Advances in Biology Laboratory Education
Publication of the Association for Biology Laboratory Education
Volume 42, Article 15, 2022

From Statistics to Ecological Analyses: An Application of the New York City East River Ichthyology Database (ERID) Using Spreadsheets

Peter J. Park[1], Devin M. Gorsen[2], Christopher D. Girgenti[3], Isa G. Del Bello[4], Christina M. Tobitsch[4], Jacqueline R. Wu[3], and Andrew J. Hong[5]

[1]Farmingdale State College, Department of Biology, Farmingdale, NY 11735 USA.
[2]New York State Office of Parks, Recreation, and Historic Preservation, New York, NY 10027 USA.
[3]Randall's Island Park Alliance, New York, NY 10035 USA.
[4]Brooklyn Bridge Park Conservancy, Brooklyn, NY 11201 USA.
[5]Independent Researcher and Software Developer from Bayside, NY 11361 USA.

(**parkp@farmingdale.edu; Devin.Gorsen@parks.ny.gov; christopher.girgenti@randallsisland.org; idelbello@brooklynbridgepark.org; ctobitsch@brooklynbridgepark.org; jackie.wu@randallsisland.org; andrew.j.hong@icloud.com**)

Undergraduate biology students often have limited opportunities in a classroom setting to apply statistics quickly and effectively to real, current biological datasets. To address this concern, we developed adaptable, customized Microsoft® Excel spreadsheets that enable instructors to teach students how to statistically explore, visualize, and analyze biological phenomena. To demonstrate the utility of these spreadsheets, data is provided from the East River Ichthyology Database (ERID, https://eastriverfishproject.org/), a database that hosts real, current, and regularly updated fish and water quality data from New York City's iconic East River strait, contributed by community scientists. ERID was initially developed in 2019 by a grant-funded partnership of environmental education organizations and academic institutions documenting fish diversity in NYC. In the present work, ERID data includes fish and water quality data, collection methods, and general locations from 2019. Broader findings and analyses from 2019 were published in Park *et al*. 2020, providing a convenient reference for instructors to contextualize student learning. Using the provided ERID spreadsheets, students can calculate descriptive statistics, generate a frequency histogram, calculate and visualize correlation and linear regression, and calculate and interpret Shannon-Wiener index values. We provide a sample activity focused on 2019 data from ERID on the popular New York sportfish striped bass (*Morone saxatilis*) to demonstrate the use of the spreadsheets together with a recently developed online web tool, *Data Explorer* by Howard Hughes Medical Institute (HHMI) BioInteractive. This activity also aims to encourage a deep appreciation for integrating community science and fish biology in general.

Key Words: ichthyology, fish biodiversity, ecology, statistics, East River, East River Ichthyology Database, ERID

Link To Supplemental Materials: https://doi.org/10.37590/able.v42.sup15

Introduction

Undergraduate biology students often have limited opportunities in a classroom setting to apply statistics quickly and effectively to real, current biological datasets. Students usually work through outdated, abstract datasets by hand or use software programs that treat statistical calculations like a mysterious “black box” without the transparency of seeing necessary steps and calculations. To address these concerns, we developed adaptable customized Microsoft® Excel (Microsoft Corporation, 2018) spreadsheets that will enable instructors to teach students basic statistics to explore, visualize, and analyze biological phenomena. Real data is provided from the East River Ichthyology Database (ERID, https://eastriverfishproject.org/). This database hosts current and regularly updated fish and water quality data from New York City’s East River strait contributed entirely by community scientists, including environmental educators, researchers, naturalists, and recreational and commercial anglers. An online web tool, *Data Explorer* (HHMI BioInteractive, 2021), can also be employed to explore ERID data. To demonstrate that the spreadsheets and *Data Explorer* can complement each other to achieve learning outcomes, we provide a sample activity focused on the fisheries status of a popular New York sportfish striped bass (*Morone saxatilis*). This activity also aims to develop in students a deep appreciation for the importance of surveying efforts by community scientists and fish biologists.

Student Outline

I. Introduction

In this lab, you will explore basic statistics and ecological analyses using a variety of customized Microsoft® Excel spreadsheets and data derived from the New York City East River Ichthyology Database (ERID, 2021, https://eastriverfishproject.org/). ERID, which hosts real, current, and regularly updated fish and water quality data from NYC, was initially developed in 2019 through a grant-funded partnership of environmental education organizations and academic institutions studying fish diversity in NYC's East River strait. The database includes fish and water quality data, collection methods, and general geographic locations of sampling sites from various programs and outings that utilized the efforts of innumerable community scientists.

II. Learning Objectives

Through proper use of the attached spreadsheets, students will be able to:

(i) Calculate descriptive statistics

(ii) Generate a frequency histogram

(iii) Calculate and visualize correlation graphs

(iv) Calculate and visualize linear regression graphs

(v) Calculate and interpret Shannon-Wiener Index values

III. Background

The East River Strait

Bordered by four of New York City's five boroughs (Brooklyn, Manhattan, Queens, and the Bronx), crossed by seven bridges, and overlaid by multiple automobile and subway tunnels, the East River is an iconic waterway with a complicated history of dynamic boom and bust periods of marine wildlife, with recent years exhibiting improving conditions with deindustrialization (Hurley, 1994; O'Neil *et al*., 2016; Steinberg, 2014; Stinnette *et al*., 2018; Waldman, 2013). Divided into an upper section and lower section by the Hell Gate, which restricts the water's flow through its narrow passages (Li *et al*., 2018), the East River lies at the heart of one of the most culturally, historically, and ecologically diverse estuaries in the world. Despite its name, New York City's East River is not a true river but rather a tidal strait that runs approximately 25 km from the northern edges of the Upper New York Harbor to the western reaches of the Long Island Sound. Geologically, it is a drowned river valley formed nearly 11,000 years ago by the retreat of the Wisconsin glaciers, the deposition of glacial debris, and subsequent inundation from rising seas. As sea levels rose, its estuarine substrate was formed by assorted glacial moraine, fine sediments, and landward erosion (Levinton and Waldman, 2006).

