



The Fourth R: Using R for Data Visualization in the Science Classroom

Brian Dennis & Lisa Washburn

To cite this article: Brian Dennis & Lisa Washburn (2026) The Fourth R: Using R for Data Visualization in the Science Classroom, The Science Teacher, 93:1, 14-23, DOI: [10.1080/00368555.2025.2589841](https://doi.org/10.1080/00368555.2025.2589841)

To link to this article: <https://doi.org/10.1080/00368555.2025.2589841>



Published online: 05 Feb 2026.



Submit your article to this journal [↗](#)



Article views: 15



View related articles [↗](#)



View Crossmark data [↗](#)

The Fourth R: Using R for Data Visualization in the Science Classroom

BRIAN DENNIS  AND LISA WASHBURN 

ABSTRACT

R, the free, open-access computer software package for scientific graphs and calculations, has become a leading tool in data science. R can be installed on most computers (or accessed for free on online servers), and the basics of its use are easy to learn. With R, students have free access to professional quality routines for data visualization, data wrangling, and data analysis. We provide two tutorial lessons on using R in the form of hands-on labs featuring earth science data (Old Faithful geyser prediction). The lessons are easily adapted for data from other STEM subjects. One of the authors (LW) successfully incorporated such lessons into live science laboratories while teaching in Idaho public schools in grades 6-12. The lessons engage numerous points in the Next Generation Standards for earth science along with points in other sciences and in Math and ELA Common Core Standards. Moreover, the lessons represent a first, inviting experience with “coding” and “modeling” that students recognize as important skills. The lessons begin with simply cutting/pasting short portions of pre-written R code, allowing students to become familiar with how R commands work. We close with a discussion of the advantages of incorporating R use in STEM education.

Keywords: Coding; computer programming; data science; functions; geyser; graphs; models; prediction; R; regression; scatter plot; visualization



R is a free, open-access computer software package for scientific graphing and calculations. It was written by scientists and is used by scientists worldwide (Vance 2009). R is not commercial, shareware, free-trial-ware, malware, or secretly-sell-your-info-ware; rather, R is maintained and distributed for free by a consortium of scientists (The R Core Team 2023). R is easy to learn and use for calculations and graphing in all high school science courses. Because R contains preprogrammed routines for virtually all statistical analysis methods and mathematical functions, R can be grasped quickly by students and can greatly enhance math-across-science and science-across-math efforts to integrate topics. Moreover, as an entry-level programming language and data management tool, R represents a great introduction to coding and data science principles. The quality and variety of data visualization routines in R are outstanding. See, for example, the use of R in a recent study of global biodiversity (Bernardo-Madrid et al. 2025).

Learning objectives

The goals of the following two Earth science labs are to introduce students to the free R programming software and to begin enabling students to create and interpret graphs of scientific data. Because the tiny bit of R presented in these lessons is focused on creating graphs from data, the learning curve is not steep. Students need only cut and paste the graphing commands provided in the labs. Students quickly become able to understand and write these commands themselves using additional sets of data acquired from other live laboratories taught in the classroom.

These labs were originally produced and taught in an Earth science course (typically ninth or eighth grade), using open-access data collected for predicting eruption times of Old Faithful geyser in Yellowstone National Park. However, converting the labs to use data from other sciences, or data collected from in-class activities, would be straightforward.

Students will:

- Use R programming to create graphs using scientific data;
- Analyze and interpret graphs created from scientific data; and
- Apply R coding and concepts to additional experiments in which scientific data are collected.

Each lab can be completed in one 50-minute classroom session. These time estimates are reliable data collected by Lisa Washburn over 10 years in public school classrooms, both brick and mortar and online. Class sizes varied from 20 in person to 100 students in the on-line classroom taught during the COVID-19 epidemic. The labs were adapted to other science courses (physical science, biology, chemistry) by substituting subject-relevant data.

Materials and methods

The materials are simply the following introductory R tutorial and Old Faithful labs, distributed as handouts or e-mailed to the students ahead of time, or made available for download, and a computer with the free R programming software installed. R versions are available for Windows, Mac OS, and Linux/Unix. Anyone can freely download and install R from the R Project website (<https://www.r-project.org/>). Installation is easy; follow directions carefully and install the correct version for the computer operating system. R is free to institutions as well as individuals, so consult your IT people at your school about making it available on your school computers.

If installing R is not an option, the teacher and students can instead link to an online free R server (<https://rdrr.io/snippets/>) opened in a separate window on their computers. Currently, R is not available for mobile operating systems (tablet/phone), but the online R server can be accessed with mobile devices.

