TDU

BOATSWAIN

Good day!

You are new to R - Welcome aboard! We are pleased to show you the ropes.

We've curated some self-study resources to help you get started. These resources are divided into three sections:

Mandatory Sea Time: The Essentials

These resources contain information that is essential to getting started with R. For instance, if you don't know how to install packages or import data you won't get very far at all. These are items that are foundational to the self-study materials captured in the next section. Please review the description of each learning item to see if it covers information that is new to you, or information where you could use a refresher.

Swab The Decks: Foundational Concepts

These resources are required learning for this self study module. These items contain information that is foundational, along with practice, to becoming proficient in R. Please review the description of each learning item to see if it covers information that is new to you, or information where you could use a refresher.

There Be Treasure: Going Further

These are optional resources to help your further your R knowledge based on your own interest level. These items are foundational to becoming an absolute swashbuckler in R!

Mandatory Sea Time: The Essentials

Virtual Training with the Training and Development Unit

Duration: **5 minutes**

This video provides an overview of the Training and Development Unit's approach to training in the virtual environment: https://www.youtube.com/watch?v=lTDS6KgIGPg

Finding Success with Virtual Training

Duration: **4 minutes**

This video provides an overview of how learners can find success in virtual training with the Training and Development Unit: https://www.youtube.com/watch?v=CNo4rIg4nU8

Install R and RStudio

Learners should have R (4.0.2) and RStudio (1.3.959.0) installed on their computers prior to joining the course. Those joining from CFEP will have this software pre-installed on their PHAC computers. Others joining from other areas within PHAC will need to open a ticket with the National Service Desk to have this software installed on their machines in advance. In the event individuals wish to join the course using computers not issued by PHAC, the following tutorial will aide them in installing R and RStudio on their computer themselves. https://techvidvan.com/tutorials/install-r/#install-r-windows

Overview of RStudio User Interface

Duration: **6 minutes**

The RStudio interface makes R much more user friendly and allows for many different aspects of the program to be available at the user's fingertips through a series of panels. However, this setup still may be foreign and intimidating to the novice user. This YouTube video indicates what is available where: https://www.youtube.com/watch?v=5YmcEYTSN7k

Install R Packages

This tutorial reviews how to install packages in R. It is critical that learners understand how to install packages prior to joining the course. Note: this tutorial covers features that will not be required in the course (e.g., R-Forge, Bioconductor, Jupyter Notebook): https://r-coder.com/install-r-packages/

Importing Data into R

Importing data into R is the first step of any data analysis and is therefore important to be comfortable with prior to starting the course. This tutorial covers several different ways of loading data into R, including flat-format files (e.g., csv, txt), files from Excel (e.g., .xlsx, .xls),

loading r-data files (.rds) and provides additional links for connecting and loading data sets from common database programs (e.g., MySQL). Learners for the Introduction to R for Public Health Investigations course should focus on loading flat files and files from Excel: https://uc-r.github.io/import

Swab The Decks: Foundational Concepts

What is Data Wrangling?

Duration: **9 minutes**

Data wrangling is too often the most time-consuming part of data science and applied statistics. Two tidyverse packages, tidyr and dplyr, help make data manipulation tasks easier. Keep your code clean and clear and reduce the cognitive load required for common but often complex data science tasks. This video is the first in a series of four, reviewing concepts such as tibbles, viewing data, the pipe operator, and data wrangling generally.
https://www.youtube.com/watch?v=jOd65mR1zfw

Dates in R

Working with dates and times can be one of the most challenging parts of any statistical software too. The following tutorial covers the basics of date, time, and date-time classes in R, and is a good place to start to begin learning about these concepts, and why they can be so challenging: https://www.neonscience.org/resources/learning-hub/tutorials/dc-convert-date-time-posix-r

The **lubridate** package, written for the expanded tidyverse, was created to help alleviate some of the common frustrations in working with date and time data. Please review Chapter 10.3 of the following resource to gain an overview of common functions within the **lubridate** package: https://bookdown.org/hneth/ds4psy/10-3-time-lubridate.html

Working with dates and times can be challenging and complex – so please don't feel that you need to master these concepts before joining the course. If you want to learn more about working with these data types – check out the additional resources included at the end of this pre-course self study guide!