For the convenience of data acquisition and discussion, the East River strait was divided into 11 physical zones (Fig. 1) based on three general criteria: (i) location of an environmental education organization study site, (ii) boundaries identified by conspicuous human-made landmarks, and (iii) presence of public access (e.g., city park, state park, marina). These zones were not originally intended to separate ecologically contrasting habitats. Each zone generally has a similar benthic substrate of sand, mud, and rocks and experiences considerable currents during tide changes. However, the lower region of the strait is narrower with generally faster current speeds. In comparison, the upper region is wider, has many more coves, and includes a zone with a relatively undisturbed marsh area (Zone 3). The 11 zones, north to south, are detailed in Park *et al*. (2020).

Figure 1. East River Ichthyological Alliance (ERIA) East River Zones of Study. Each zone is bordered by commonly recognized city landmarks (e.g., bridges, tunnels). Illustration © Hannah Ahn.

East River Ichthyological Alliance (ERIA) and East River Ichthyology Database (ERID)

The data provided in this work is a subset of fish data collected in 2019 by the East River Ichthyological Alliance (ERIA), a network of community scientists invested in understanding the ecology and fauna of the East River. Details about this network and the full findings from 2019 were published in Park *et al*. (2020). One significant outcome of the 2019 ERIA research was the development of the East River Ichthyology Database (ERID). This database hosts real, current, and regularly updated fish and water quality data from New York City's East River strait (ERID, 2021).

IV. Methods and Data Collection - Data Analysis Spreadsheets

*A. Initial Exploration of Data Analysis Spreadsheets***.** All students have access to the following data analysis spreadsheets. Please download each onto your computer.

(i) Descriptive Statistics (See Appendix A)
(ii) Frequency Distribution (See Appendix B)
(iii) Correlation with Scatter Plot (See Appendix C)
(iv) Linear Regression with Scatter Plot (See Appendix D)
(v) Ecological Analyses - Shannon-Wiener Diversity Index (See Appendix E)

B. General Features of ERID Spreadsheets. Each datasheet already includes real data. As you explore these data, you will get proficient at navigating and utilizing the features in each spreadsheet.

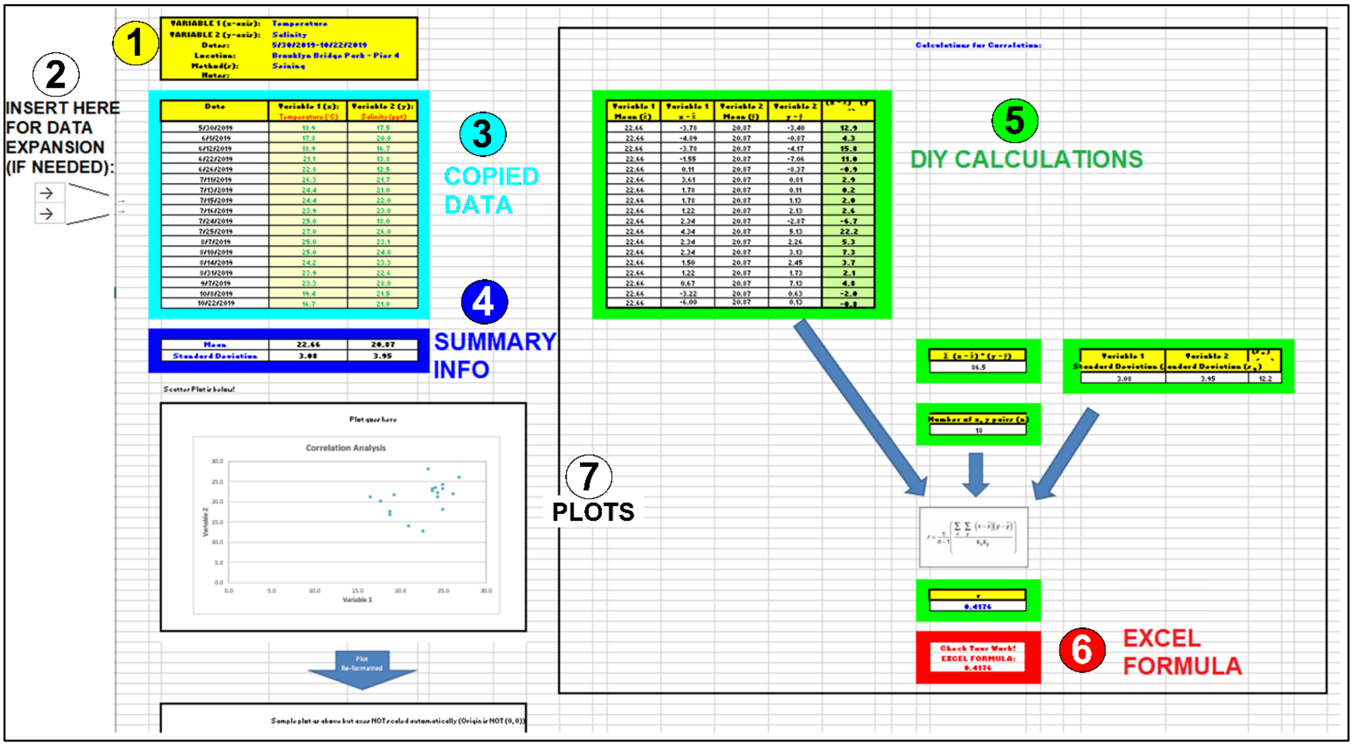

Figure 2. General Features of ERID Spreadsheets. See text for description of each numbered subsection.

General features of ERID spreadsheets are indicated in Fig. 2 and detailed below:
(1) General Information Table. This table is filled in with bright yellow. It includes general descriptors about the data.
(2) The double-arrows indicate the location where additional rows can be inserted to allow for the “Copied Data” table (see **(3)** below) to be expanded to accommodate larger datasets. Inserting rows here generally does not alter other aspects in the spreadsheet (e.g., DIY calculations, the format of tables, plots), but careful checking is recommended. For example, particular info (e.g., ranks) or calculations (e.g., formula answers) associated with the newly added rows of raw data entries may still need to be updated/revised manually. A good way to identify errors upon inserting new rows is to check the “DIY Calculations” table derived calculations (see **(5)** below) with the “Check Your Work” table (see **(6)** below).
(3) The “Copied Data” Table, surrounded by a light blue border, is the table where 2019 data from ERID can be pasted. In the spreadsheets, the to-be-copied data resides in the “REAL DATA” worksheet as green text within light yellow-filled cells. You will copy these data into the empty and light yellow filled cells of the “- STUDENT” worksheet; these to-be-pasted-into filled cells of the “-STUDENT” worksheet are also indicated with large bright green downward-pointing arrows near their column headers.
(4) If present, the “Summary Info” table(s), surrounded by a dark blue border, summarizes data in the “Copied Data” Table, usually in the form of means and standard deviations.
(5) The “DIY Calculations” table, surrounded by a bright green border, is often enclosed within its own broader box and displays all the step-by-step calculations involved in deriving any final key statistical values.
(6) The “Check Your Work” table, surrounded by a bright red border, provides a check of key result(s) of the “DIY Calculations” table. This red-bordered table can be a formula such as the correlation coefficient in the correlation

