OVERVIEW

Data Explorer is a web tool that provides a quick and easy way of visualizing and analyzing data without advanced technological requirements. You can open the tool in any major web browser, and on most devices, to use the different functions. With *Data Explorer*, students can easily create a variety of plots to visualize data. They can also quickly generate statistical summaries and perform hypothesis testing with a variety of tests.

Using *Data Explorer* allows students to build their skills in data literacy and science practices by exploring real research data. The tool includes a curated collection of authentic data sets supplemented with background information and in-depth BioInteractive activities. Students can also import their own data sets to use within the tool. Projects can be saved and/or downloaded at any time.

The "How to Use *Data Explorer*" document provides a more in-depth guide on using the tool. Additional information related to pedagogy and implementation can also be found on this resource's webpage, including suggested audience, estimated time, and curriculum connections. Additional resources and help, including frequently asked questions (FAQs), can be found in the footer menu for *Data Explorer*.

KEY CONCEPTS

- Plots (graphs) are useful tools for visualizing data.
- Data can be analyzed to test hypotheses and address research questions.

STUDENT LEARNING TARGETS

- Explore data sets through visualizations and statistical tests.
- Analyze and interpret data.

PRIOR KNOWLEDGE

Data Explorer can be used by students with any level of experience. However, students with more prior knowledge in data analysis will better understand the purposes, requirements, and results of the different plots and statistical tests. *Data Explorer* contains brief descriptions of the plots and tests, but is not designed to teach graphing or statistics from scratch.

Depending on your class's needs and objectives, you may want to provide additional support for certain tasks, such as the following:

- To **interpret a plot**, students should understand the basic composition of a plot, including axes and variables.
- To **select a plot type**, students should know the purposes and requirements for each plot type. The "Plot Selection Guide" on the *Data Explorer* webpage can be used as a supporting resource; similar information is provided in the "Help" page of *Data Explorer* (link located in the footer menu).
- To **interpret a statistical analysis**, students should have a basic understanding of the normal distribution, hypothesis testing, statistical significance, and *P* values.
- To **select a statistical analysis**, students should know the purposes, requirements, and assumptions of each statistical analysis. The "Statistical Test Selection Guide" on the *Data Explorer* webpage can be used as a

supporting resource; similar information is provided in the "Help" page of *Data Explorer* (link located in the footer menu).

BACKGROUND ON THE TOOL

Data Explorer contains three main tabs at the top: "Data," "Visualize," and "Analyze." These tabs are independent of each other and can be mixed and matched based on your class's needs and objectives.

The **"Data"** tab is for exploring the data table. It includes the following features, which can be accessed from the gray bar at the top:

- Data table
- Cross tabulation (also called pivot table)

The **"Visualize"** tab is for creating plots (graphs). It includes the following types of plots, which can be accessed through the "Add a new plot" button:

- Histogram
- Scatterplot
- Bar graph
- Line graph
- Box plot

The **"Analyze"** tab is for performing statistical tests. It includes the following tests, which can be accessed through the "Add a new analysis" button:

- Descriptive statistics
- Chi-square test of goodness of fit
- Chi-square test of independence
- Two-sample *t* test assuming equal variance (Student's *t* test)
- Two-sample *t* test *not* assuming equal variance (Welch's *t* test)
- Paired *t* test
- One-way ANOVA assuming equal variances (Fisher's ANOVA)
- One-way ANOVA *not* assuming equal variances (Welch's ANOVA)
- Mann-Whitney *U* test (independent samples)
- Wilcoxon signed rank test (paired samples)
- Linear regression
- Linear correlation

TEACHING TIPS

- For students with less experience in data analysis and quantitative reasoning, provide support through additional resources and/or consider differentiating tasks.
- Consider introducing different parts of *Data Explorer* at various points of instruction. Exposing students to certain skills at different times helps build student capacity.
- Consider having students work in pairs or small groups to support risk-taking and discourse.
- You can find more data sets to use by going to the "Help" page of *Data Explorer* (located in the footer menu), then scrolling down to the "Publicly available data" section under "Other Resources."

CURATED DATA SETS AND ACTIVITIES

Under the "Choose data to explore" option on the landing page of *Data Explorer*, you will find a curated collection of authentic data sets supplemented with background information and in-depth BioInteractive activities. This collection will be continually updated to add more data sets over time.

When you open a curated data set, there will be an extra tab at the top titled **"Background."**

- This tab introduces how the data was collected and what the data includes.
- It intentionally does not discuss the analyses done by the scientists who collected the data, in order to give students the freedom to do any analyses that interest them. To learn more about the scientists' analyses, consult the "References" at the end of the tab.

The **"Related Activity"** section in the "Background" tab contains a link to an accompanying activity developed by BioInteractive.

- Each activity has its own "Student Handout" document to guide students' exploration of the data. There may be multiple versions of the handout to engage students with different levels of graphing and/or statistical knowledge.
- Each activity also has its own "Educator Materials" document with additional background information and implementation suggestions.
- Use of this activity is *completely optional*. You can use the activity as is, customize it based on your class's needs, or design a new activity to use with the data.

As examples, two of the curated data sets and their related activities are described below.

Data Set 1: Elephant populations under poaching

This data set was collected from populations of African elephants that have suffered historically heavy poaching and recent population declines. It includes data on the elephants' physical characteristics, including tusk size, over two time periods (1960s and 2000s).

Related activity: "Using Data to Investigate Elephant Evolution"
This activity uses a guided inquiry approach. Students come up with their own ways to investigate the data at the beginning and are then guided through more specific graphical and statistical analyses. There are two versions of the "Student Handout."

In "Version 1," students:
- generate their own research question, hypothesis, and plot
- construct bar graphs to compare means (error bars not required)

In "Version 2," students:
- generate their own research question, hypothesis, plot, and statistical analysis
- construct bar graphs to compare means (including error bars)
- determine whether a difference in means is statistically significant by using 95% confidence intervals and Student's *t* test

Data Set 2: Finches in the Galápagos

This data set was collected from a population of medium ground finches that experienced a severe drought in 1977. It includes beak and body measurements from 100 birds from 1973 to 1980.

Related activity: "Evolution in Action: Data Analysis"
This activity uses a direct instruction method. Students follow a linear procedure to perform graphical and statistical analyses. There are two versions of the "Student Handout."

In "Version 1," students:
- interpret histograms
- calculate means and standard deviations
- construct bar graphs to compare means (error bars not required)

In "Version 2," students:
- calculate mean, standard deviation, standard error of the mean, and 95% confidence interval
- construct bar graphs to compare means (including error bars)
- perform a *t* test
- (optional) construct a scatterplot and use linear regression

REFERENCES

McDonald, John H. *Handbook of Biological Statistics*, 3rd ed. Baltimore, Maryland: Sparky House Publishing, 2014. Available online at http://www.biostathandbook.com/.

Whitlock, Michael C., and Dolph Schluter. *The Analysis of Biological Data*. Greenwood Village, CO: Roberts and Company Publishers, 2009.

Zar, J. H. *Biostatistical Analysis,* 5th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2010.

CREDITS

Written by Li Yao, Esther Shyu, HHMI
Edited by Kim Parfitt, Central High School, Cheyenne, WY (retired)