If the teacher is new to R, know that this article is meant to be *done*, not read! Install R (or find the online server), fire it up, and type the commands as you go through this article. Make e-copies of the scripts (lists of multiple commands) in the form of plain text files to distribute to students.

In Lab 1, the teacher should begin with a brief demonstration of some capabilities of R (below), with the students repeating the commands on their machines. In Lab 2, the resulting scatter plot of the Old Faithful data will lead into a class discussion of, and subsequent hands-on lab exercise in, modeling methods used to predict the eruption times.

If you are working with an installed version of R, just click the blue R icon to launch the program. Inside the resulting R window will be a sub-window, the command console, with a prompt `>` waiting for your instructions. If you are using the online R server, just put the URL into your web browser and go. You will see a window (the snippets box) with some sample illustrative R commands. Remove the sample commands from this box (highlight them and delete) so that you can insert and run your own R commands.

Lab 1. R tutorial: Math calculations and graphs

The following presentation serves as scaffolding for Lab 2, in which the students move into more independent use of R to analyze data. In Lab 1, the teacher demonstrates simple R statements by either projecting a computer screen with the R window open in the brick and mortar, or by presenting the teacher screen to the online class, enabling online students to share their own screens if any difficulties arise. The students should type the commands on their own machines as led by the teacher. R can be used as an ultra-powerful graphing calculator. At the prompt in the command console, type: `6 + 8` and hit **enter**, or in the snippets box, type `6 + 8` and click the **Run** key:

```
6 + 8
```

```
[1] 14 (this is the printed answer: don't type this part)
```

Row 1 of the answer is 14. Answers to complicated calculations often come in many rows, and so R prints row numbers along with the answers on the screen.

Try subtraction: type `6-8` at the new prompt, or erase what is in the snippets box, type `6-8`, and click **Run**:

```
6-8
[1] -2
```

The minus sign can be used as a sign for negative numbers:

```
6+-3
[1] 3
```

Multiplication is an asterisk:

```
6*8
[1] 48
```

Division is a forward slash, so 6 divided by 8 is:

```
6/8
[1] 0.75
```

Raising to a power: “six to the eighth power,” written 6^8 , means $6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6$. The carat symbol “ \wedge ” denotes raising to a power:

```
6^8
[1] 1679616
```

You can put a string of calculations all in one statement. The calculations with `*` and `/` are done first, then `+` and `-` are done, and the calculations are done left to right:

```
6+8*2-12/4-1
[1] 18
```

Raising to a power is performed first, even before multiplication and division:

```
2+5*3^2
[1] 47
```

The order of operations can be altered by using parentheses:

```
(6+8)*2-(12/4-1)
[1] 26
```

Parentheses inside of parentheses will be done first. Just make sure that every left paren “(” has a companion right paren “)”:

```
(6+8)*(2-12/(4-1))
[1] -28
```

R will store *everything* for you. Just give your calculations names:

```
ellen=6+8
scott=5-1
ariel=ellen-scott
ellen
[1] 14
scott
```

```
[1] 4
ariel
[1] 10
```

R names are sensitive to lower and upper case. In R, the name `ellen` is different from `Ellen`.

They will disappear when you run new commands in the snippets box. If you use a version of R installed on your computer, the stored values will linger in the computer memory during your R session, but they will disappear when you exit R (unless you elect to save your workspace, which we do not recommend as it is easy to create errors in future R sessions).

You can enter multiple commands which will be executed sequentially from top down. Such a series of commands is called a *script*, another name for a computer program. Installed versions of R (Windows or MacOS) have text editor windows that can be opened for writing and running scripts. Linux R users will need a separate plain text editor, running scripts by copying and pasting them at the command console prompt. Word processors are not good places to write scripts, because they add hidden formatting codes into the scripts which cause errors when submitted to be run in R. The snippets window online accepts scripts of any length, either typed directly into the window or pasted in as plain text (Figure 1).

