

Third_exercises

Joe Thompson

2018-07-25

Specifying function inputs

You may have already noticed that a single line of code can become awkward or hard to read. For example,

```
summary(rnorm(10,0,1))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.4835 -1.2834 -0.7316 -0.3435  0.3217  2.2344
```

This code randomly samples ten data points from the standard normal distribution, and then provides summary statistics for the sample. One method for making this code more readable is to use the named arguments available within `rnorm`. For example,

```
summary(rnorm(n=10,mean=0,sd=1))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.9993 -0.5300  0.3701  0.1301  0.6728  0.9927
```

Using these arguments can make code much more readable, and mistakes less common. These names are specified by the `rnorm` package, as can be seen if you use the help command.

You may have noticed, however, that some functions do not take numbers as inputs. For example, consider the `sample` function.

```
?sample
```

`Sample` chooses `n` items from the vector `x`, but has a parameter called 'replace', which will result in sampling with replacement when `replace=TRUE` e.g.,

```
sample(5,5,replace=FALSE)
```

```
## [1] 1 3 4 2 5
```

```
sample(5,5,replace=TRUE)
```

```
## [1] 1 4 5 2 3
```

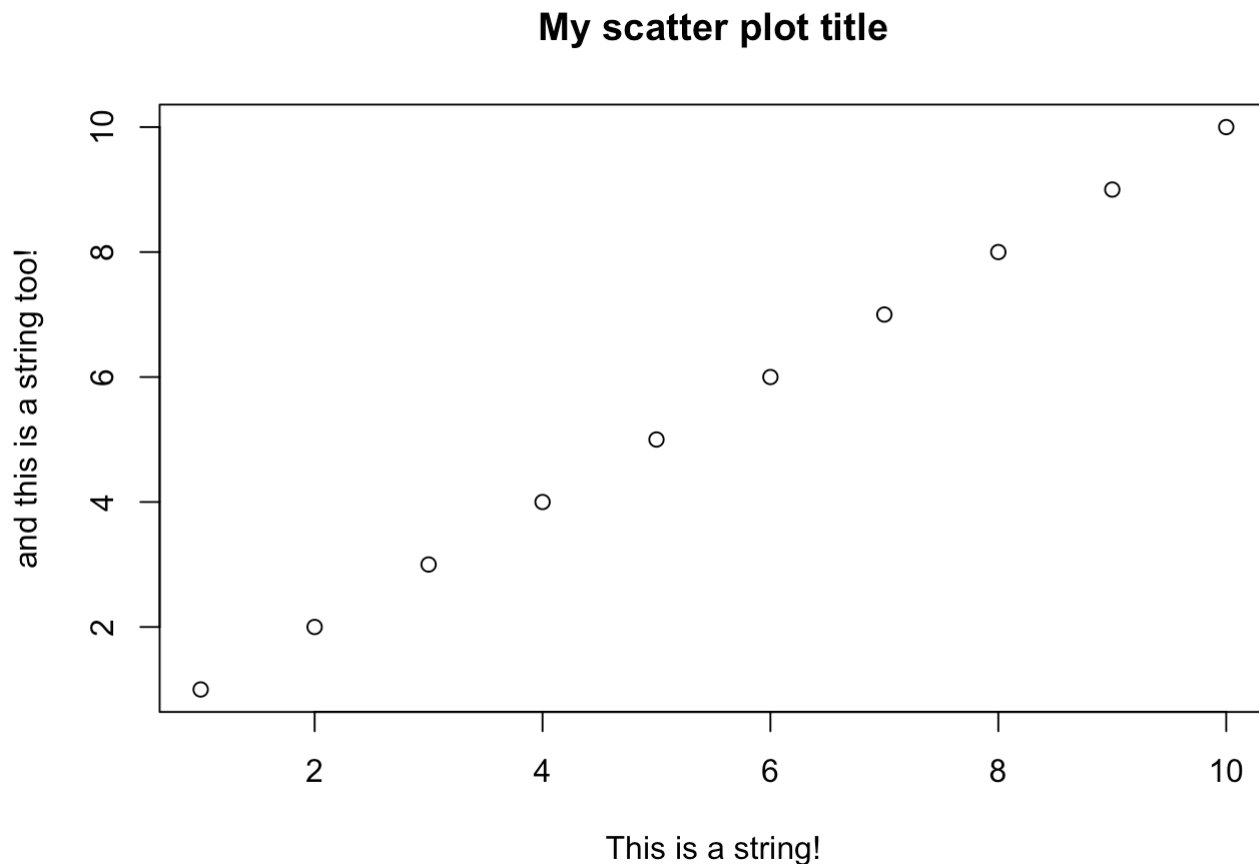
Notice that `TRUE` is not a number, but a logical.

```
class(TRUE)
```

```
## [1] "logical"
```