spreadsheet - in this case, the one-cell Excel formula shortcut is used to compare the value calculated in the DIY calculations. In other cases, such as in the Shannon-Weiner Index spreadsheet, it is simply the sum of proportions used to track calculations before calculating the index.
(7) If present, the "Plots" area includes graphs of data (e.g., scatterplot, histogram). Plots are first shown using Excel's default axes settings (which always displays the origin (0,0)), and then, if applicable, a large white downward-pointing arrow points to a second version of the plot, which has better formatted fitted axes.

(NOTE: Not all provided spreadsheets will have the information described above because some features will not apply to specific analyses.)

C. Let's Explore!

Your instructor will provide you with a scenario and East River fish data from which you can begin exploring patterns, visualizations, and interpretations most appropriate for the research question(s) to be answered!

V. References

East River Ichthyology Database (ERID). 2021. Available online at https://eastriverfishproject.org/. Accessed 26 Sept 2021.

Hurley A. 1994. Creating ecological wastelands: Oil pollution in New York City, 1870–1900. Journal of Urban History 20: 340–364.

Levinton JS, Waldman JR. 2006. The Hudson River Ecosystem. New York (NY): Cambridge University Press. 471 p.

Li Y, Meseck SL, Dixon MS, Wikfors GH. 2018. The East River tidal strait, New York City, New York, a high-nutrient, low-chlorophyll coastal system. International Aquatic Research 10: 65–77.

O'Neil JM, Taillie D, Walsh B, Dennison WC, Bone EK, Reid DJ, Newton R, Strayer DL, Boicourt K, Birney LB, Janis S, Malinowski P, Fisher M. 2016. New York Harbor: Resilience in the face of four centuries of development. Regional Studies in Marine Science 8: 274–286.

Park PJ, Girgenti CD, Del Bello IG, Tobitsch CM, Gorsen DM, Stanner KC, Van Horn D, Wilding KC, Gonzalez LF, Wu JR, Adams JJ, Reeve EJ, Bender ME, Bowser C, Turrin MK, Lake T. 2020. New York City East River fish species inventory and emergence of a unique fish community science network. Urban Naturalist 38: 1-27.

Steinberg T. 2014. Gotham Unbound: The ecological history of Greater New York. New York (NY): Simon and Schuster. 544 p.

Stinnette I, Taylor M, Kerr L, Pirani R, Lipuma S, Lodge J. 2018. NY/NJ HEP The State of the Estuary 2018. New York, (NY): Hudson River Foundation. Available online at https://www.hudsonriver.org/NYNJHEPStateoftheEstuary.pdf. Accessed 12 Nov 2020.

Waldman J. 2013. Heartbeats in the muck. New York (NY): Fordham University Press. 180 p.

Materials

Required equipment is only a computer (PC or Mac) or electronic device that has Microsoft® Excel.

Notes for the Instructor

Introduction

This section includes background information about New York City's East River, the East River Ichthyology Database (ERID) and its origin, general information about data analysis, detailed descriptions of the Microsoft® Excel ERID spreadsheets, and extensions of the present work. The latter includes an original activity, "Investigating Striped Bass Biology in New York City's East River," which incorporates the use of ERID spreadsheets and an online web tool, *Data Explorer* (HHMI BioInteractive, 2021), to explore 2019 data from ERID.

Background Information

Overview of East River Ichthyological Alliance (ERIA) and East River Ichthyology Database (ERID)

The East River Ichthyological Alliance (ERIA) was established in 2019 by a group of dedicated estuarine educators and researchers who realized the potential to be gained from enhanced research and knowledge, specifically of the East River strait's fishery. ERIA represents a first for New York City's iconic East River strait. Stakeholders from throughout the strait, including environmental educators, researchers, naturalists, and recreational and commercial anglers, gathered around a common goal to better understand the underwater life in the East River strait to serve as a proxy for ecological health and support increased engagement with its resources. The data provided in this work is a subset of the summary of fish data collected by ERIA in 2019, which was published in Park *et al*. (2020). One significant outcome of the 2019 ERIA research was the development of the East River Ichthyology Database (ERID, https://eastriverfishproject.org/). This database hosts real, current, and regularly updated fish and water quality data from New York City's East River strait (ERID, 2021). ERID was funded by a grant through a partnership among New York Sea Grant, the New York State Department of Environmental Conservation (NYSDEC), and the Marine and Coastal District of New York Conservation, Education, and Research Grants Program.

Caveats of ERID Data

Various collection methods were used to collect ERID data, and standardizing variables that serve as proxies for effort were not always clearly defined for many methods. Therefore, standardizing variables for catch data can be limited. Collection methods consisted of netting (e.g., seine net, cast net, or dip net), angling (e.g., fishing clinics, recreational angling), trapping, or personal observation. All seining and trapping data were acquired by an environmental education organization or academic institution. Seining employs a long, rectangular net dragged along a shoreline by at least two people (Říha *et al*., 2008). Depending on the organization, beach seine nets varied in length from 15 to 30 ft (457 to 914 cm), but all were ¼ inch (6.35 mm) mesh. Trapping involved the use of oyster mesh cages, Gee minnow traps, or crab traps. Angling involved hook-and-line bait fishing or lure fishing. Data submitted by recreational anglers required supplementation with photographs. Personal observations were made in the field only by ichthyologists. Before performing any parametric statistical tests, instructors should check to see if the data used conform to the assumptions of those tests. For example, many parametric tests assume that data values are normally distributed and homoscedastic, and data values often need to be transformed to align to test assumptions.