If you reuse a name in the same R session for something different, R will erase the old value and replace it with the new value. To illustrate, run the following six commands as a script:

```
jane=9
jane
jane=10
jane
jane=jane+1
jane
[1] 9
[1] 10
[1] 11
```

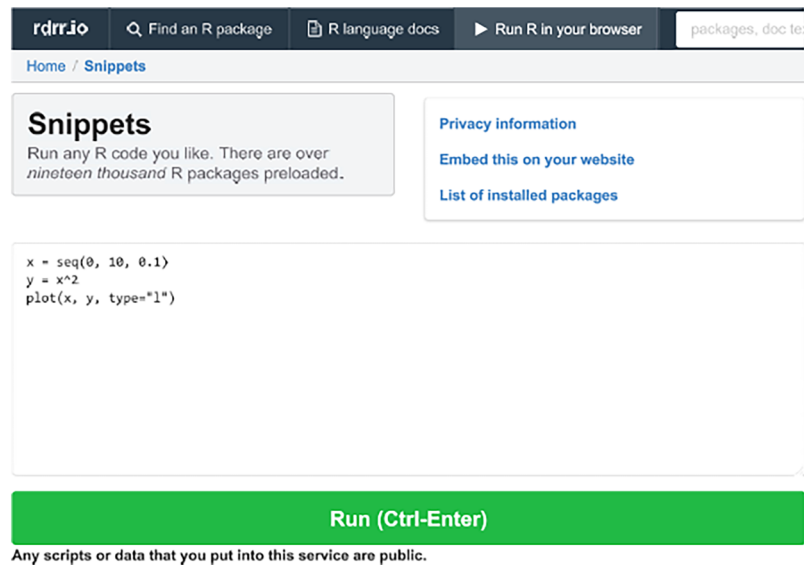
You likely will have noticed the seemingly weird command `jane=jane+1`. There of course is no real number which equals itself plus one. Interestingly, the equals symbol “`=`” in R (and in many other computer programming languages) does not mean “equals.” Instead, it means: *Calculate or do what is on the right, and store the result using the name on the left.*

A statement with the equals symbol is called an *assignment statement*. In old versions of R, assignment statements used the symbol “`<-`” instead of “`=`”, which is supposed to look like an arrow pointing left, and the old way still works in newer R versions. Many books and tutorial websites use the arrow-like symbol.

All the calculations in R can be done with vectors! A *vector* is a list of numbers and is a basic unit of data in R. An ordinary single number like “3” is called a scalar (“scale-er”). In R, calculations are vectorized: for vectors, each element of a vector is used in calculations with the corresponding elements of other vectors, and in a calculation with a scalar and a vector, the scalar

FIGURE 1

Screen shot of the “R Snippets” online server. The statements in the snippets box will produce **Figure 2** when the “Run” button is clicked.



will operate on all the elements in the vector. Vectors can be built with the “c ()” function (c for “combine”). Try this script and see if you understand each result:

```
sam = c(2, 5, 1)
bev = c(3, 4, 6)
deb = c(7, 3, 2)
pat = 2
deb
pat
sam + bev
bev / deb
sam * pat
deb + bev * sam - 1
```

In R, typed spaces are mostly ignored (but not in names of vectors, data sets, etc.). Most programmers prefer to use lots of spaces in their scripts like in the above script for easier reading.

In R, the symbol (0:10) is a special vector containing the integers from 0 to 10:

```
(0:10)
```

Do you want to see all the powers of 2 from 0 to 20? Of course you do!

```
2^(0:20)
```

R contains many mathematical and statistical functions pre-programmed and available for use. Functions often can

calculate a result for every element in a vector. For instance, let’s try the function for square root:

```
nums = (0:20)
sqrt(nums)
```

More generally, a vector containing a sequence of numbers can be produced with the seq () function, for example:

```
seq(10, 20, 0.5)
```

creates a vector starting at 10, ending at 20, with increments of 0.5. Naming the vector stores it for plotting or calculating.

Speaking of plotting, let’s try a graph! Run this script:

```
x = seq(0, 10, 0.1)
y = x^2
plot(x, y, type="l")
```

What each line of the above R graph plotting code tells R to do

The first command sets up a vector named *x* that contains the sequence 0, 0.1, 0.2, ..., all the way up to 10. The second command calculates the square of each number in *x*, storing the results in a vector named *y*. The plot () statement draws a graph with the values in *x* on the horizontal axis and the corresponding values in *y* on the vertical axis. The type="l" option (lower case letter L) in the plot () statement produces a “line plot” in which the plotted points are connected with lines but without any symbols for the points themselves. There are

many options that can be entered in the `plot()` statement for changing all aspects of the graph's appearance. The graph appears in a new window and can be saved in a variety of graphic files and formats, for further use in documents or presentations (Figure 2).

You can save the graph! In snippets, right-click on the graph and select **Save image as** from the menu. A window will pop up. In the window, select and open the folder in your computer where you want the graph saved. Choose a file name, perhaps `faithfulgraph.png`, and enter it in the area designated **File name**. Use the `.png` extension for your file name; `png` is a graphic format that is suitable for ordinary scientific graphs (for high-quality printing of high-detail figures, `tiff` format is better, but the files are often much larger). Click the **Save** button! In an installed version of R, the File menu at the top of the R window will have a **Save as** facility (but first make sure that the graph window is the active window by clicking on it).

