Title

DistPlotter: An Application that Creates High-Quality Plots of Univariate Distributions

Journal
Technology Innovations in Statistics Education, 17(1)

Authors
Casement, Christopher
Innerst, Grant
Innerst, Melissa

Publication Date
2026-06-20

DOI
10.5070/T5.1885

Peer reviewed

DistPlotter: An Application that Creates High-Quality Plots of Univariate Distributions

Christopher Casement[1], Grant Innerst[2], Melissa Innerst[3]
Fairfield University[1], Shippensburg University[2], Dickinson College[3]

Abstract

Coverage of distributions, such as normal and *t* (among others), is prevalent in courses focused on probability and/or statistics irrespective of the discipline. In this article, we discuss existing software that can be used to create plots of univariate distributions and highlight several limitations of these tools. We then introduce a free, web-based, point-and-click application that we developed to address these limitations. The application is designed for both instructor and student use and supports interactive exploration of distributions. Visualizations of distributions can support conceptual understanding, and hands-on interactions with such tools can help students develop deeper insight into distributional behavior. To facilitate classroom implementation, we provide sample activities that can be incorporated at the introductory, intermediate, and advanced levels.

Keywords: statistics education, educational technology, univariate distributions, Shiny

1. INTRODUCTION

In the classroom, visualizations can enhance a student's ability to understand complex topics in the statistics curriculum (Bobek and Tversky 2016; Chance et al. 2007; Pea 1987). One such topic is the concept of a probability distribution. It is common for instructors of probability and statistics courses to cover multiple univariate distributions throughout a course. For instance, at the introductory level, normal and *t* distributions are commonly covered, as are other distributions – e.g., binomial, uniform, chi-squared, and *F* – depending on the instructor. In more advanced, math-based probability and statistics courses, distributions such as beta, gamma, and Poisson, among others, are also covered. Due to the inherent importance of the visual aspects of a probability distribution, like the shape of a probability density or histogram, visualizations play an important role in understanding key attributes of a distribution. Understanding how differing parameter values affect the shape of a distribution is a fundamental concept that would be nearly impossible without visualizations. In addition, concepts like the skewness and modality of a distribution can be challenging to understand abstractly, but can easily be explained through visualizations.

In addition to helping students understand distributions themselves, visualizations can illustrate the foundational role that distributions play in other concepts encountered throughout the statistics curriculum. For example, visualizations that represent probabilities as shaded areas under a curve are helpful when calculating a probability based on cutoff(s) or finding a cutoff (e.g., a percentile) based on a probability. Thus, it is important that instructors can create visualizations to include in class materials such as notes and activities (Gentleman 2012).

In a broader sense, the inclusion of visualizations in the statistics curriculum and classroom is important because graphical competency has become an essential skill for any consumer of data. Due to the presence of data, and therefore graphical visualizations, in nearly every facet of our lives, reading information from visualizations is a key feature of statistical literacy (Friel et al. 2001; González et al. 2011). Despite the importance, students are known to have trouble reading information from visualizations and going beyond the initial level of graphical competence (Biehler et al. 2018; Friel et al. 2001). Research has also shown that teachers can be lacking in graphical competency as well (González et al. 2011). Fortunately, increased access to technology may result in improved understanding.

Technology is changing the landscape of statistics education due to its ability to allow students to conceptualize statistical concepts through visualization (Wild et al. 2011). As high-quality tools become readily available for students and teachers alike, recommendations on when and how to use them have followed. Green and Blankenship (2015) include the use of technology and visualizations in their list of tools that can be used to foster active learning and conceptual understanding of the concepts discussed in a traditional mathematical statistics classroom. Further, the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report, written by Carver et al. (2016), now recommends the use of technology to explore conceptual ideas and enhance student learning. Frameworks also have been created to evaluate the tools and how they fit into the statistics education landscape (Biehler 2019; McNamara 2018). The GAISE guidelines and others mentioned above suggest that when selecting a tool, instructors should keep various things in mind, such as the tool's use in supporting the intended educational objectives, its user-friendliness (including when inputting data), its capacity to import existing data stored in various formats, and its interactivity.

Interactivity is becoming standard on the web and should be encouraged in the classroom as well. Instructors should "teach about, and with, interactive graphics" (Ridgeway 2015). Similarly, the direct manipulation of a graphic is recommended over the adjustment of a script (Biehler 1997). This especially makes sense for exploring probability distributions. The resulting reduction in the amount of time between making a change and seeing the results can be helpful in making connections between the value of a parameter and the characteristics of a probability distribution. For example, being able to interactively vary the standard deviation of a normal distribution would hopefully lead to a more nuanced understanding of the parameter, which is notoriously challenging for students to interpret.

Most of the instructional technology commonly used in the statistics classroom has the capability to plot distributions. However, those technological tools do not all offer the same functionalities. Ideally, any tool would be free, would be easy-to-use, and would produce highly customizable plots. Unfortunately, many of the existing tools cannot satisfy all three of these needs. For instance, tools that produce highly customizable plots often come with a cost or a steep learning curve.

In this article, we discuss a free application we developed to address the limitations discussed previously. The application enables users to create customizable, high-quality plots of univariate distributions through a point-and-click interface. The article continues as follows. Section 2 discusses popular existing tools to create plots of distributions and drawbacks associated with them. Section 3 describes our application and the main features it offers. Section 4 identifies the intended audience for our application and contains examples of probabilistic and statistical concepts that can be demonstrated visually using our application. Lastly, Section 5 concludes with a summary of the article.

2. EXISTING TOOLS FOR PLOTTING UNIVARIATE DISTRIBUTIONS

The current landscape of tools with the capability of plotting distributions is vast. The tools range from specialized applications that focus only on distributions to general-purpose software where distributions represent only a small subset of the focus and capabilities. To make an objective comparison of these tools, it is important to formulate a framework for evaluation. Such frameworks exist in the statistical literature (Biehler 2019; McNamara 2018). McNamara's work lays out ten key attributes for assessing statistical computing tools (McNamara 2018, p. 5). Because our analysis is focused on plotting distributions, we adapt and condense their framework to five attributes that fit our situation:

- Accessibility
- Easy entry for novice users
- Plot interactivity
- Probability distribution plotting capability
- Data plotting and fitting capability

The first three attributes are taken directly from McNamara's work and Repenning et al.'s before them (McNamara 2018; Repenning et al. 2010). In terms of accessibility, we agree that tools should be free (or at least affordable), work on different operating systems, and be easy to install or gain access to. In terms of easy entry, we agree that the tools should be easy to understand and use, even for novices. For our focus, users should be able to explore and plot different distributions within minutes of being introduced to the tool. Our plot interactivity attribute is adapted from McNamara's attribute called "Interactivity at Every Level" to focus solely on plotting. We agree that tools should provide 'dynamic-interactive' graphics, where users can interact with any component of the plot and have the results update in real time (McNamara 2018). Users should be able to interact with every aspect of the plot as they are completing their exploration. This includes being able to modify plot parameters (parameters of a distribution, bins of a histogram, etc.), zoom in/out on a plot, access information like the height of density curves or histogram bars, and shade areas under density curves or probability. Users should also be able to interact with plot aesthetics like colors, axes, line types, and annotations.