Types of Data Variables

Two types of data commonly explored and visualized with spreadsheets are numerical data and categorical data: (i) Numerical data, also called quantitative data, represent numerical values, which can either be discrete or continuous. Discrete values can only take on exact values that can be counted but not measured (e.g., number of students in a class). Typical visualizations for this type of data are scatter plots or bar graphs. Continuous values have an infinite number of probable values which can be selected, cannot be counted, but can be measured (e.g., height, weight, length). Typical visualizations for this type of data are scatter plots. (ii) Categorical data, also called qualitative data, are organized into categories representing characteristics, which can be nominal or ordinal. Categorical data can take on numerical values, but these values are often only organizational and do not carry mathematical meaning. For example, if data were coded as 1=red and 2=blue, the numbers cannot be logically added. Nominal values are discrete and do NOT follow a natural order (e.g., types of movies). Typical visualizations for this type of data are pie charts or bar graphs. Ordinal values are discrete and follow a natural order (e.g., letter grades, stages of a disease). Typical visualizations for this type of data are pie charts or bar graphs.

Classroom Implementation

Overview Microsoft® Excel East River Ichthyology Database (ERID) Spreadsheets

East River Ichthyology Database (ERID) data in the spreadsheets include fish and water quality data, collection methods, and general locations from a subset of data collected in 2019. Park *et al*. (2020) provide a convenient reference for instructors to tailor in-class applications of the database and contextualize student learning. Datasets from ERID can be downloaded at any time, thus permitting lessons to incorporate the most up-to-date data available. Using the provided Microsoft® Excel spreadsheets (hereafter referred to as “ERID spreadsheets”) and external tools, students will be able to calculate descriptive statistics, generate a frequency histogram, calculate and visualize correlation and linear regression, and calculate and interpret Shannon-Wiener Index. To develop a deep appreciation for the importance of surveying efforts by community scientists, a sample activity focused on the popular NYC sportfish striped bass (*Morone saxatilis*), is provided, which demonstrates the use of most of the ERID spreadsheets and HHMI BioInteractive *Data Explorer*.

Types of ERID Spreadsheets

Use of the ERID spreadsheets provided in this work assumes basic knowledge of Microsoft Excel® spreadsheets, but if instructors or students need tutorials on basic elements of Excel, Howard Hughes Medical Institute (HHMI) BioInteractive published comprehensive tutorials on formulae, function, and averages (HHMI BioInteractive, 2016a), autofilling data, cell references, and standard deviation (HHMI BioInteractive, 2016b), column graphs, error bars, and standard error of the mean (HHMI BioInteractive, 2016c), t-test (HHMI BioInteractive, 2016d), and histograms (HHMI BioInteractive, 2016e).

ERID spreadsheets provided in this work are described below. In general, short instructions, descriptors, and/or comments are provided in the files on many of the table header cells, identifiable with a red triangle on the upper-right corner of relevant cells and accessible via hovering over the red triangle with a mouse.

(1) ERID Descriptive Stats.xlsx (Appendix A) - Students can apply this spreadsheet to explore descriptive statistics. One example is provided of the size (total length – tip of snout to tip of caudal fin) of striped bass (*Morone saxatilis).*

(2) ERID Frequency Histogram.xlsx (Appendix B) - Students can apply this spreadsheet to partition data into bins and generate a frequency histogram. One example is provided of the size (total length) of striped bass (*Morone saxatilis*).

(3) ERID Correlation.xlsx (Appendix C) - Students can apply this spreadsheet to calculate a correlation coefficient and generate a scatterplot of two variables. Two examples from Brooklyn Bridge Park Pier 4 are provided - the relationship of marine water temperature and salinity (Example 1) and the relationship between marine water temperature and dissolved oxygen (Example 2).

(4) ERID Regression.xlsx (Appendix D) - Students can apply this spreadsheet to calculate a linear regression model (least squares estimation), which includes calculation of the slope, Y-intercept, and predicted values of the linear regression model. Students will also generate a scatterplot of the dependent variable and independent variable. One example is provided - the relationship of the size (total length) of black sea bass (*Centropristis striata*) and time (number of days).

(5) ERID Shannon-Wiener.xlsx (Appendix E) - Students can apply this spreadsheet to calculate Shannon-Weiner Diversity Index values. Three examples are provided to permit a comparison of biodiversity estimates of one site (Zone 9) across three different fish collection methods (seining, fishing clinic, recreational angling).

Types of Worksheets within ERID Spreadsheets

Each datasheet already includes real 2019 data from ERID, described as “examples” in the previous section. Students can learn to explore these example data to become proficient at navigating and utilizing the features in each spreadsheet. Each spreadsheet is generally organized into a set of worksheets, accessed by clicking on tabs at the bottom of the file. The “REAL DATA” worksheet contains 2019 data from ERID, located in the table with light yellow-filled cells and green text. The student version of the worksheet has a prefix that is an abbreviation of the spreadsheet (e.g., “FreqHist” for Frequency Histogram) and the base word “- STUDENT.” Students can simply copy data from the “REAL DATA” worksheet into the corresponding data entry columns in their “- STUDENT” worksheet (e.g., FreqHist – STUDENT); columns, where these data will be pasted, are conveniently coded similarly as the copied data in the “REAL DATA” worksheet – the empty cells where the data will be pasted are filled with a light yellow background. There are also large bright green arrows in the "- STUDENT" worksheet that indicate the columns where data is to be pasted. The instructor version of the statistic spreadsheet is indicated with the same prefix but with the base word "- INSTRUCTOR" (e.g., FreqHist – INSTRUCTOR); this file is essentially the answer key, and thus, it is recommended that this worksheet(s) be deleted before sharing the Excel ERID spreadsheet files with students. If a RESOURCES worksheet is also available, any relevant resources such as the HHMI BioInteractive Excel tutorials (HHMI BioInteractive, 2016a-e) are listed in that worksheet.

Time Requirements

Each Microsoft® Excel ERID spreadsheet file may include up to three student worksheets. Each spreadsheet file can be utilized as a stand-alone activity or in combination with others. Each spreadsheet file can be expected to take a minimum of 15 minutes as a class activity. The spreadsheets are designed to allow the instructor to determine how they are to be used when teaching, which will affect the time estimates for teaching particular lessons.

Extensions of the East River Ichthyology Database

Use of ERID Spreadsheets for Any Dataset

ERID spreadsheets do not need to be limited to fish data. Instructors can use the provided datasets simply as exemplars to teach a particular data representation method or analysis. Students can then be tasked to independently find and explore different types of datasets acquired from textbooks or other databases. Equipped with the blank student worksheets of ERID spreadsheets, new investigations can become student-driven independent research projects.