Lab 2. Predicting eruptions of Old Faithful

Teacher and students will read out loud through the following narrative and then will cut and paste (or carefully type) the graphing R script from their lab (this document) provided by the teacher:

Yellowstone National Park sits on top of a super volcano, resulting in the park having over 500 geysers and numerous hot springs popping up through the Earth's crust. Archeological excavation revealing prehistoric tools and

artifacts indicate that people have inhabited this area for over 11,000 years! Native Americans cooked their food, bathed for medicinal purposes, and occasionally buried their dead in the bubbling waters. Yellowstone had great spiritual significance to the Native Americans, especially the Crow, who named Yellowstone *bidemahpe*, meaning sacred or powerful water.

At the Old Faithful Visitor Center, a few hundred meters from Old Faithful geyser in Yellowstone National Park, the park rangers post a prediction of the time when the next eruption of the geyser will take place, so that the crowds of visitors can gather at the right time. While the average time until the next eruption is somewhere around 75 minutes, the actual times from eruption to eruption are quite variable. Old Faithful is fairly faithful for a geyser, but not so faithful for impatient visitors on a tight schedule.

Scaffolding questions to be discussed with the class by the teacher

How do the rangers know when the next eruption will be? Do they have some sort of valve for turning on the eruption, like turning on a slide show in the Visitors Center theater? (Figure 3).

Graphing the Old Faithful eruption data

Let's use a graph to look for any useful pattern in Old Faithful eruptions. The following three lines of code are copied and pasted by both the students and the teacher into the R snippets window or into the script editor window for the installed R version.

```
attach(faithful)
head(faithful)
plot(eruptions, waiting)
```

FIGURE 2

Graph of $y = x^2$, for a range of values of x from 0 to 10, produced by the statements in Figure 1.

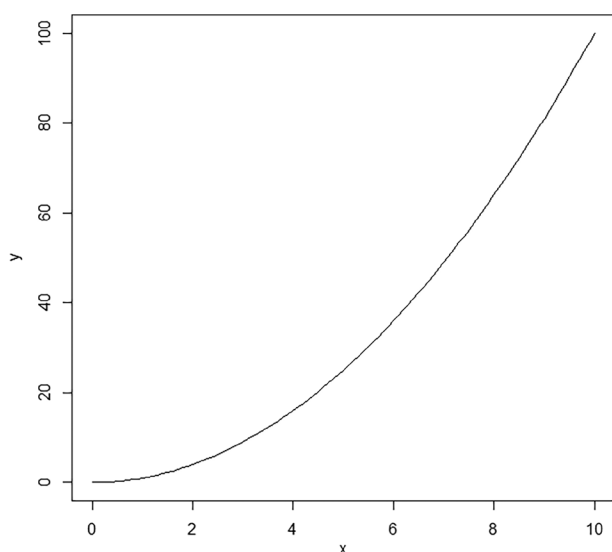


FIGURE 3

An eruption of Old Faithful geyser, Yellowstone National Park, Wyoming (photo by Brian Dennis).



What each line of code tells R to do with the Old Faithful data

R contains many sets of data that are useful for illustrating data science. One data set is named `faithful`, and it contains two vectors: `eruptions` has durations (lengths of time) of 272 eruptions (in minutes), and `waiting` has the 272 corresponding times (in minutes) until the next eruption. The above R script contains just three commands. The command `attach(faithful)` puts the two vectors in the data set into the R workspace so that they are available for calculations or plotting in R, `head(faithful)` lists the first few elements in the data for inspection, and `plot(eruptions, waiting)` plots the eruption durations on the horizontal axis and the resulting waiting times until the next eruption on the vertical axis. The type of plot is the default kind in R, called a *scatter plot*, produced when the `type=""` option is omitted from the `plot()` statement (Figure 4).

Following are formative assessment questions to be discussed with the class after all the students/groups of students have successfully run the graphing command in the R window. Refer to the Differentiation section for recommended groupings for students to work together in both the brick and mortar and online classrooms:

1. Does the graph of Old Faithful eruption data appear to show a relationship between the eruption length and the waiting times until the next eruption?
2. If there appears to be a relationship, is it shaped like a line or like a curve?
3. Are the data points on the graph relatively close to a straight line or curve, or are the data points scattered above and below a line or curve (what scientists would call a “noisy” set of data)?
4. If there were no relationship between the eruptions and the waiting period for eruptions (in other words, if scientists could not use eruptions to predict when the next eruption of Old Faithful takes place), what would the graph look like?
5. If an eruption lasted 4.5 minutes, when would you tell the visitors to expect the next eruption?
6. How do the Old Faithful eruptions relate to the Crow term *bidemahpe*?