The last two attributes – probability distribution plotting and data plotting/fitting ability – we introduce as attributes that are more specific to distribution plotting-focused tools such as ours. We describe the rationale of each attribute below.

Because plotting probability distributions is the focus of both this article and the application we present, we thought it important to include an attribute that directly focuses on the capability of tools to plot probability distributions, which we call probability distribution plotting capability. In general, we believe that any statistical computing tool should have the capability to plot probability distributions commonly taught in introductory statistics courses. Tools that are focused solely on probability distributions should go beyond that and contain other, more advanced distributions that one might encounter throughout a statistical education or career. Tools also ideally should be able to calculate probabilities using that distribution and shade in corresponding areas all in the same window.

A modern statistical computing tool needs to be able to plot data well. For this attribute, we are interested in the role of plotting data in relation to probability distributions. The main task in this area is fitting a probability distribution to a set of data. We define the data plotting and fitting capability attribute as a check to see if the tool has the capability to plot a set of data, estimate the parameters of a distribution with the data, and then overlay the fitted distribution.

In the discussion that follows, we partition the tools into different relevant groups, evaluate them based on the above criteria, and offer a comparison to DistPlotter's functionalities. We note that this list is not exhaustive. We do not include a discussion on programming languages like R, Python, Julia, MATLAB, and others that have many of the capabilities we are assessing. We chose not to include them because these tools have a steep learning curve for new users. There is also no discussion of learning platforms like CourseKata, WebAssign, and MyLab as their primary role is to provide access to course materials and assessments and not to serve as specialized plotting software.

2.1 Point-and-click Statistics Software

A popular set of tools often used in statistics classrooms is point-and-click statistics software. These tools include JMP, Minitab, StatCrunch, Microsoft Excel, Fathom, Jeffreys's Amazing Statistics Program (JASP), and many others. Due to their popularity and prevalence, these tools have been the focus of similar analyses in the past. For instance, Biehler's work on the analysis of software for learning and doing uses McNamara's criteria to compare Fathom, TinkerPlots, and Codap (Biehler 2019). We will focus on JMP, Minitab, StatCrunch, and JASP because each possesses key features of interest.

Cost is a consideration in terms of a tool's accessibility. All of the tools mentioned above, with the exception of JMP, for which a free Student Edition is available for academic use, require users to purchase licenses, which is the case for most point-and-click software. This can be cost-prohibitive for instructors and students who are looking to make and use such plots. Another consideration in terms of a tool's accessibility is its modality and compatibility with a user's operating system. Some software, like StatCrunch, are purely web-based, while others, like Fathom and JASP, require an application on the user's machine, and others, like Minitab, have both options. In terms of compatibility, the web-based software should work on all popular operating systems, yet some application-based software is not compatible with all popular operating systems. For instance, Fathom is not supported on macOS versions 10.15 and beyond. Another attribute to consider is ease of entry for novice users. Due to the high-quality user interfaces, most point-and-click software is extremely easy to use for novice users. Within minutes of gaining access, any user will be able to plot data and distributions.

Generally, point-and-click statistics software offers the user limited options to change the properties of a plot. Users can add plot annotations, change colors, and change orientation for most plots. These options will change based on the type of plot being created. For instance, in both StatCrunch and Minitab, users have all default customization options when plotting data and overlaying distributions. In contrast, their probability density plotting options have quite limited options. In terms of interactivity, the plots made by the software are usually static. However, many of the software try to make parts of the process interactive. StatCrunch gives the user the ability to click on part of a plot, such as a particular bin of a histogram, and will highlight the data that make up that bin. Minitab and JMP offer interactive graph building options where the user can drag and drop variables to explore different plotting options. Each software gives the user the ability to download the plots in multiple formats.

Point-and-click software varies widely in terms of probability distribution plotting ability and data plotting and fitting ability. Some software, such as JMP, Minitab, and StatCrunch, possess valuable functionalities for both fitting curves to existing data and for plotting distributions. Minitab includes options to plot single or multiple distributions and shade areas under curves. StatCrunch and JMP include a probability calculator that will calculate probabilities associated with a chosen distribution while displaying a corresponding plot of the probability distribution with the proper area shaded. Of the four pieces of software in this comparison, JASP offers the most capability for distribution plotting and fitting by far. JASP has a distributions module that allows the user to interact with 40 different distributions. The user can plot the density or mass function, cumulative distribution function, and a quantile function if the

Attributes	*DistPlotter*	JMP	Minitab	StatCrunch	JASP
Accessibility	Free to use; web-based	Regular software not free to use; Student Edition free to use; download required; works on Macintosh and Windows operating systems	Not free to use; web-based	Not free to use; web-based	Free to use; download optional; works on multiple operating systems; webbased option available
Easy entry for novice users	✓	✓	✓	✓	✓
Plot interactivity	Can change plot aspect ratio, axis labels, text properties, geometries (curves and bins); can change distribution parameters	Can change plot color, orientation, and annotations; offers interactive graph builder mode to explore different plot types	Can change plot color, orientation, and annotations; includes interactive graph builder mode to explore different plot types	Can change plot color, orientation, and annotations	Can customize axis properties (title, breaks, plotted range, etc.); can interact with graphics by adding shaded areas under density curves and histograms
Distribution plotting	66 distributions plottable individually; can plot multiple distributions on a single graph	26 distributions plottable individually using probability calculator	24 distributions plottable individually; can plot two distributions on the same graph	16 distributions plottable individually using probability calculator	40 distributions plottable individually
Data plotting and fitting	✓	✓	✓	✓	✓

Table 1: A comparison of the *DistPlotter*, JMP, Minitab, StatCrunch, and JASP tools.

distribution is continuous. In addition, the user can calculate probabilities and shade proper areas under the density/mass. The module also allows the user to fit each distribution to data using the maximum likelihood estimate for each parameter.

A comparison of the proposed *DistPlotter* application with four other point-and-click tools – JMP, Minitab, StatCrunch, and JASP – is provided in Table 1.

2.2 Web-based Calculators

Desmos and GeoGebra are two popular, interactive web-based programs commonly used in mathematics and statistics classrooms. While not the only options for web-based calculators, these two offer distribution plotting features. Both programs are free-to-use and require just a web browser for use. Both also offer application versions for mobile devices and are easy to use; even novice users will be able to understand and use the software in minutes.

Desmos allows the user to change a plot's aspect ratio, add plot labels, and background properties, and it also allows the user to download a plot as a png image or the entire report as a pdf file. GeoGebra's probability calculator, by default, offers no customization options to the user. To save and download a plot, the user must create a free account.

Desmos has built-in functions for a few common distributions – the normal, t, uniform (continuous), binomial, and Poisson – and the user can overlay as many distributions as they see fit. GeoGebra has a probability calculator in its calculator suite that allows users to visualize and calculate probabilities from 14 different distributions. The probability calculator also has a feature that will overlay a normal distribution on the distribution curves. Neither allows for the inclusion of data or the option to fit distributions to data, however.