Create Lessons Using ERID Data

An original lesson derived from utilizing 2019 data in ERID, titled "Investigating Striped Bass Biology in New York City's East River," was created by Devin M. Gorsen and Peter J. Park and is provided in Appendix F. This activity provides a brief overview of the biology of the sportfish striped bass (*Morone saxatilis*) which includes the species life history and foraging ecology, monitoring efforts of the fishery in the Hudson River Estuary by the New York State Department of Environmental Conservation, surveying efforts by community scientists, and recreational angling of the species. This is followed by a section titled "Research Questions," which includes a set of five exercises, each inspired by observations recorded by environmental educators and researchers during past community science events. These observations were turned into research questions that were explored using 2019 data from ERID via provided Microsoft® Excel spreadsheets and/or the online web tool *Data Explorer* (HHMI BioInteractive, 2021). The data to be used in these exercises are provided in Appendix G. Students must incorporate data from Appendix G into the respective ERID spreadsheets and modify them as needed. Exercises 1-5 of this activity are summarized below.

<u>Exercise 1: 2019 Distribution of Striped Bass by Age Class</u>

Generally, it is believed that Atlantic striped bass spawn in the main part of the Hudson River near Albany, which is far north of New York City's East River. If so, one might expect young-of-year (one-year-old) bass to be rare in the East River because they are born in the Hudson River (and presumably not in the East River). Furthermore, striped bass generally are thought not to begin migrating outside the Hudson River until they reach sexual maturity, which in males occurs at 2-3 years and in females at 5-6 years. For this exercise, the Frequency Histogram ERID spreadsheet (Appendix B) is used to explore the distribution of striped bass size (total length) in the East River. The resulting histogram is provided in Fig. 3, and a key is provided in Appendix H.

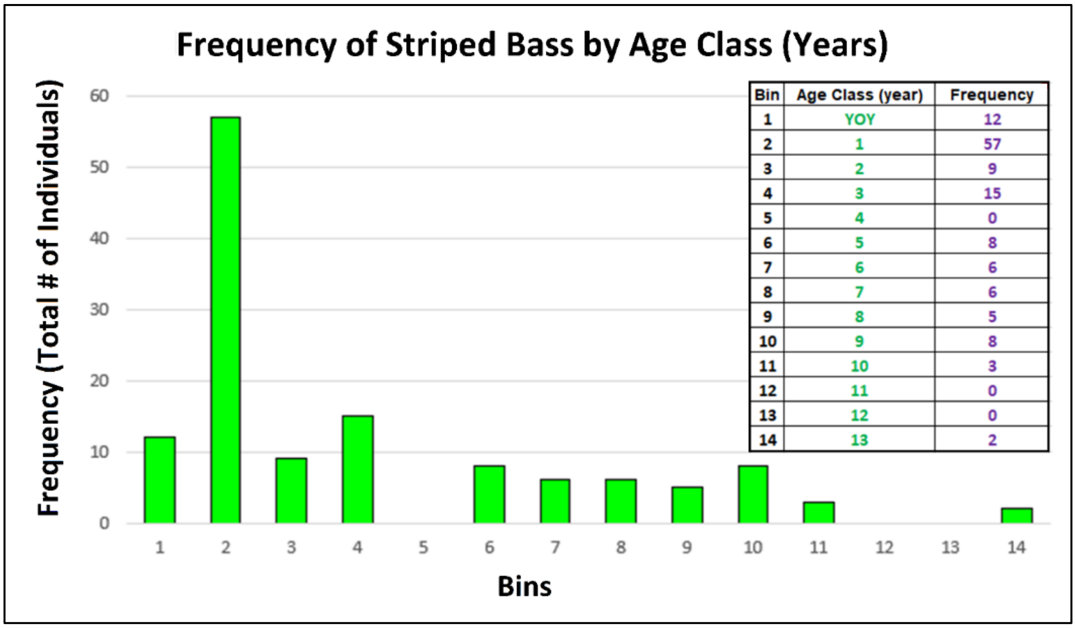

Figure 3. Histogram generated from Microsoft® Excel spreadsheet for Exercise 1. Inset, table of bins and their corresponding total length ranges in inches.

Exercise 2: Fate of the Striped Bass Fishery

In 2020, a fishing slot limit was imposed for striped bass for the first time in New York. Set by NYSDEC, this type of limit prohibits fish above or below certain lengths from being harvested. Thus, only bass within a specified range can be harvested. In 2020, a bass measuring between 28 inches and 35 inches can be harvested in the East River; bass measuring below 28 inches or above 35 inches must be released. The slot limit aims to protect large specimens of striped bass, which are mostly sexually mature females. Students must determine which spreadsheet to use to assess current distributions of striped bass sizes best and then predict the future of striped bass sizes in subsequent seasons. Answers for this exercise will vary.

Exercise 3: Average Size of Striped Bass and Time of the Year Caught

Some East River anglers believe that the biggest striped bass in the East River are caught only in the Fall. Students first use the Linear Regression ERID spreadsheet (Appendix D) to identify the relationship of average size (total length) of striped bass caught by month. Then, students use the HHMI BioInteractive *Data Explorer* web tool to reproduce the plot and compare it with the one in their spreadsheet. The input data file (Appendix G) is also formatted so that *Data Explorer* can color-code the plot data by collection method, which is substantially more challenging to accomplish using Microsoft® Excel. In general, the linear regression model of all observations in the dataset was not statistically significant, but data submitted by recreational anglers do suggest that larger bass were caught not only during the Fall but also during Spring months (Fig. 4). Exercise 3 was also designed to prepare students to transition from the Microsoft® Excel spreadsheet to HHMI BioInteractive *Data Explorer*.

