

Learning R

Q-Step Labs

Installing R

R is a free software package and programming language for statistical computing. For this course, we'll be running R in the software environment "RStudio". To download and set up RStudio on your own computer/laptop, visit: <https://www.rstudio.com/products/RStudio/> and follow the instructions.

Basic Commands

Command	What it does
<code>class(x)</code>	See the class of an object
<code>cor.test(x,y)</code>	Conduct a correlation test between variables between x and y
<code>head(data)</code>	Shows first six rows of the data
<code>hist(x)</code>	Create a histogram of a variable
<code>length(x)</code>	Compute the length of a vector
<code>ls()</code>	Lists all the objects that are stored in your environment
<code>mean(x)</code>	Compute the mean of a variable
<code>median(x)</code>	Compute the median of a variable
<code>quantile(x)</code>	Compute the quantiles of the variable
<code>read.csv(data)</code>	Load data into R (remember to assign this to an object, by using " <code><-</code> !")
<code>scatterplot(y x)</code>	Create a scatterplot that show the relationship between variables x and y
<code>sum(x)</code>	Calculate the sum of the vector's values
<code>summary(data)</code>	Compute summary statistics of the data/variable (mean, standard deviation, etc.)
<code>sd(x)</code>	Compute the standard deviation of a variable
<code>tail(data)</code>	Shows last six rows of the data

- To look up how to use commands/options, use the help function in R: `help(command)` or `?command`.
- Use "\$" to call a variable. For example, if I have a dataset called `data` that contains a variable called `economic_growth`, use `dataset$economic_growth`. (HINT: you can use auto-complete by using the `tab` button on your keyboard once you have entered the first letter of the variable name after \$.)
- Use "<-" to assign a value or name to an object. For example, to assign the value 5 to an object called `a`, use `a <- 5`.
- The general formula for running a regression is: `lm(dependent_variable ~ independent_variable)`.
- When you want to calculate the mean/median/standard deviation of a variable that has missing values, use the `na.rm=TRUE` option. E.g. `mean(dataset$economic_growth,na.rm=TRUE)`.
- To combine multiple plots, use the `par(mfrow=c(number_of_rows,number_of_columns))`.

Extra Resources

Online

There are a number of websites that offer interactive courses. Have a look at:

- Datacamp: <http://www.datacamp.com> (see in particular “Introduction to R”, “Intermediate R”, and “A Hands-On Introduction to Statistics with R”.)
- CodeSchool: <http://tryr.codeschool.com/>

You may also want to make use of the following resources if you have specific questions:

- Quick R: <http://www.statmethods.net/>
- Stackoverflow: <http://stackoverflow.com/questions/tagged/r>
- R-Seek: <http://rseek.org>

Interactive courses (Swirl)

Swirl (www.swirlstats.com) is an R package that offers interactive courses that can be run directly from RStudio. You may find the following courses useful (note that we’ll recommend specific courses for each lab):

- R Programming
- Data Analysis
- Regression Models

To set up Swirl, use the following commands:

```
install.packages("swirl") #installs the swirl package on your computer/laptop  
library(swirl) #loads the package  
install_from_swirl("Course Name", mirror = "bitbucket") #installs the specific course  
swirl() #starts up swirl and allows you to select from installed courses
```