Prediction model

The points in the Old Faithful graph almost beg for a line to be drawn through the center of the points to help our eyes perceive the relationship and to help us issue actual numerical predictions. Such a line is easy to draw in R!

But which line? The equation for a line is: $y = mx + b$, where y is the value of the quantity to be predicted (response variable), x is the quantity used for prediction (predictor variable), m is the slope of the line (how much does y change when x changes by one unit), and b is the intercept of the line (value of y on vertical axis when $x = 0$). Each pair of values of m and b give a different line; imagine lines going every which way on the graph. Which is the best line to use for prediction?

In 1805, the French mathematician Legendre proposed a solution for linear prediction, nowadays called *least squares* or *regression* (Stigler 1986). Regression is used today daily by data analysts all over the world. Legendre derived a unique line that will minimize a measure of the total amount of prediction error for a given data set. (Algebra 2 students might be curious to know that the derivation involves solving two linear equations for the two unknown values of m and b ; see Dennis 2013).

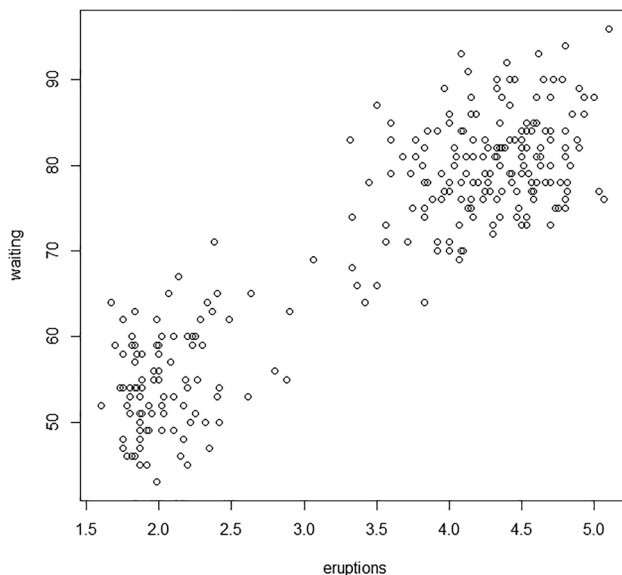
The cool thing is that we do not have to dive into any mathematical details to use this tool. R has a built-in statement called `lm()` (linear models) that will do the calculations for us.

In R snippets online, clear the snippets box. In an installed version of R, close the window of the previous graph and clear the script editor window. We will expand the tutorial graphing script slightly to plot the data and draw a prediction line. The response variable data will be the vector `waiting`, the predictor variable will be `eruptions`. Here is a small R script that will do the trick for the Old Faithful data (note to teacher: Again, students will cut and paste the following 7 lines of R code into the R window and hit the “Run” button. Make sure that any other scripts in the R window are deleted before pasting these new lines of code!):

```
attach(faithful)
head(faithful)
results = lm(waiting ~ eruptions)
summary(results)
plot(eruptions, waiting)
abline(results)
detach(faithful)
```

FIGURE 4

Old Faithful scatter plot.



The `lm()` statement builds a list of calculation results for the prediction analysis and stores it in the computer. The list here is named “results”, but any valid R name can be used. The “`waiting ~ eruptions`” notation means “use eruptions to predict waiting.” The `summary()` command prints a statistical summary calculated from the list named `results`, the `plot()` command makes the now familiar scatterplot of the data, and the `abline()` command draws the prediction line from the `results` list on the scatter plot (Figure 5). The `detach(faithful)` command removes the vectors in the `faithful` data set from the R workspace so that the workspace is clean and fresh for the next analysis.

In the printout of numerical results (Table 1), the intercept and slope of the prediction line are the first column (labeled “Estimate”) of numbers in the sub-table labeled “Coefficients.” The intercept is the top number, and the slope is the bottom number. The table has many other details calculated that would be explained in a statistics or data science course. One detail in particular is the “adjusted *R*-squared” near the bottom. It is a number between 0 and 1 measuring the quality of prediction for the data used, with 1 being perfect (all the points are on the prediction line) and 0 being useless (predictor variable is no help for prediction). The value of around 0.81 suggests that the line gives pretty good predictions for the Old Faithful data.