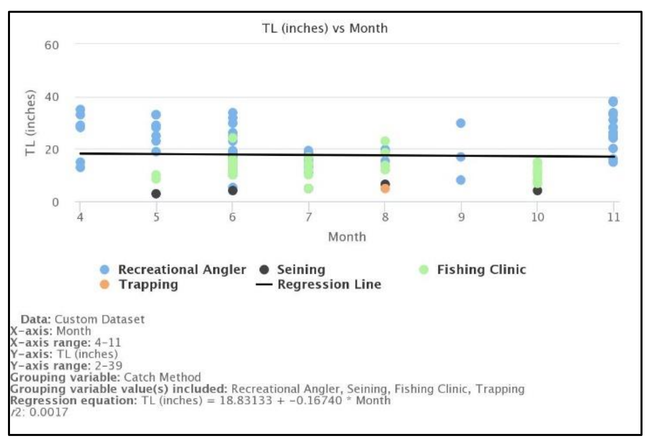

Figure 4. Output from HHMI BioInteractive *Data Explorer* for Exercise 3. Plot summarizes the total length of striped bass by month caught. The linear regression line is shown for all observations in the dataset. Collection methods are coded by different colors.

<u>Exercise 4: Collection Method and Average Size of Striped Bass Caught</u>

Some environmental educators believe that the average size of striped bass caught is influenced by the type of fish collection method employed. The types of collection methods in the 2019 dataset from ERID are hook-and-line fishing during a fishing clinic, hook-and-line fishing by recreational anglers during personal outings, seining during a public event, and trapping during a public event. Students use the HHMI BioInteractive *Data Explorer* web tool to determine the best visualization plot for the data presented in this exercise. The resulting plot is provided in Fig. 5.

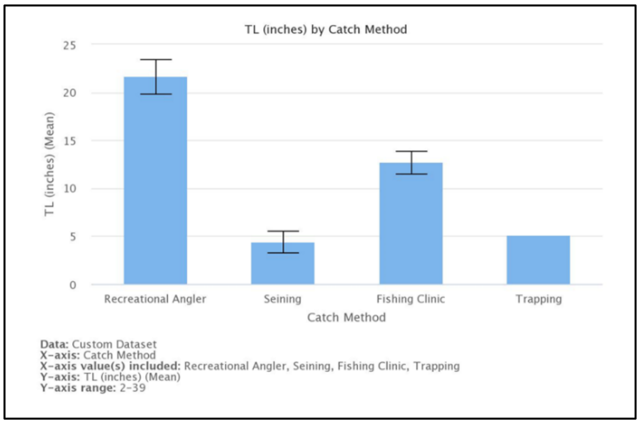

Figure 5. Output from HHMI BioInteractive *Data Explorer* for Exercise 4. The total length of striped bass by fish collection method is shown.

<u>Exercise 5: Geographic Area and Average Size of Striped Bass Caught</u>

Some recreational anglers believe that the average size of striped bass caught could differ across geographic regions. Using the HHMI Biointeractive *Data Explorer* web tool, students explore the 2019 data from ERID to determine the best plot to visualize the relationship between the geographic region of the East River and the distribution of average sizes (total lengths) of striped bass. For the convenience of this analysis, the East River can be divided into eleven different zones, labeled Z1-Z10, based on natural and/or anthropogenic landmarks (see Fig. 1). The resulting plot is provided in Fig. 6.

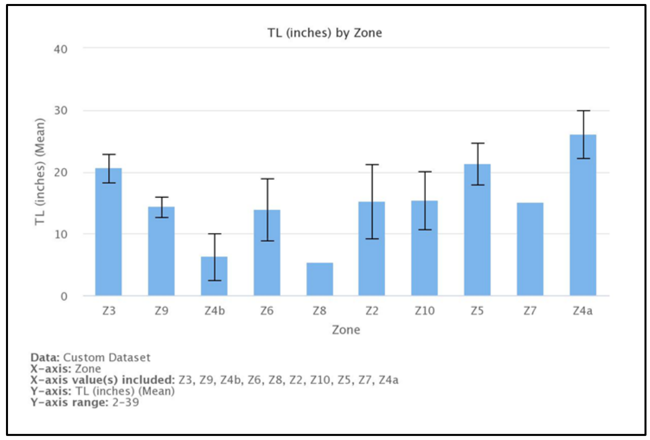

Figure 6. Output from HHMI BioInteractive *Data Explorer* for Exercise 5. The total length of striped bass by East River zone (See Fig. 1 for zones) is shown.

Instructor notes and suggestions for exercises are provided in the activity as red text (See Appendix F), which should be deleted before sharing or discussing with students. In addition to the "Research Questions" section, a couple of additional extension exercises are provided. The first focused on differences in recreational fishing slot limits sizes in different parts of the Hudson River Estuary. The second focused on human consumption advisories of striped bass.

Discussion

While many biology instructors aim to instill strong analytical skills anchored in statistics, biology students do not usually get to explore datasets relevant to real-life applications. We introduced the East River Ichthyology Database (ERID) and a series of customized Microsoft® Excel spreadsheets to address these concerns. While the ERID spreadsheets can be used for any dataset, we demonstrate their utility when combined with an emerging statistical web tool and in the context of a community science story centered on a popular New York City sportfish, the striped bass (*Morone saxatilis*). The exercises in this activity also demonstrate top-down and bottom-up approaches to teaching statistics. In addition, all research questions in this activity were driven by authentic inquiries from the general public or environmental educators during actual community science events. Thus, the integration of community science, ERID, and statistical tools provides a powerfully novel teaching model to engage students as they learn complex topics at the interface of biology and statistics.

Two approaches to introduce statistics were presented in this work. Using the Microsoft® Excel ERID spreadsheets, students learn how to calculate descriptive statistics and statistical analyses via a bottom-up approach, with raw data converted to calculations that can be easily tracked through final test statistical calculations and data visualizations (e.g., plots, graphs). In addition, we introduced the "Investigating Striped Bass Biology in New York City's East River" activity in the Extensions section. In exercises 3-5, students import data into HHMI BioInteractive *Data Explorer* and bypass detailed step-by-step calculations to choose the best data visualizations for a given dataset, which is more of a top-down approach to teaching statistics. Even though we introduced the spreadsheets and *Data Explorer* in this manner, we acknowledge that each tool can be used in either top-down and bottom-up ways. Regardless, the combination and complementary nature of both approaches, presented simultaneously, can provide compelling learning experiences for students.

ERID introduces students to the efforts of real community scientists. In tandem with its scientific benefits, community science also has the power to create diverse constituencies who are excited about their local ecosystems and, ultimately, more scientifically literate and engaged in conservation efforts. Fishes are a particularly vital subject for community-based research because they live adjacent to even the most urbanized and populous cities, they capture the imaginations of people all over the world, and they provide a strong link to other environmental concerns (Brink *et al*., 2018). The exercises in the "Investigating Striped Bass Biology in New York City's East River" activity were inspired by real inquiries from the general public that could not be explored before establishing a database like ERID. Thus, integrating the personal and social benefits of science-based informal education programs with statistics learning in formal education settings presents students with real-world context to learning that may be less intimidating than solving textbook problems based on abstract datasets.