Attributes	*DistPlotter*	Desmos	GeoGebra
Accessibility	Free to use; web-based	Free to use; web-based	Free to use; web-based
Easy entry for novice users	✓	✓	✓
Plot customization and interactivity	Can change plot aspect ratio, axis labels, text properties, geometries (curves and bins); can change distribution parameters; can download plots as png's	Can change plot aspect ratio, axis labels, and calculator background; can download plots as png's or pdf's	No ability to customize plots; must create a free account to save and download plots
Distribution plotting	66 distributions plottable individually; can plot multiple distributions on a single graph	7 distributions plottable individually; can plot multiple distributions on a single graph	15 distributions plottable individually using probability calculator
Data plotting and fitting	✓		

Table 2: A comparison of the *DistPlotter*, Desmos, and GeoGebra tools.

A comparison of the proposed *DistPlotter* application with two web-based calculators, Desmos and GeoGebra, is provided in Table 2.

2.3 Statistics and Probability Web Applications

A variety of web applications exist that have the capability to plot distributions. Two of note, due to their focus on probability distributions, are the Statistics Online Computational Resource (SOCR) project and the Probability Distributome project (Dinov et al. 2015). All websites provided by these projects are freeto-use and require only a web browser. The tools in these projects are all extremely easy to use, and even novice users will be able to understand and use the software in minutes.

The customization capabilities for these two projects are quite limited. The SOCR project and Distributome project do not allow users to customize plots. In contrast to their limited customization capabilities, these projects are very interactive. Each has the capability to dynamically change the properties of the plot based on user input. For instance, both the Distributome and SOCR projects allow users to change the parameters of a distribution with sliders. Additionally, the SOCR project allows users to hover their mouse over a plot and see important information. For example, users can hover their mouse over probability mass functions to see the probability at that specific value or probability density functions to see the value of a density. In terms of downloading the resulting plots, both projects allow users to download plots as png's.

Each application offers users the ability to explore and plot a wide array of probability distributions. The SOCR project houses several different applications that include probability distributions. The distribution applications, which most closely resemble our application, have 75 distributions listed. The previous versions of the applications were powered by Java and are now deprecated as of 2021. However, replacements were built using R. The SOCR project also has applications that plot the bivariate normal distribution and previously plotted mixture distributions (built using Java), to list just a few capabilities. The Probability Distributome project houses calculator, simulator, and experiment applications for 47 different probability distributions. The project also includes an interactive navigator application that allows users to interact with properties of a distribution and visualize relations between distributions. However, the Distributome project does not currently have the capability to fit distributions to usersupplied data. The Distributome project contains experiment applications that will simulate data from a distribution and then overlay that same distribution on the data, but it will not accept user-defined data. The SOCR project, on the other hand, has multiple applications that plot user-defined data. The same application that plots univariate distributions also has a tab where users can fit a distribution to data, although the focus seems to be on simple linear regression between a predictor and a response. The project also includes an application called SOCR Analytical Toolkit (SOCRAT) that has the capability to accept user-defined data and overlay distributions on that data as well.

A comparison of the SOCR Distribution and Distributome Project's applications with our DistPlotter application can be found in Table 3.

Attributes	*DistPlotter*	SOCR Distribution application	Distributome Project applications
Accessibility	Free to use; web-based	Free to use; web-based	Free to use; web-based
Easy entry for novice users	✓	✓	✓
Plot customization and interactivity	Can change plot aspect ratio, axis labels, text properties, geometries (curves and bins); can change distribution parameters; can download plots as png's	No ability to customize plots, but allows for interaction at every level; can download plots as png's	No ability to customize plots, but allows for interaction at every level; can download plots as png's
Distribution plotting	66 distributions plottable individually; can plot multiple distributions on a single graph	75 distributions plottable individually using probability calculator	47 distributions plottable individually
Data plotting and fitting	✓	✓	

Table 3: A comparison of the *DistPlotter*, SOCR web apps, and Distributome Project tools.

3. DESCRIPTION OF THE DISTPLOTTER APPLICATION

We now describe the *DistPlotter* application and its ability to make plots of univariate distributions. The application, which was created using the *Shiny* package (Chang et al. 2026) in R (R Core Team 2025), can be accessed for free online at https://educationapps.shinyapps.io/distPlotter. While the application incorporates various R packages, including *rio* (Chan et al. 2023), *VGAM* (Yee 2025), and *extraDistr* (Wolodzko 2026), among others, the most notable is *ggplot2* (Wickham 2016), which is used to create the high-quality plots. After launching the application, users are taken to the "Home" tab, which specifies the features offered in the three tabs used to make plots: "Plot a Single Distribution," "Plot Multiple Distributions," and "Include Your Data." We now describe the features implemented in these three tabs in depth.

The application enables users to plot numerous univariate distributions – 66 total, 48 of which are continuous and 18 discrete, including various options not implemented in other software already discussed – and was designed with a particular focus on education. Users can plot distributions based on either shape or a common named discrete or continuous family. They also have the ability to shade area(s) underneath curves (such as areas corresponding to general probabilities, *p*-values, or confidence levels) and customize plots. Additionally, users can add multiple superimposed density curves to the same plot or even plot their own data, overlay a fitted density curve (found via maximum likelihood estimation), and shade area(s) underneath. When downloading the plot, users can elect for a transparent background, something particularly desirable if they plan on inserting the plot into a file (e.g., a slide) with a non-white background. We now further describe these features.

3.1 Plotting a Single Distribution

Users can plot a single distribution in the "Plot a Single Distribution" tab. There, they can plot either a probability density function (PDF) as a density curve or a probability mass function (PMF) as a histogram. They can also customize plots using a variety of options. A screenshot of this tab can be seen in Figure 1.

Distribution Options

Users first select whether they want to plot a distribution with a general shape or one from a chosen family. If they select to work with a general shape, for which a density curve is plotted, the following shapes and corresponding parameters are offered:

- symmetric #1 (a normal distribution)
- symmetric #2 (a continuous uniform distribution)
- symmetric #3 (a mound, which is a beta distribution with two shape parameters larger than one)
- symmetric #4 (a tub, which is a beta distribution with two shape parameters between zero and one)
- skewed (a general unimodal option with four parameters: minimum, maximum, and two shape parameters that allow for symmetry or skewness), and
- bimodal (four parameters: a mean and standard deviation for each of two subpopulations).

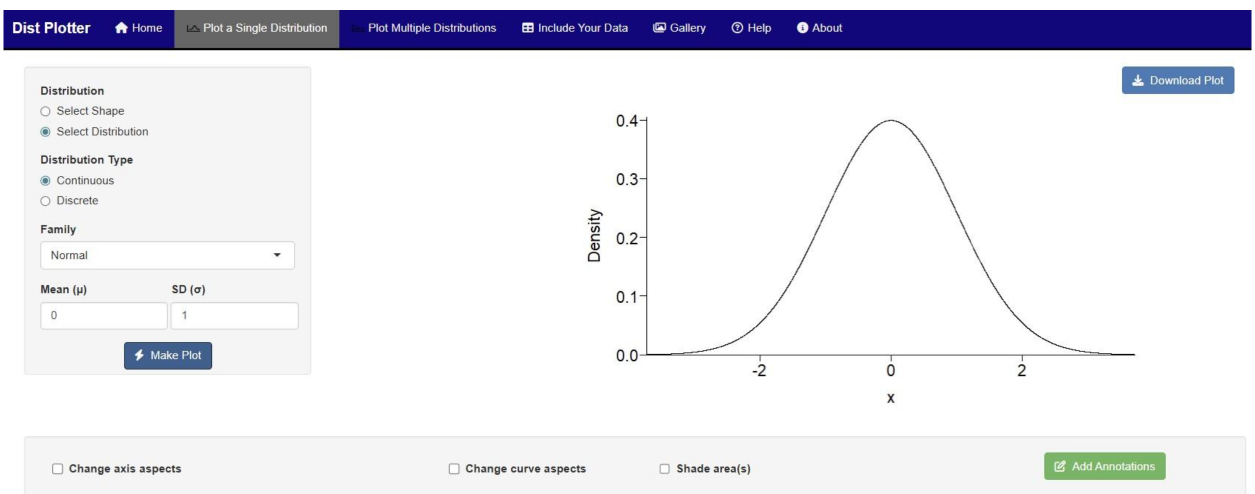

Figure 1: Screenshot of the "Plot a Single Distribution" tab. Inputs for the distribution can be found in the left-hand well panel, with the generated plot appearing on the right. Plot customization options can be found in the bottom well panel. While the plot shown is not customized, various options to do so can be viewed when the appropriate checkboxes are checked.