FIGURE 5

Old Faithful scatter plot with prediction line.

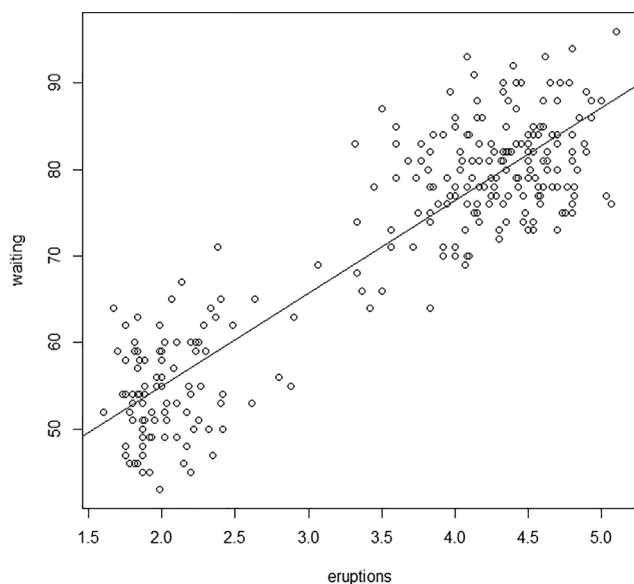


TABLE 1

Numerical output of the linear models script for the Old Faithful data.

```
Call:
lm(formula=waiting~eruptions)
Residuals:
    Min       1Q   Median       3Q      Max
-12.0796 -4.4831  0.2122  3.9246 15.9719
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.4744     1.1549   28.98  <2e-16 ***
eruptions   10.7296     0.3148   34.09  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5.914 on 270 degrees of freedom
Multiple R-squared:  0.8115, adjusted R-squared:  0.8108
F-statistic: 1162 on 1 and 270 DF, p-value: < 2.2e-16
```

Summative assessment

Students will use the R graphs they created to answer the following questions:

1. What are the values of m and b that give the best prediction line for the Old Faithful data?
 $m =$
 $b =$
2. What is the equation of the prediction line (insert the values of m & b that you calculated above into the slope/intercept equation $y = mx + b$)?
3. If the duration of an eruption is increased by 1 minute, how much longer must we wait until the next predicted eruption?
4. An eruption lasts 4.5 minutes, ending at 1:22PM. You are the ranger. How many minutes from now should park visitors expect the next eruption? (Use R to calculate the answer! Show the R command you used).
5. Is the above prediction perfect? What would the graph look like if the prediction were better? What would a worse prediction look like?
6. Look at the scatter of points on the graph directly above and below the height of the line at eruption length of 4.5 minutes. By inspection, could you recommend a “window” of time for predicting the next eruption that would include most if not all the waiting values occurring for an eruption value of around 4.5 minutes?
7. In a study involving data from 23 colleges, a line for predicting college GPA from a student’s high school SAT score would have an adjusted *R*-squared value of around 0.12 (Bridgeman et al. 2000). What would you conclude about the predictive value of using SAT in college admissions?

Assessments

Multiple types of assessments are provided for both Laboratories 1 and 2. Formative assessment points are provided in the labs using suggested questions for the teacher to present to the students. Formative assessments are also as simple as “can the students get the copies lines of code to run in the R program” or can the students explain in their own words what the R code is asking R to do, and can they apply and use the R code to other sets of scientific data.

The summative assessment takes the form of a quiz or test after both labs have been completed, and requires the student: (a) to analyze and evaluate the graph they have created using R programming and the Old Faithful eruption data, (b) to use the linear model results to calculate the mathematical formula for the slope of the line on that graph, lending a “teaching across the curriculum” theme to this lesson, and (c) to bring ELA standards into the lesson as relating to equity, diversity and inclusion readings and discussions (Table 2).

Differentiation

In the brick-and-mortar setting, students diagnosed with dyslexia or other learning disabilities are paired with students not diagnosed with learning disabilities to correct any R coding or data interpretation errors. If working online, students not able to get the R program to run correctly or with graph interpretation issues can share their computer screen and ask other students in the class to help troubleshoot the problem. The teacher can also share the demonstration onscreen, with students following and repeating on their own home computers. The modes of collaboration between students will depend on the features of the online platform. ELL students are accommodated during the laboratories with the definitions of basic mathematical terms defined by the teacher as they are presented in R. Google translate can also be used to aid ELL students.