Community science can have several goals, including introducing and promoting environmental education, exposing communities to scientific research, creating ownership of the scientific process, surveying biodiversity together, and providing data for teaching applications. Moving forward, ERIA and its partnerships with public participants and anglers and the further growth of ERID create a model to build lasting connections between stakeholders and the health of local ecosystems. Armed with real ecological data from ERID, instructors can construct innovative activities relevant to students' lives to inspire them to participate in environmental initiatives. As extensions of this work, assignments can be written to assess higher-level Bloom's learning. Examples of such assignments can be writing a mock grant proposal or evaluating existing environmental engineering green infrastructure project plans. The present work also aims to inspire student participation in local community science efforts. Tying together community science, a current fish database, and in-classroom teaching may promote rich, profound outcomes that benefit students, instructors, community stakeholders, policymakers, and land managers of watersheds and beyond.

ERID Data Use and Permissions

Please contact the corresponding author regarding inquiries about the most current East River Ichthyology Database data. Instructors are permitted to use the data within the East River Fish Database for in-classroom educational purposes only. Any other use of data from the database external to in-classroom educational use (e.g., publications, conference presentations) must obtain expressed written consent directly from the East River educational organizations involved in acquiring and submitting the relevant data. Upon receiving consent, specific instructions, which may vary by organization, will be provided regarding how to acknowledge the relevant organization(s) in the external work.

Acknowledgements

We would like to thank the entire East River Ichthyological Alliance network, which currently includes Farmingdale State College (Farmingdale, NY), Randall's Island Park Alliance (New York, NY), New York Office of Parks, Recreation, and Historic Preservation (New York, NY), Brooklyn Bridge Park Conservancy (Brooklyn, NY), New York State Department of Environmental Conservation - Region 2 Office (Long Island City, NY), Alley Pond Environmental Center (Oakland Gardens, NY), Battery Park City Authority (New York, NY), City Parks Foundation - Coastal Classroom (New York, NY), Lower East Side Ecology Center (New York, NY), and Nyack College (New York, NY). We thank Hannah Ahn for permission to use the image "East River Ichthyological Alliance East River Zones of

Study." We thank Sarah Gross and the Department of Biology at Farmingdale State College for support and encouragement of this work. We are grateful to the New York State Department of Environmental Conservation (NYSDEC), especially Carol Hoffman and Caitlin A. Craig, for providing a list of striped bass fishing regulations for New York State from 1976-present and Melissa K. Cohen and her team at NYSDEC Region 2 Office (Long Island City, NY) for their I FISH NY data contributions and constant guidance and feedback on various components of this work. The I FISH NY program in the NYSDEC Bureau of Fisheries is an outreach program promoting angling in urban areas. Region 2 staff conducts free fishing events in all boroughs of New York City, in freshwater and saltwater. We are also grateful to the New York State Department of Health, especially Audrey Van Genechten for providing data on striped bass consumption advisories. We thank Laura Bonetta, Melissa Csikari, and the entire HHMI BioInteractive team for their helpful guidance on implementing *Data Explorer* and Rebecca Orr for directing us to the HHMI Microsoft® Excel tutorials from 2016. We thank Erwin Cabrera and the Farmingdale State College Research Aligned Mentorship (RAM) Program and Risa Stein and the Farmingdale State College Collegiate Science and Technology Entry Program (CSTEP) for the opportunity to utilize the ERID spreadsheets. We also thank Monica Cordova Hernandez and Matthew DeMilt for feedback on the ERID spreadsheets. Lastly, we are indebted to New York Sea Grant, especially Antoinette O. Clemetson, for encouragement and feedback on previous versions of exercises within the "Investigating Striped Bass Biology in New York City's East River" activity. This project was supported by a grant through a partnership among New York Sea Grant, the New York State Department of Environmental Conservation, and the Marine and Coastal District of New York Conservation, Education, and Research Grants Program. Funding for ERID was provided by the Marine and Coastal District License Plate, which is administered by the Marine and Coastal District of New York Conservation, Education and Research Board, and authorized through NYS Environmental Conservation Law Article 13, Title 5 Section 13-0503; this grant is a collaboration among Farmingdale State College, Nyack College, Brooklyn Bridge Park Conservancy, and Randall's Island Park Alliance.

Cited References

Brink K, Gough P, Royte J, Schollema PP, Wanningen H (Eds.). 2018. From Sea to Source 2.0: Protection and restoration of fish migration in rivers worldwide. Groningen (The Netherlands): World Fish Migration Foundation. 360 p.

East River Ichthyology Database (ERID). 2021. Available online at https://eastriverfishproject.org/. Accessed 26 Sept 2021.

HHMI BioInteractive. 2016a. Microsoft® Spreadsheet Excel Tutorial 1 - Formulae, Function, and Averages. Available online at https://www.biointeractive.org/classroom-resources/spreadsheet-tutorial-1-formulae-functions-and-averages. Accessed 26 Sept 2021.

HHMI BioInteractive. 2016b. Microsoft® Spreadsheet Tutorial 2 - autofilling data, cell references, and standard deviation Available online at https://www.biointeractive.org/classroom-resources/spreadsheet-tutorial-2-autofill-data-cell-references-and-standard-deviation. Accessed 26 Sept 2021.

HHMI BioInteractive. 2016c. Microsoft® Spreadsheet Tutorial 3 - column graphs, error bars, and standard error of mean Available online at https://www.biointeractive.org/classroom-resources/spreadsheet-tutorial-3-column-graphs-error-bars-and-standard-error-mean. Accessed 26 Sept 2021.

HHMI BioInteractive. 2016d. Microsoft® Spreadsheet Tutorial 4 - t-test Available online at https://www.biointeractive.org/classroom-resources/spreadsheet-tutorial-4-ttest. Accessed 26 Sept 2021.

HHMI BioInteractive. 2016e. Microsoft® Spreadsheet Tutorial 5 - histograms Available online at https://www.biointeractive.org/classroom-resources/spreadsheet-tutorial-5-histogram. Accessed 26 Sept 2021.