If they instead prefer to select a named distribution, users then choose the type of distribution – i.e., continuous or discrete. The families of 48 continuous distributions which users can work with include: ARGUS, beta, beta prime, Birnbaum-Saunders, Burr, Cauchy, chi, chi-square, exponential, *F*, Fréchet, gamma, generalized beta, generalized extreme value (GEV), Gompertz, Gumbel, half-Cauchy, half-normal, half-t, Huber density, inverse-gamma, Kumaraswamy, Landau, Laplace, Lévy, log-normal, logistic, lomax, Maxwell, Nakagami, noncentral *t*, normal, Pareto, PERT, power, Rayleigh, reciprocal, skew-normal, skewed *t*, slash, *t*, Tracy-Widom, triangular, truncated normal, Tukey lambda, uniform (continuous), Wald, and Weibull. The families of 18 discrete distributions available to users include: Benford, Bernoulli, betabinomial, binomial, Delaporte, geometric, hypergeometric, logarithmic, negative binomial, Poisson, Skellam, truncated binomial, truncated Poisson, uniform (discrete), zero-inflated binomial, zero-inflated negative binomial, zero-inflated Poisson, and Zipf.

Plot Customization Options

After making a base plot of the distribution of interest, users have various options available for customizing the plot. For instance, users can change aspects of the axes (such as the minimum, maximum, locations of the tick marks, label, and text size) or remove them entirely on an individual basis (e.g., they might remove the *y*-axis line, label, and tick marks). Users can also shade areas of interest and change colors. The application additionally offers an annotation feature, which enables users to add text, line segments, or arrows to plots. Further, whenever adding text to a plot – either as a label or an annotation – users can include special symbols and text options, such as Greek letters, a hat or horizontal bar above a letter, mathematical notation such as ≥, ≤, and ≈, subscripts, superscripts, bold text, and italicized text.

3.2 Plotting Multiple Distributions

The option to plot multiple distributions in the same plot is available in the “Plot Multiple Distributions” tab. There, users can include multiple superimposed density curves in the same plot, including curves from the aforementioned 48 different families. After creating a plot, users can continue to add more curves or update existing curves as desired instead of needing to start from scratch each time. The distributions available are those from the named continuous families listed in Section 3.1. In terms of customization options, users can change any of the axis elements mentioned in Section 3.1, as well as

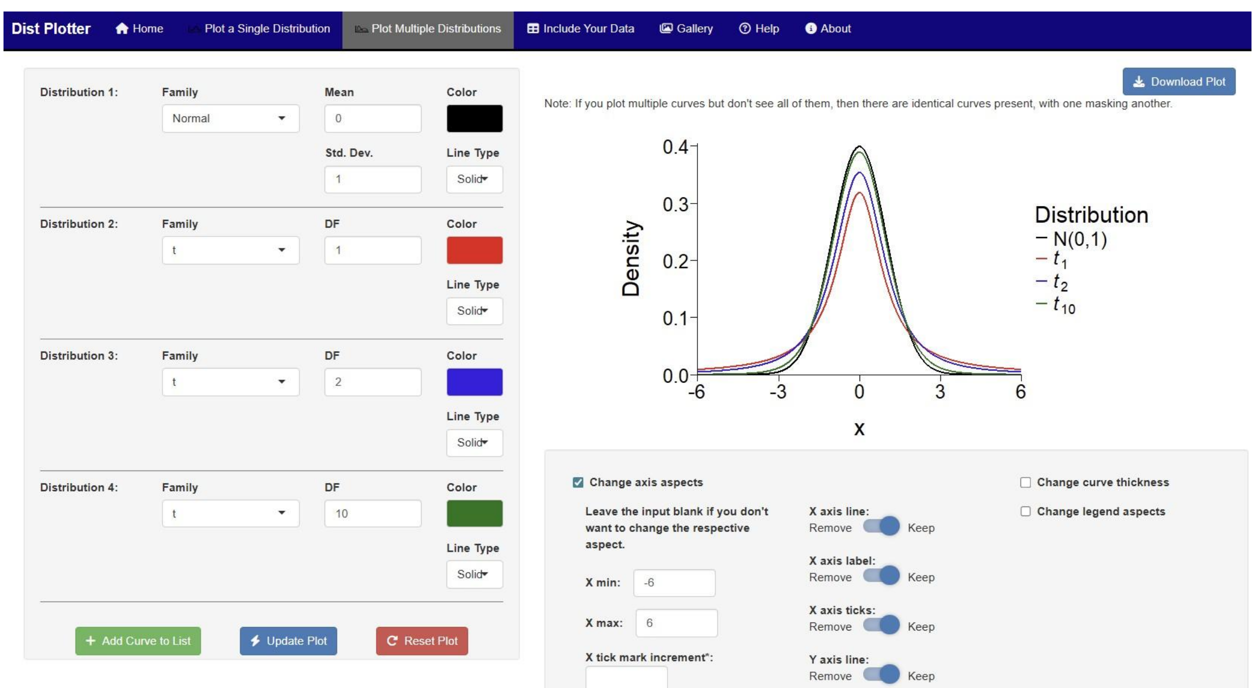

Figure 2: Screenshot of the “Plot Multiple Distributions” tab. Inputs for the distributions can be found in the dynamic left-hand well panel, with the generated plot appearing on the right. Plot customization options can be found in the well panel located below the plot.

aspects involving the legend – the main label, the label size, the text size, and the color of each curve. By default, the application includes text – using common statistical notation – in the legend for each distribution plotted. Figure 2 includes an example of multiple distributions plotted at once, which could be used to show, for instance, how *t* distributions converge to the standard normal distribution.

3.3 Including User Data

If users have their own quantitative data that they wish to plot, they can do so in the "Include Your Data" tab. There, they input their data in the "Input Data" tab by either uploading a dataset or manually inputting it. For anyone uploading a dataset, the application is compatible with various common file types used with datasets, including .csv (CSV), .xlsx (Excel), .mat (MATLAB), .R (R), and .sav (SPSS), among others listed in the application. After inputting their data, users continue to the "Make Plot" tab, where they select the variable of interest and then make a histogram. An example can be found in Figure 3.

