl>
## 1 Blue Angelfish       1  1998 CH                   0
## 2 Blue Angelfish       1  1998 DG                   2
## 3 Blue Angelfish       1  1998 SB                   0
## 4 Blue Angelfish       1  1998 SC                   0
## 5 Blue Angelfish       1  1998 TF                   0
## 6 Blue Angelfish       1  2000 CP                   0
```

The `head` function shows that the five variables consist of two character variables in SPECIES_NAME and LOCATION, and three numeric variables in FISH_ID, YEAR, and PER_ABUND. Overall, 12 locations were studied throughout the duration of data collection. These sites include, from southeast clockwise to northeast, Bodden Bay Lagoon (BL), Beach Bay (BB), Smith's Cove (SC), Sunset House (SH), Sea View (SV), Casuarina Point (CP), Devil's Grotto (DG), Eden's Rock (ER), Cemetery Reef (CR), Turtle Farm (TF), Spanish Bay (SB), and Mangroves (MG). A majority of these sites were described and shown geographically in 2018 by Alec Timpe, a Marine Term student of Professor De Stasio. His dive site image and site description paper can be seen below.

Grand Cayman, B.W.I.
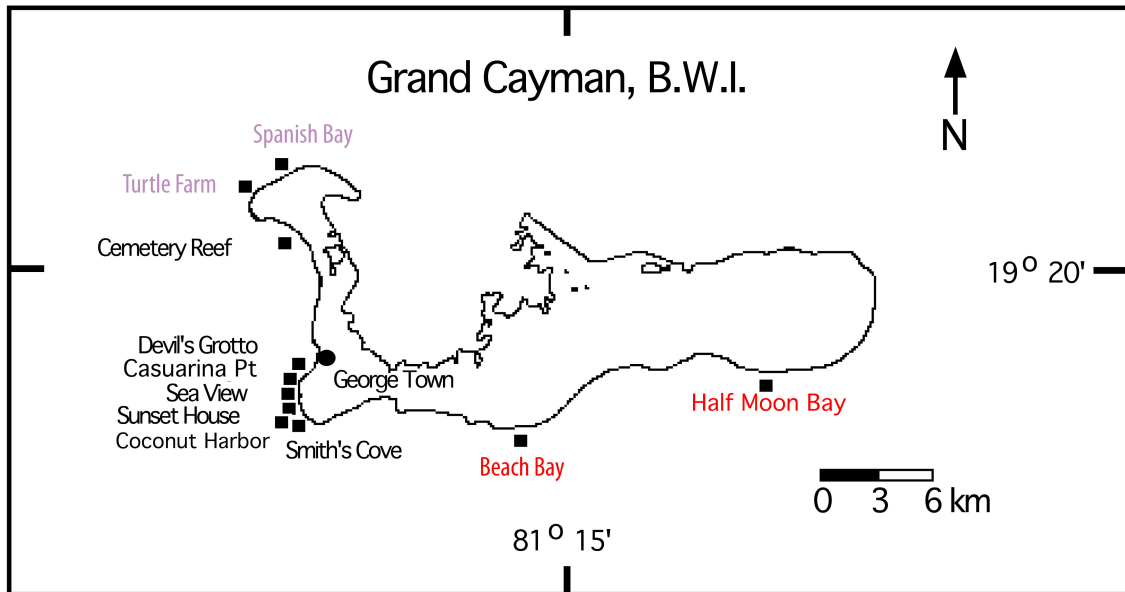
Timpe_Description

Exploratory Analysis

Table 1: Species With an Average Density Index Across All Dive Sites of 2.5 or Greater by Year

| YEAR	SPECIES_NAME	MEAN_DEN	STD_DENSITY
1998	Bicolor Damselfish	2.580000	0.4266146
1998	Blue Chromis	3.504365	0.2865079
1998	Fairy Basslet	3.404167	0.2389696
1998	Sergeant Major	2.525000	1.4777282
1998	Stoplight Parrotfish	2.592064	0.3545586
2000	Blackear Wrasse	2.779286	0.2107698
2000	Blue Chromis	3.352917	0.1647768
2000	Fairy Basslet	2.991607	0.2485778
2000	Rainbow Parrotfish	3.000000	0.0000000
2000	Redtail Parrotfish	2.635714	0.3439473
2000	Reef Squirrelfish	2.537559	0.2596139
2000	Squirrelfish	2.829583	0.1411313