HHMI BioInteractive. 2021. HHMI BioInteractive *Data Explorer.* Available online at https://media.hhmi.org/biointeractive/tool/dataexplorer/v1/#/home. Accessed 26 Sept 2021.

Hurley A. 1994. Creating ecological wastelands: Oil pollution in New York City, 1870–1900. Journal of Urban History 20: 340–364.

Levinton JS, Waldman JR. 2006. The Hudson River Ecosystem. New York (NY): Cambridge University Press. 471 p.

Li Y, Meseck SL, Dixon MS, Wikfors GH. 2018. The East River tidal strait, New York City, New York, a high-nutrient, low-chlorophyll coastal system. International Aquatic Research 10: 65–77.
Microsoft Corporation. 2018. Microsoft Excel. Retrieved from https://office.microsoft.com/excel.
O'Neil JM, Taillie D, Walsh B, Dennison WC, Bone EK, Reid DJ, Newton R, Strayer DL, Boicourt K, Birney LB, Janis S, Malinowski P, Fisher M. 2016. New York Harbor: Resilience in the face of four centuries of development. Regional Studies in Marine Science 8: 274–286.
Park PJ, Girgenti CD, Del Bello IG, Tobitsch CM, Gorsen DM, Stanner KC, Van Horn D, Wilding KC, Gonzalez LF, Wu JR, Adams JJ, Reeve EJ, Bender ME, Bowser C, Turrin MK, Lake T. 2020. New York City East River fish species inventory and emergence of a unique fish community science network. Urban Naturalist 38: 1-27.
Říha M, Kubečka J, Mrkvička T, Prchalová M. 2008. Dependence of beach seine net efficiency on net length and diel period. Aquatic Living Resources 21: 411–418.
Steinberg T. 2014. Gotham Unbound: The ecological history of Greater New York. New York (NY): Simon and Schuster. 544 p.
Stinnette I, Taylor M, Kerr L, Pirani R, Lipuma S, Lodge J. 2018. NY/NJ HEP The State of the Estuary 2018. New York, (NY): Hudson River Foundation. Available online at https://www.hudsonriver.org/NYNJHEPStateoftheEstuary.pdf. Accessed 12 Nov 2020.
Waldman J. 2013. Heartbeats in the muck. New York (NY): Fordham University Press. 180 p.

About the Author(s)

Peter J. Park is an Assistant Professor, with expertise in evolutionary ichthyology, in the Department of Biology at SUNY Farmingdale State College in Long Island, New York. He teaches majors Introductory Biology II (Organismal diversity and evolution) and coordinates its lecture and laboratory course. He is an HHMI BioInteractive Ambassador Academy member, enthusiastic contributor to NYC fish community science programs, and co-developer of instructional tools such as Laboratory Aquaponics and the HHMI Stickleback Evolution Virtual Lab.

Devin M. Gorsen currently manages Marsha P. Johnson State Park in Williamsburg, Brooklyn, where she oversees all park maintenance, operations, programming, and stewardship. Prior to that, she oversaw the NYC Regional Environmental Education team, which develops and leads environmental education programs at five of the eight State Parks in New York City, including Marsha P. Johnson State Park, Gantry Plaza State Park, Roberto Clemente State Park, Denny Farrell Riverbank State Park, and Bayswater Point State Park. She received her BA in Environmental Studies from Davidson College and her MA in Biology through Miami University and the Wildlife Conservation Society's Advanced Inquiry Program.

Christopher D. Girgenti is the Director of Environmental Programs with expertise and over a decade of experience in urban restoration ecology at the Randall's Island Park Alliance in New York City. He oversees the adaptive management, research, restoration, and educational programming for the 330-acre park. He has a Master's in environmental management from Duke University, received the Environmental Champion Award from EPA Region 2 for the Waterfront Stewardship Program in 2017, and passionately supports engagement with communities and their local environment throughout New York City.

Isa G. Del Bello works to introduce numerous students to urban ecology and wild spaces found within and around the Brooklyn waterfront as Director of Education with Brooklyn Bridge Park Conservancy. She holds a M.A. in cognitive and developmental psychology with a concentration in risk, resilience and prevention and a B.A. in elementary education with dual concentrations in science and literacy.

Christina M. Tobitsch is the Education Manager at Brooklyn Bridge Park Conservancy. She often engages students and families in estuary education along the East River. She holds a B.S. in Environmental Science from the University of Connecticut and a Master's in Environmental Conservation Education from NYU. Her thesis investigated the value to having a centralized community science platform for local marine life; she has been thrilled to collaborate with a team of East River partners who share that vision.

Jacqueline R. Wu is the Park-as-Lab Research Coordinator at Randall's Island Park Alliance in New York City where she develops and conducts environmental monitoring projects around the Island.

She received her B.S. in Natural Resources with a concentration in applied ecology from Cornell University and her M.S. in Marine Biology from Northeastern University.

Andrew J. Hong is an independent researcher and the Software Developer of the East River Ichthyology Database (ERID). His professional background focuses on creating workflows for content and data management with an expertise in using automation to streamline processes. He holds a B.A. in Politics and Sociology from New York University.

Appendices

Filenames of Appendices:
- Appendix A - ERID Descriptive Stats.xlsx
- Appendix B - ERID Frequency Histogram.xlsx
- Appendix C - ERID Correlation.xlsx
- Appendix D - ERID Regression.xlsx
- Appendix E - ERID Shannon-Wiener.xlsx
- Appendix F - Investigating Striped Bass.docx
- Appendix G - IMPORT DATA - ERIA_HHMI_DE - StripedBass2019.csv
-Appendix H – Investigating Striped Bass Ex1 Histogram Key.xlsx

Mission, Review Process & Disclaimer

Citing This Article

Peter J. Park, Devin M. Gorsen, Christopher D. Girgenti, Isa G. Del Bello, Christina M. Tobitsch, Jacqueline R. Wu, and Andrew J. Hong. 2022. From Statistics to Ecological Analyses: An Application of the New York City East River Ichthyology Database (ERID) Using Spreadsheets. Article 15 In: Boone E and Thuecks S, eds. *Advances in biology laboratory education*. Volume 42. Publication of the 42nd Conference of the Association for Biology Laboratory Education (ABLE). https://doi.org/10.37590/able.v42.art15