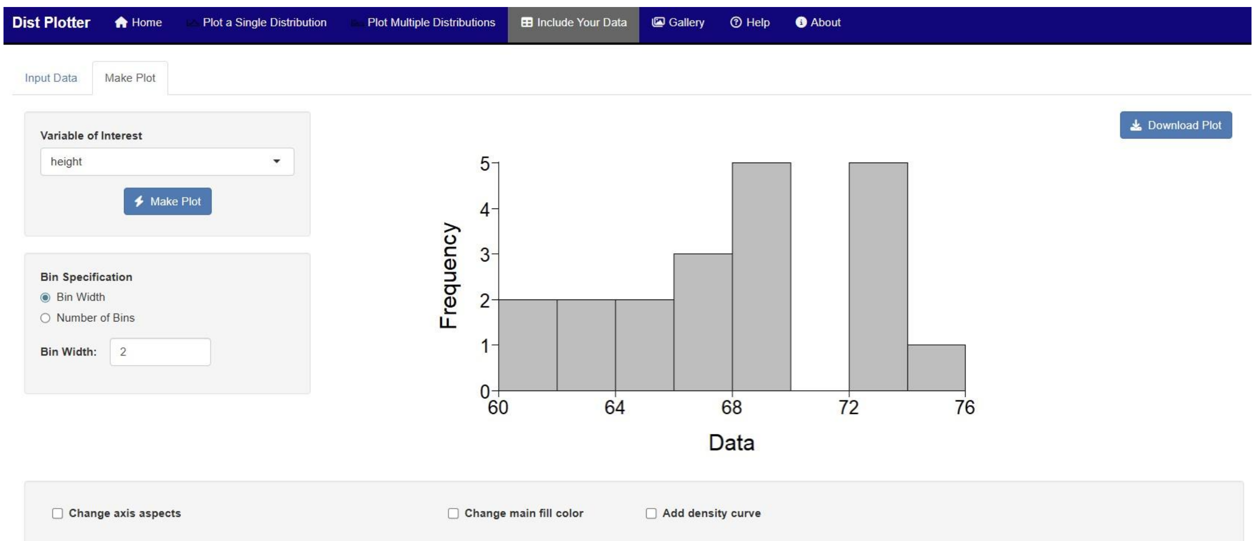

Figure 3: Screenshot of the "Include Your Data" tab. Inputs for the variable and bins of the histogram can be found in the two left-hand well panels, with the generated plot appearing on the right. Plot customization options can be found in the well panel located at the bottom of the screen. While the plot shown is not customized, various options to do so can be viewed when the appropriate checkboxes are checked.

After making a histogram of their data, users can adjust various aspects of the plot or even add to it. For example, users can specify the bin width (or number of bins) used in the histogram, add a density curve on top of the histogram, shade area(s) under the density curve (along with selecting the color and opacity used in order to make both the shaded area and the histogram underneath it visible), and customize various aspects of the plot, such as the axes (using the same customization options offered in the other main tabs) and the color of the histogram's bars. When it comes to adding a density curve, users can choose from a list of 34 different continuous families, with the application also displaying the maximum likelihood estimates of the respective parameters given the user's data. This list of 34 stands out in particular due to the fact that only JASP offers similar extensive capability.

Section 4 discusses the intended audience for the tool and contains a set of examples that the authors have found to work well in their own classrooms and that highlight the different features of the application.

4. INTENDED AUDIENCE AND SAMPLE CLASSROOM ACTIVITIES

4.1 Intended Audience

We believe that *DistPlotter* can be used across the statistics curriculum by both students and instructors, with potential applications in courses ranging from AP Statistics and undergraduate introductory statistics to more advanced undergraduate courses such as probability and statistics for engineers, mathematical probability, and courses about Bayesian statistics.

In the K-12 setting, it is uncommon for students and instructors to know a tool, like R, that could facilitate comprehensive explorations of probability distributions. Further, learning to use those tools, which often require a steep learning curve, would take away valuable learning time. Many of the same issues exist for undergraduate introductory statistics courses. Although instructors may be more likely to be familiar with R, this cannot be assumed, as many instructors teaching undergraduate introductory statistics courses are not trained statisticians (Soler 2010). Furthermore, the use of programming-based tools in an introductory course may place additional cognitive demands on students and could hinder their learning of basic statistical concepts (Meletiou-Mavrotheris 2003). *DistPlotter* offers a free, easy-entry experience to gain insight about common probability distributions and fitting a probability model to data.

In more advanced courses, both instructors and students are likely to be more comfortable with programming-based tools. Even so, we still believe *DistPlotter* could be useful in these courses. *DistPlotter* includes a large collection of probability distributions that allows for the easy exploration of concepts found in these more advanced courses, such as the relationships between various distributions. Further, an interactive tool such as *DistPlotter* allows students to explore these relationships without the extra cognitive burden resulting from needing to write and modify their own code.

As described in Section 3, the three main features offered by the application are: plotting a single distribution, plotting multiple distributions, and including user data. We now provide a set of examples that highlight the utility of these features by demonstrating how they can help deepen student understanding of important concepts in the statistics classroom.

4.2 Understanding Shape, Center, and Variability

An important set of concepts in the introductory statistics classroom is understanding the shape, center, and variability of real data and, later in the course, probability distributions. It is also a set of concepts that can be difficult for students to grasp, especially when trying to identify or understand them graphically (Cooper and Shore 2008; Kaplan et al. 2014). One example where we observe this in our courses is when introducing the normal or Gaussian family of distributions. Depending on the level and focus of the course, this might be the first time students are introduced to the concept of a probability distribution, and it is likely to be the first instance of working with continuous probability. Thus, it is understandable that students struggle with new notation, a new concept of probability, and new problem types. However, we find it interesting that students struggle with the graphical representation and interpretation of the distribution's parameters and probabilities. For instance, after several guided examples of parsing the $N(\mu, \sigma)$ notation and using the requisite software or table correctly, we found that students can reliably calculate normal probabilities in a variety of different problem setups. Those same students, though, were less likely to successfully illustrate the probability on a normal distribution curve (i.e., draw a correct distribution curve and shade the proper area under it). Additionally, those students could understand that the two parameters define the mean and standard deviation and could also substitute those values into equations, such as when calculating z-scores, but some could not connect the parameters to graphical elements like the peak of the distribution.

Motivated by the difficulties described above, one possible assignment instructors could implement is to ask students to use the "Plot Multiple Distributions" tab in the *DistPlotter* application to plot a variety of normal distributions by varying the values of the mean and standard deviation. An example of a plot created for this activity can be found in Figure 4. After the exploration, the instructor could ask the students to articulate what role the mean and standard deviation play in determining the shape of the normal distribution. By enabling students to simulate the behavior of such distributions interactively, our hope is that the students would gain a more complete understanding of the distributions' behavior as well as connect the formulaic and graphical representations.