In other words, some functions may use numbers as inputs and others may use other data types, such as logicals. Some functions require factors or categorical variable inputs, or even formula. We will talk about those other data types throughout the workshop. Here we will focus on strings, which are required by many functions, including plot.

```
plot(x=seq(1,10), y=seq(1,10), main='My scatter plot title', xlab='This is a string!', ylab='and this is a string too!')
```

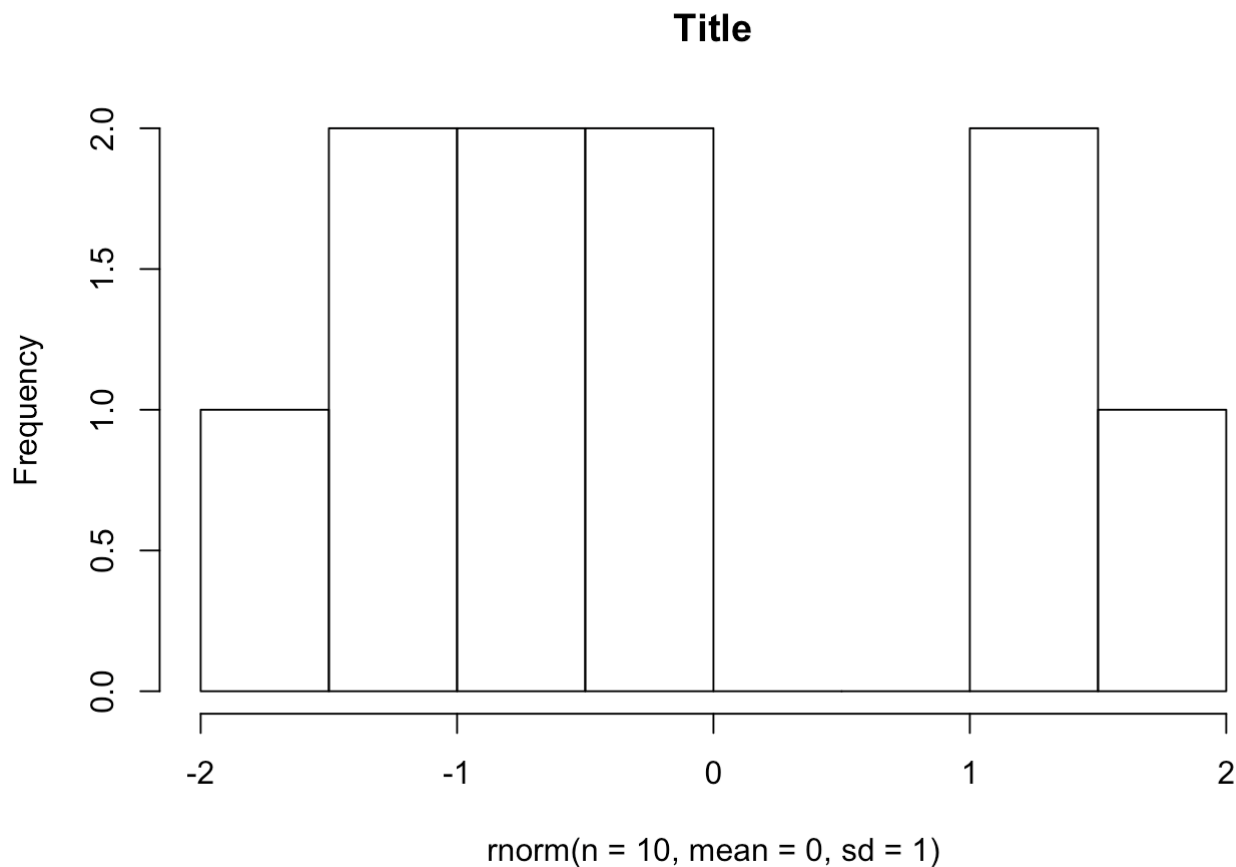


The plot function, not surprisingly, calls for a string of characters when one is specifying a title or an axis label. In order for R to know that you are specifying string data, you need to enclose your strings in quotation marks. This means that

```
hist(x=rnorm(n=10,mean=0, sd=1), main=Title)
```

will result in an error if R cannot find a variable in the workspace called 'Title'. If enclosed in quotation marks, however, R will interpret 'Title' as a string.

```
hist(x=rnorm(n=10,mean=0, sd=1), main='Title')
```



More advanced use of strings

It is possible to create an array of strings. Compare, for example, examine this string, and then close the view window.

```
MyString='study study study'  
View(MyString)
```

Compare this string of three word to this array of three strings.

```
MyString=c('study','study','study')  
View(MyString)
```