2002	Bicolor Damselfish	2.666667	0.2581989
2002	Blue Tang	2.875000	0.3061862
2002	Brown Chromis	3.250000	0.2886751
2002	Fairy Basslet	2.500000	0.1490712
2002	Sergeant Major	2.694444	0.4002314
2002	Stoplight Parrotfish	2.652778	0.3000772
2002	Threespot Damselfish	2.583333	0.5845226
2004	Fairy Basslet	2.814286	0.1978305
2004	Reef Squirrelfish	2.604762	0.2449644
2004	Sunshinefish	3.248753	0.1618911
2006	Cubbyu	2.629464	0.5303945
2006	Fairy Basslet	2.635714	0.2607954
2006	Reef Squirrelfish	2.525893	0.3573870
2006	Shortstripe Goby	2.791667	0.2920915
2006	Sunshinefish	2.763988	0.3205049
2008	Creole Wrasse	2.583333	0.4364358
2008	Cubbyu	2.934829	0.3088982
2008	Fairy Basslet	2.664660	0.1762466
2008	Jolthead Porgy	2.712500	0.1774489
2008	Longfin Damselfish	2.741667	0.2522911
2008	Longjaw Squirrelfish	2.555458	0.2212560
2008	Queen Triggerfish	2.645518	0.3388357
2008	Reef Squirrelfish	2.736810	0.1692300
2008	Sunshinefish	2.701042	0.2361412
2010	Cubbyu	2.592593	0.5327543
2010	Fairy Basslet	2.769577	0.2873558
2010	Jolthead Porgy	2.527778	0.2089574
2010	Queen Triggerfish	2.529894	0.3790748
2010	Reef Squirrelfish	2.611111	0.3929942
2010	Sunshinefish	3.156878	0.2603973
2010	Yellowtail Reeffish	2.875000	0.4432026
2012	Blackcap Basslet	2.600000	0.5163978
2012	Blue Tang	2.841667	0.2734383
2012	Cubbyu	2.763095	0.3167875
2012	Jolthead Porgy	2.700000	0.2334062
2012	Longfin Damselfish	2.751905	0.2874064
2012	Reef Squirrelfish	2.564762	0.3051854
2012	Scrawled Cowfish	2.900000	0.3162278
2012	Sunshinefish	3.199405	0.1895906
2012	Yellowtail Reeffish	2.501852	0.2923141
2014	Bicolor Damselfish	2.569405	0.2680085
2014	Blue Chromis	3.482143	0.1401024
2014	Blue Tang	2.981905	0.2941677
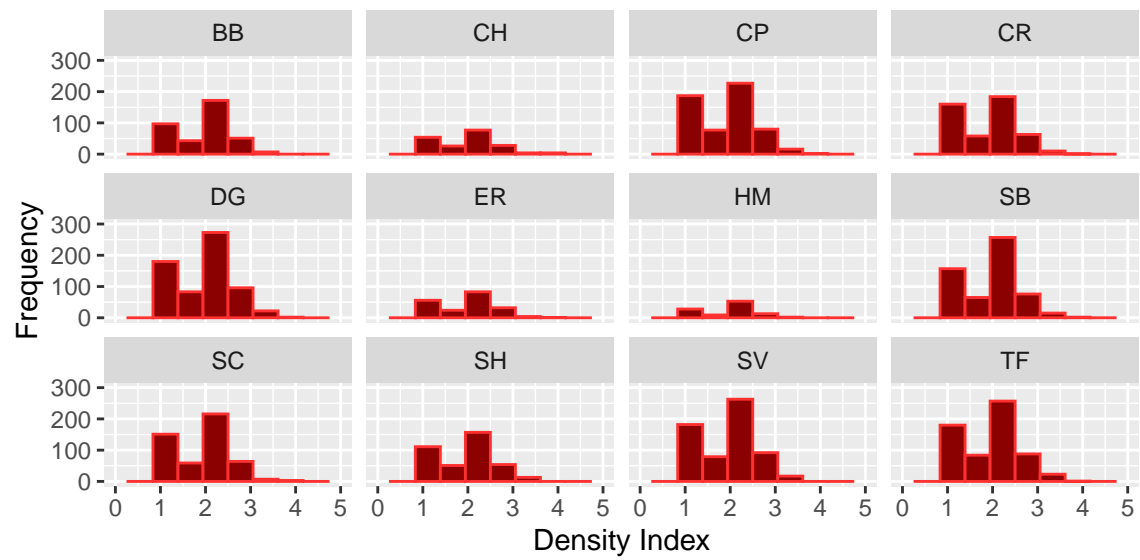
Figure 1: REEF Fish Density Index at Twelve Sampling Location on Grand Cayman Island