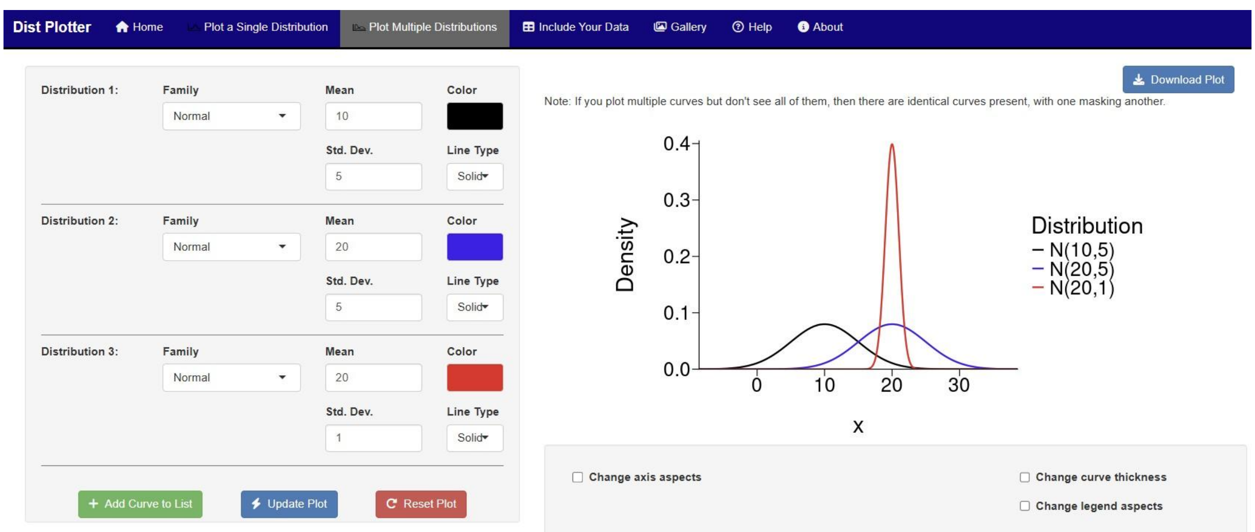

Figure 4: Screenshot of the "Plot Multiple Distributions" tab, with a focus on exploring the role of the mean and standard deviation for normal distributions.

4.3 Modeling Probability Distributions

We have found uses for the three main features in the *DistPlotter* application in a variety of different situations throughout the statistics curriculum. For instance, at all levels, the ability to model the natural variability in data with a probability distribution is an important skill that students need. It is also a skill that can be aided by using visual tools. When students are learning the concept of a probability distribution, we have found it helpful for students to explore overlaying fitted densities to match the pattern of data and really drive home the fact that probability distributions are meant to model real-life processes. Further, at the introductory level, we have also found plotting helpful when connecting various important concepts: population distributions, sample/data distributions, sampling distributions, and hypothesis testing. And in more advanced courses, such as those covering Bayesian methods, where modeling with probability distributions takes a more prominent role, the *DistPlotter* application can also be employed. We now provide examples of how the application can be used in classroom settings.

Sampling Distribution Activity

As a concrete example of the utility of the application in an introductory statistics classroom, consider a semester-long activity involving the heights of students in a class, where the class will be treated as the population. At the beginning of the semester, the instructor could collect data on the students, in particular each student's height, and compile the data in a dataset. When covering univariate distributions, the instructor could have their students use the *DistPlotter* application to make a histogram of the heights and explore what histogram settings, like number of bins or bin width, make the most readable graph. Figure 5 shows a histogram of height data for a class of twenty students. Later in the semester, when students have been exposed to probability distributions, the instructor could return to the activity and ask students to replot their data, this time with a normal density curve overlaid, as Figure 6 displays. Students would also be asked to assess the normal curve's fit and potentially calculate an area under the curve.

Figure 5: A histogram of the heights of the students in a relatively small class.

Next, when introducing the concept of the sampling distribution of the sample mean, the instructor could extend the activity again. They could ask students to make a list of all samples of two heights (to keep things manageable) when sampling with replacement and then calculate the mean of each sample. To facilitate this process for a larger class, the instructor could divide the work into small groups of students and then combine the resulting data into a dataset that they then would share with the class. Students would then input the dataset containing the sample means into the application and make a histogram with a normal distribution overlaid, as is seen in Figure 7. The instructor could next ask the students to explore the relationship between the population distribution examined previously and the sampling distribution of the sample mean, and even ask students to calculate areas under the curve. Even further, when introducing the idea of hypothesis testing, the instructor could extend the activity once again by presenting a scenario of interest that involves testing the mean height of the entire class (while emphasizing hypothesis testing involves testing an *unknown* population mean). They could then connect sampling distributions to hypothesis testing by asking students to plot the sampling distribution of the sample mean, add a normal curve to the plot, and shade the area under the normal density curve that corresponds to the area of interest. An example of such a plot can be found in Figure 8. The instructor could also ask students to comment on whether the shaded area accurately reflects the *p*-value, as well as what could improve the accuracy and reliability of the results (such as an increase in the sample size).

Figure 6: A histogram of the heights of the students in a relatively small class, with a poorly-fitting normal density curve overlaid.

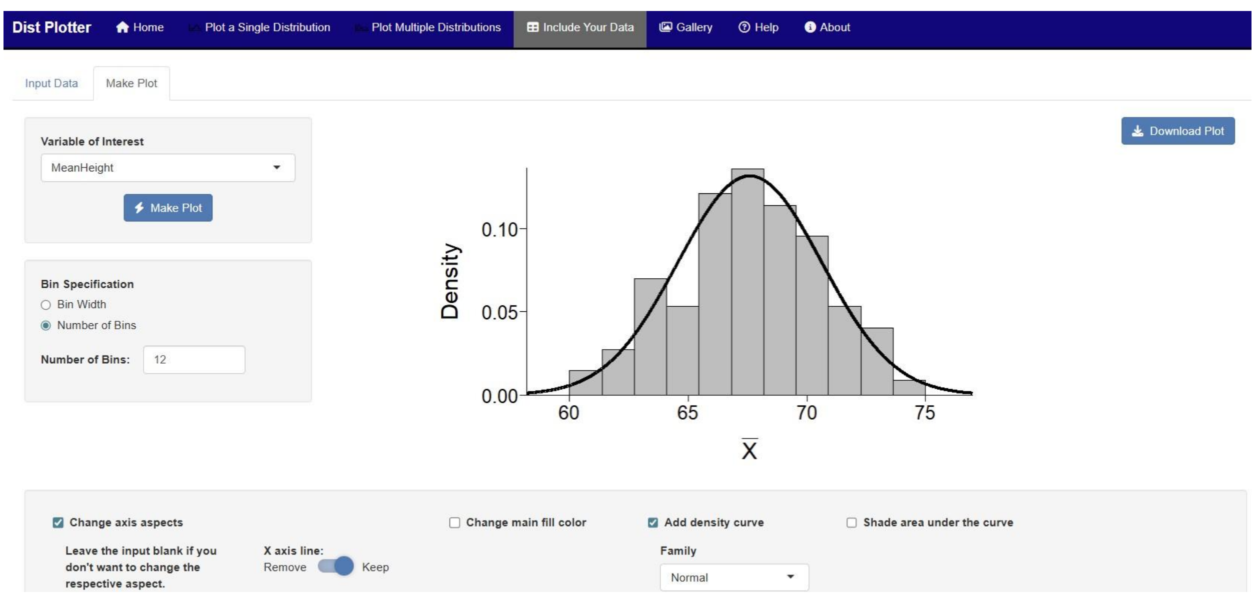

Figure 7: A histogram of the sampling distribution of the sample mean for the height data seen in Figures 5 and 6, assuming a sample size of two and sampling with replacement. A normal curve is overlaid on the plot, with the parameters determined by maximum likelihood estimation.