Diversity, equity, and inclusion

These labs embed scientific investigation with human stories. Note the initial discussion in Laboratory 2 relating to the spiritual and daily living significance of Yellowstone National Park and its bubbling waters. Formative and summative evaluation questions in this laboratory can increase students’ level of engagement by encouraging students to discover relevance between what they are learning and their personal lives (O’Donnell et al. 2021). Summative evaluation question 7 combines scientific investigation with human stories and seeks to dispel the myth that students with lower SAT scores are less able to succeed in college.

Concluding remarks

Intriguing aspects of R seem to have almost revolutionary potential for high school and early college STEM education:

TABLE 2

Grading rubric for Old Faithful laboratory.

Graphing data using R (10 points possible)

Student successfully ran R graphs = 10 points _____

Student did not attempt to run R graph = 0 points _____

Total points for graphing: _____

Summative Assessment Questions (7 points possible)

Student answered 7 questions successfully _____

Student answered 6 questions successfully _____

Student answered 5 questions successfully _____

Student answered 4 questions successfully _____

Student answered 3 questions successfully _____

Student answered 2 questions successfully _____

Student answered 1 question successfully _____

Total points for answering questions: _____

Total points on Old Faithful laboratory (17 points possible) _____

- (1) R use could help students greatly improve their understanding of quantitative STEM concepts. With R, scientific calculations and graphs are fun and easy to produce. A student using R is freed to focus on the scientific and mathematical concepts without having to pore through a manual of daunting lists of calculator keystroke instructions.
- (2) R could be used across a wide variety of STEM courses, promoting the integration of STEM subjects that has been much discussed in principle but elusive in practice. R skills would follow a student from course to course. Teach it once, use it everywhere.
- (3) For students, R is now a nearly universally available tool for scientific computation. All that is needed is access to a computer. Many students nowadays access computers for social

media and written assignments, and many schools and colleges make institutional machines available to the students. R runs on most platforms, so R can be made instantly available to every student in every course. No additional proprietary software or calculators need to be purchased. Once R is installed on a machine, the internet is not needed to use R.

- (4) R invites collaboration among students. Like scientists sharing their applications, students can work in groups to conduct projects in R, build R scripts, and improve each other's work, and they can collect, accumulate, and just plain show off exemplary graphical analyses. Results on a computer screen are much easier to view in groups than results on a calculator, and R scripts are much easier to save and alter cooperatively than calculator keystroke lists. At home, students can message each other about their latest R scripts, working cooperatively online. Computer-savvy students will turn up additional interesting and useful R resources in their work on class projects, and instructors can then have them share their discoveries with the class members. Every new class can take what previous classes have done and build new accomplishments upon the old. As scientists have found, *R builds on itself*.
- (5) R skills will follow a student to college and professional life. College statistics and advanced science courses are increasingly teaching R, and a student entering college with some R knowledge will have a head start. Many forward-looking companies are now using R (Vance 2009), and R skills are a valuable professional credential. The fast-growing field of data science, which combines advanced statistics, data management, and intensive computation, uses R as a main tool. High school students using R would be analyzing data and depicting equations just as scientists are doing in labs all over the world.
- (6) Students learning some R will be learning coding structures present in most computer languages, such as assignment statements, built-in functions, custom functions, if-then-else conditional structures, loops, and logical comparisons. Also, R has easily implemented facilities for input and output of data files, and excellent routines for "data wrangling" (merging, subsetting, etc.).
- (7) This article covers just the tip of the tip of the iceberg of R capabilities. However, there is excellent, free documentation available online. Many college science and statistics courses have posted free tutorial websites (a few examples are listed in "On the web," below). The tutorials vary considerably in quality, and the sheer quantity of web resources for R is bewildering, but a little exploration will turn up something suitable for every purpose. Tutorial videos on YouTube and elsewhere are burgeoning. Also, there is an extensive selection of instructional R books that are inexpensive paperbacks (examples are de Vries and Meys 2015; Dennis 2013; Adler 2012; Kabacoff 2011; Leemis 2022).
- (8) The authors strongly recommend sticking to what is now called "base R" for teaching beginning programming skills.

Base R is the standard R package downloaded from the R project website. Many add-ons to R have become popular in recent years, especially RStudio (a development environment for R work) and the "tidyverse" (a suite of prewritten data management functions). These add-ons are emphasized (and base R deemphasized) in many current instructional media. However, these add-ons are better suited to

TABLE 3

Next-generation STEM standards addressed by these lessons (from <https://www.nextgenscience.org/>) and Common Core Math and ELA standards (<https://www.sde.idaho.gov/academic/shared/math/ICS-Mathematics.pdf>).