Sometimes it is also useful to create a string from a variety of data types. For example, if one wants to enclose a mean in a histogram title, they can do so using the paste function.

```
Data=rnorm(100,0,1)  
Title=paste('My sample of the standard normal, sample mean =', round(mean(Data), digits=  
2))  
Title
```

```
## [1] "My sample of the standard normal, sample mean = 0.06"
```

Paste can also be used to create long strings out of many substrings. e.g.,

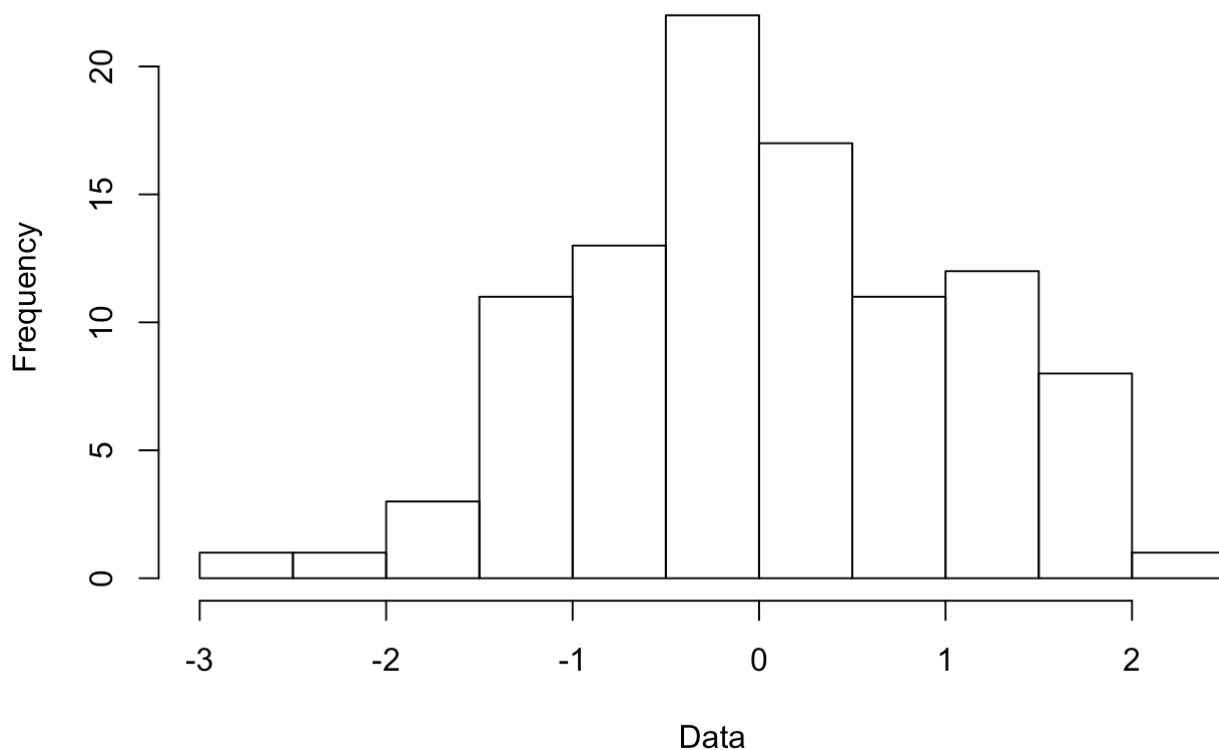
```
paste('The', 'dog', 'went', 'to', 'the', 'market')
```

```
## [1] "The dog went to the market"
```

Thanks to the paste function, the variable 'Title' will be a string that can be used as input to a histogram.

```
hist(Data,main=Title)
```

My sample of the standard normal, sample mean = 0.06



Exercises

1. Use named arguments to randomly sample data from any distribution you like.
2. Create a plot with a user-specified title.
3. I sometimes like to teach students about the gambler's fallacy by having them play rock-paper-scissors against a computer program which randomly samples one of three strings: 'rock', 'paper', or 'scissors'. Students who initially try to predict the computers next move based on its previous move are succumbing to the Gambler's fallacy. Use the sample function and your knowledge of strings to create this program.
4. Try creating a vector of Forty strings. The first twenty should say 'Control Condition' and the second twenty should say 'Experimental Condition'
5. [hard] Joe likes having the sample mean in the title of his histogram, but he wants it to simply be enclosed in parantheses. e.g., 'My sample of the standard normal (0.01).' How can Joe do this using the paste command and without specifying '0.1' manually. Hint: By default, the paste function will put spaces between the strings you input to the function. To complete this exercise you will need to look at ?paste for details about how to change this default separation.