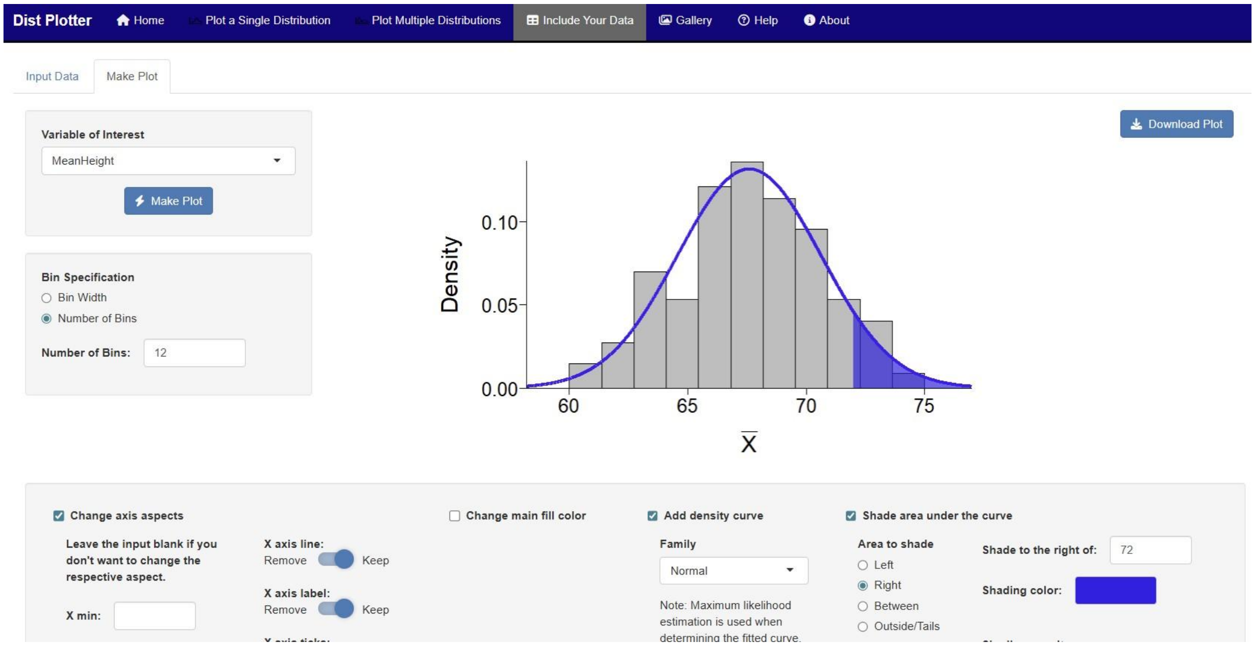

Figure 8: A plot of the sampling distribution of the sample mean based on the height data seen previously, assuming a sample size of two and sampling with replacement. The area under the normal curve corresponding to the *p*-value for a hypothetical scenario is shaded. Students could be asked to assess the accuracy of that value.

Bayesian Prior Exploration

The *DistPlotter* application can additionally be used in classroom settings beyond the introductory level. For instance, in a course in which Bayesian methods are covered, instructors could use the application when beginning a discussion on the choice of prior distributions and prior elicitation. They could task students with determining a reasonable beta distribution for a set of situations about probabilities. The instructor could ask their students to vary the parameters of the beta distribution until they find a plot that they think matches a situation the best. Afterward, the instructor could hold a class discussion where students defend their choice of prior compared to other choices the class produces. We find this activity to be beneficial because students are engaging in two critical steps of a Bayesian analysis: choosing a reasonable prior distribution and being able to defend the choice of a prior. The *DistPlotter* application also has value in this activity because the students can interactively update the parameters and resulting plot, allowing them to quickly explore a wide set of distribution choices. Figure 9 displays beta distributions for three potential scenarios: a situation where all outcomes are equally likely (black), a situation in which an outcome seems to be between about 25-75% likely (blue), and a situation in which an outcome is quite likely (red).

Relationships Between Distributions

Distributions, and the relationships between them, are often of interest in advanced undergraduate courses. Discussion of these relationships often centers on proving them mathematically. However, students may still lack an intuitive understanding of them. Being able to explore these relationships through dynamic visualizations can help strengthen students' understanding.

One possible assignment instructors could implement is to ask students to use the "Plot Multiple Distributions" tab in the *DistPlotter* application to plot a sequence of *t* distributions with increasing degrees of freedom and overlay the standard normal distribution. Afterward, the instructor could hold a class discussion where the students discuss how the t distribution changes as the degrees of freedom increase, particularly in the tails, and explain how these observations support the theoretical result that the standard normal distribution is the limiting distribution of the *t* distribution. Figure 9 displays an example of this.

Figure 9: A plot of three beta distributions reflecting differing prior beliefs regarding a probability.

Another possible assignment instructors could implement is to ask students to use the “Plot Multiple Distributions” tab in the *DistPlotter* application to plot a gamma distribution with a shape parameter of one and their choice of rate parameter and overlay an exponential distribution with the same rate parameter. Afterward, the instructor could hold a class discussion where students discuss how parameters of the two distributions relate and how exponential is a special case of the gamma distribution. Figure 10 highlights this relationship by showing only one curve, but in reality there are two identical curves for the two distributions. A note above the plot is included as a reminder of such possible behavior.

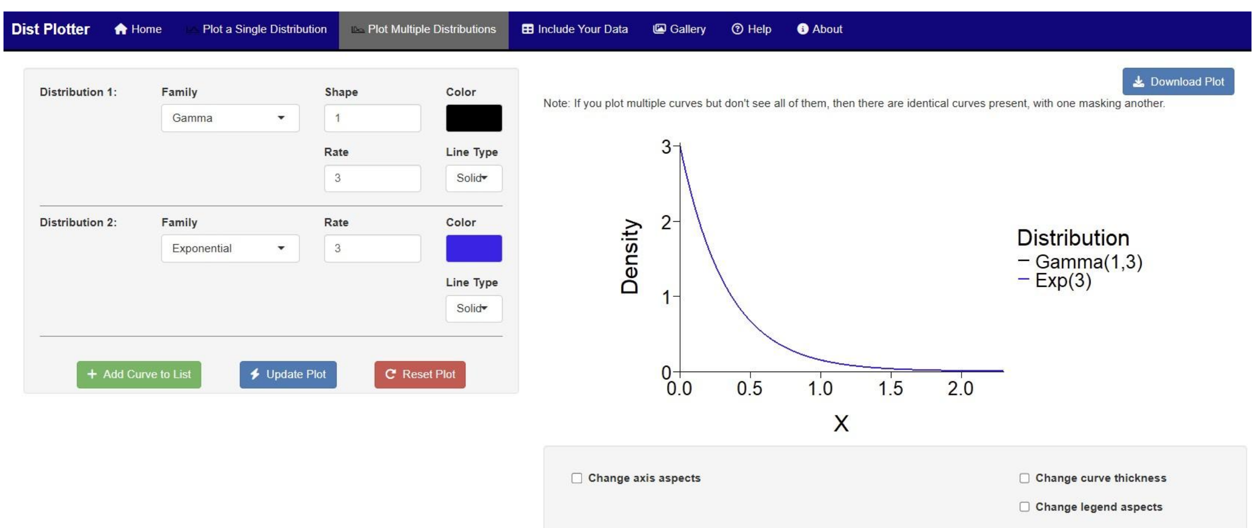

Figure 10: A plot of a gamma distribution with a shape parameter of one and a rate parameter of three as well as an exponential distribution with the same rate parameter of three. Only one curve is displayed due to the two distributions being identical.