Next generation stem: Science and engineering standards

Developing and using models

Modeling in grades 9–12 builds on K–8 experiences and progresses to using, synthesizing, and developing models to predict and show relationships among variables between systems and their components in the natural and designed world(s).

- Develop a model based on evidence to illustrate the relationships between systems or between components of a system. (HS-ESS2-3), (HS-ESS2-6).

Analyzing and Interpreting Data

Analyzing data in 9–12 builds on K–8 experiences and progresses to introducing more detailed statistical analysis, the comparison of data sets for consistency, and the use of models to generate and analyze data.

- Analyze data using tools, technologies, and/or models (e.g. computational, mathematical) in order to make valid and reliable scientific claims or determine an optimal design solution (HS-ESS2-2).

Common core standards

Math

8.F. Use functions to model relationships between quantities

S-ID Interpret linear models

ELA

ODC.8.2. Analyze the purpose of information presented in diverse media and formats (e.g., visually, quantitatively, orally) and evaluate the intent (e.g., social, political, commercial) behind its presentation.

intermediate and advanced users who have already learned programming basics. The books listed above mainly cover base R.

- (9) In the authors' experience, the students greatly enjoy R and recognize their new skills as genuinely important in their futures. Various students will engage with R on their own, once initial exposure has been experienced, for class projects and outside interests. With R, students will become the teachers, and teachers will become the students. In STEM, educators can hardly ask for more (Table 3).

R is more available to students, is far cheaper, is easier to use, is more powerful, is more transportable across STEM courses, and is a more valuable STEM career skill than calculators or commercial software products. So why not use R in science and mathematics courses in high school and early college?

ON THE WEB

R for Biologists at U Tennessee: <http://www.nimbios.org/products/RforBiologistv1.1.pdf>

Intro to R at the R Project website: <http://cran.r-project.org/doc/manuals/R-intro.html>

DataCamp has some free R resources: <https://www.statmethods.net/r-tutorial/index.html>

TutorialsPoint provides an extensive free R study: <https://www.tutorialspoint.com/r/index.htm>

GeeksforGeeks will take you from beginning R to data science: <https://www.geeksforgeeks.org/r-tutorial/>

The R Graph Gallery provides code examples for a vast array of data displays: <https://r-graph-gallery.com/index.html>

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

- Adler, J. 2012. *R in a Nutshell*. 2nd ed. Sebastopol: O'Reilly Media.
- Bernardo-Madrid, R., M. González-Suárez, M. Rosvall, M. Rueda, E. Revilla, M. Carrete, J. L. Tella, J. Astigarraga, and J. Calatayud. 2025. "A General Rule on the Organization of Biodiversity in Earth's Biogeographical Regions." *Nature Ecology & Evolution* 9 (7): 1193–1204. <https://doi.org/10.1038/s41559-025-02724-5>.
- Bridgeman, B., L. McCamley-Jenkins, and N. Ervin. 2000. *Predictions of Freshman Grade-Point Average from the Revised and Recentered SAT I: Reasoning Test*. New York: College Board.
- de Vries, A., and J. Meys. 2015. *R for Dummies*. 2nd ed. Chichester: John Wiley.
- Dennis, B. 2013. *The R Student Companion*. Boca Raton: CRC Press.
- Kabacoff, R. I. 2011. *R in Action: Data Analysis and Graphics with R*. Shelter Island: Manning.
- Leemis, L. M. 2022. *Learning Base R*. 2nd ed. Overland Park, KS: Ascended Ideas.
- O'Donnell, C., E. Prather, and P. Behroozi. 2021. "Making Science Personal: Inclusivity-Driven Design for General Education Courses." *Journal of College Science Teaching* 50 (3): 68–77. <https://doi.org/10.1080/0047231X.2021.12290510>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Stigler, S. M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge: Harvard University Press.
- Vance, A. 2009. "Data Analysts Captivated by R's Power." *The New York Times*, January 7, 2009, B6. <http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?pagewanted=all>

© 2025 National Science Teaching Association

Brian Dennis (brian@uidaho.edu) (ORCID: 0000-0001-7917-1424), a professor emeritus in the Department of Mathematics and Statistical Science and the Department of Fish and Wildlife Sciences at the University of Idaho, has taught undergraduate courses featuring R for two decades and has authored over 100 scientific publications. **Lisa Washburn** (ORCID: 0009-0006-1755-4191), a retired high school science teacher and the NSTA Shell Lab Challenge Regional Winner (Idaho, Montana, and Wyoming) in 2018, taught chemistry, biology, earth science, and physical science in both brick-and-mortar and online high schools in Idaho, featuring R use in all her courses.