5. CONCLUSION

In this article, we discuss the coverage of distributions in the classroom, existing tools for making plots of such distributions, and limitations to these tools. We also describe a free, online, point-and-click application we developed to create high-quality plots of distributions for various instructional purposes in probability and statistics courses ranging from the introductory to the advanced levels. The application innovates upon existing point-and-click tools by combining the following, all in a single application: (1) the ability to access a large collection of distributions for a variety of different probability distribution-specific tasks, (2) ease-of-use, and (3) an increased level of interactivity, which is often seen only in more technical and comprehensive tools.

We believe that our application is innovative and useful because it moves the field closer to what an optimal distribution visualization tool should be, while also providing various distribution options not found in other software. In the era of AI tools, such as ChatGPT, our application encourages hands-on interaction. Students cannot rely on AI tools to complete tasks within the application and must instead interact with it directly to explore distributions, adjust parameters, and interpret visualizations.

In Section 2 of this work, we put forth a framework for evaluating such tools, and this framework guides our view of what an optimal tool looks like. The tool should be free, easy-to-use, highly interactive, have the capability to plot probability distributions, and have the capability to plot and fit a distribution to data. The design principle of such a tool should be to minimize a user's interaction with the machinery of the tool without sacrificing the user's ability to customize options or interact with the output. Additionally, the user should be able to easily navigate through intuitive menus to make any distribution plot they want without needing an intricate knowledge of the tool. Further, the user should be able to easily customize a plot to illustrate a certain feature or relationship without a significant tool-specific learning curve. Lastly, the tool should include features that can enhance the user's knowledge about distributions without forcing those features on them.

While we believe that our application meets most of the criteria for an optimal tool, we also acknowledge the application's limitations. For instance, the current application does not allow the user to vary the parameters of the distribution fit to data in the data fitting feature, and it also does not allow the user to plot discrete and continuous probability distributions together in the plotting multiple distributions feature. These two capabilities were lower on our priority list and more challenging to implement. The application can also experience performance issues as the number of users increases, so we suggest students using the app in larger classes work together in groups. There also remain potential areas for future work beyond addressing the limitations. In terms of studies, one might involve asking educators and/or students for their feedback after using the application, while another could evaluate the impact of the application on student learning of the associated concepts. Additional areas of future work include the implementation of cumulative distribution functions (CDFs) in the application, the calculation of probabilities, and the development of another application designed specifically for bivariate distributions.

REFERENCES

Biehler, R. (1997), "Software for Learning and for Doing Statistics," *International Statistical Review*, 65(2), 167-189.

Biehler, R. (2019), "Software for Learning and for Doing Statistics and Probability - Looking Back and Looking Forward from a Personal Perspective," In J. M. Contreras, M. M. Gea, M. M. López-Martín, and E. Molina-Portillo (Eds.), Proceedings of the Third International Virtual Congress of Statistical Education. University of Granada.

Biehler, R., Frischemeier, D., Reading, C., and Shaughnessy, M. (2018), "Reasoning About Data," In D. Ben-Zvi, K. Makar, and J. Garfield (Eds.), International Handbook of Research in Statistics Education (pp. 139-192): Springer International.

Bobek, E., and Tversky, B. (2016), "Creating Visual Explanations Improves Learning," *Cognitive Research: Principles and Implications*, 1.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Rowell, G. H., Velleman, P., Witmer, J., and Wood, B. (2016), "Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016," American Statistical Association.

Chan, C. H., Leeper, T., Becker, J., and Schoch, D. (2023), rio: A Swiss-army Knife for Data File I/O, R package version 1.0.1.

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007), "The Role of Technology in Improving Student Learning of Statistics," *Technology Innovations in Statistics Education,* 1(1).

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Aden-Buie G., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2026), shiny: Web Application Framework for R, R package version 1.13.0.

Cooper, L., and Shore, F. S. (2008), "Students' Misconceptions in Interpreting Center and Variability of Data Represented via Histograms and Stem-and-Leaf Plots," *Journal of Statistics Education*, 16(2), 1.

Dinov, I. D., Siegrist, K., Pearl, D. K., Kalinin, A., and Christou, N. (2015), "Probability Distributome: A Web Computational Infrastructure for Exploring the Properties, Interrelations, and Applications of Probability Distributions," *Computational Statistics*, 31(2), 559-577.

Friel, S. N., Curcio, F. R., and Bright, G. W. (2001), "Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications," *Journal for Research in Mathematics Education*, 32(2), 124-158.

Gentleman, J. F. (2012), "It's All a Plot (Using Interactive Computer Graphics in Teaching Statistics)," *The American Statistician*, 31(4), 166-175.

González, M. T., Espinel, M. C., and Ainley, J. (2011), "Teachers' Graphical Competence," In C. Batanero, G. Burrill, and C. Reading (Eds.), Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education (pp. 187-197). Dordrecht/Heidelberg/London/New York: Springer.

Green, J. L., and Blankenship, E. E. (2015), "Fostering Conceptual Understanding in Mathematical Statistics," *The American Statistician*, 69(4), 315–325.

Kaplan, J. J., Gabrosek, J. G., Curtiss, P., and Malone, C. (2014), "Investigating Student Understanding of Histograms," *Journal of Statistics Education*, 22(2).

McNamara, A. (2018), "Key Attributes of a Modern Statistical Computing Tool," *The American Statistician*, 73(4), 375-384.

Meletiou-Mavrotheris, M. (2003). Technological tools in the introductory statistics classroom: Effects on student understanding of inferential statistics. *International Journal of Computers for Mathematical Learning*, *8*(3), 265-297.

Pea, R. D. (1987), "Cognitive Technologies for Mathematics Education," In A. Schoenfeld, *Cognitive Science and Mathematics Education* (pp. 89-122). Erlbaum.

R Core Team (2025), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Ridgeway, J. (2015), "Implications of the Data Revolution for Statistics Education," *International Statistical Review*, 84(3), 528-549.

Repenning, A., Webb, D., and Ioannidou, A. (2010), "Scalable Game Design and the Development of a Checklist for Getting Computational Thinking into Public Schools," *SIGCSE '10: Proceedings of the 41st ACM Technical Symposium on Computer Science Education.* https://dl.acm.org/citation.cfm?id=1734357.

Soler, F. P. (2010). Who Is Teaching Introductory Statistics? *The American Statistician*, *64*(1), 19–20. https://doi.org/10.1198/tast.2010.09183

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer.

Wild, C. J., Pfannkuch, M., Regan, M., and Horton, N. J. (2011), "Towards More Accessible Conceptions of Statistical Inference," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2), 247-295.

Wolodzko, T. (2026), extraDistr: Additional Univariate and Multivariate Distributions, R package version 1.10.0.2.

Yee, T. W. (2025), VGAM: Vector Generalized Linear and Additive Models, R package version 1.1-14.