ggplot

Duration: **5 minutes**

This YouTube video reviews the use of the ggplot2 package to create a scatter plot and histogram. This video makes use of a pre-installed dataset allowing for viewers to follow along: https://www.youtube.com/watch?v=ccLi41JwkbQ

R Markdown

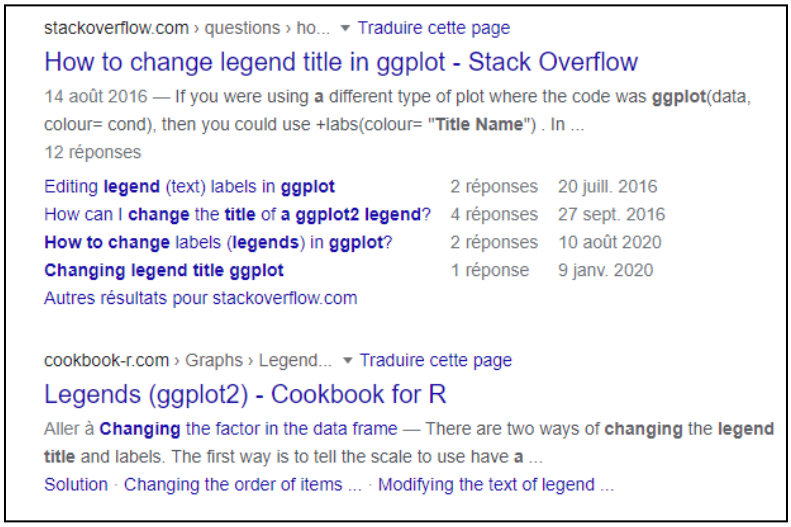
Duration: **7 minutes**

This YouTube video provides a demo of R Markdown:
https://www.youtube.com/watch?v=DNS7i2m4sB0

Troubleshooting

A few ways/resources for troubleshooting R issues:

- Every analyst has had the experience of hitting a wall when trying to code. What to do?
- R requires an almost detective-like attitude: the answer is out there online, you just need to sleuth it out.
- Some ideas:
    - Just plop your question, verbatim, into Google and see what you get. Include the package you're using and "R". Ex. "how to change legend name in ggplot2 in r". Try typing this in google. You should get a few excellent hits:



    - Get comfortable searching **Stack Overflow**, and learning how to apply non-epi solutions to your epi problems.
    - If you are using the tidyverse, there are a TON of resources, including Cookbook for R, R for Data Science, etc.

Tips:

1. Start your search statement with the software name (and version as needed)
2. Follow-up with question, function, or error message
    a. In relation to error messages, they are sometimes long and in such cases it may be most useful to take the more generic portions of the returned message
3. Identify and make note of sources/websites that keep popping up that have useful information, such as:
    - https://stackoverflow.com/

- https://www.rdocumentation.org
- https://www.tidyverse.org/
- https://www.dummies.com/programming/r/r-for-dummies-cheat-sheet/
- http://www.cookbook-r.com/
- https://www.reconlearn.org/

Tutorial on troubleshooting and common errors:
https://ourcodingclub.github.io/tutorials/troubleshooting/

Silly Mistakes We All Make in R/R Studio

Duration: **4 minutes**

This YouTube video reviews some basic programming challenges encountered when running code in R/RStudio, specifically unbalanced parentheses and quotation marks. “When in doubt, escape [Esc] out”: https://www.youtube.com/watch?v=xQ9SJvuzg0A

Common R Errors

This webpage reviews some errors that are commonly encountered when working with R. Some of the content of this page may be inaccessible to novice or beginner R users. If that is the case we recommend (1) a high level review of the webpage, and (2) filing the webpage away for later use as a resource in your R learning journey when these errors are encountered:
https://www.programmingr.com/r-error-messages/

There Be Treasure: Going Further

- The Epidemiologist R Handbook: https://epirhandbook.com/
- R for Data Science: https://r4ds.had.co.nz/
- R Cheat Sheets: https://rstudio.com/resources/cheatsheets/
- YaRrr! The pirate's guide to R: https://bookdown.org/ndphillips/YaRrr/
- R Shiny Tutorials: https://shiny.rstudio.com/tutorial/
- Comprehensive R Archive Network: https://cran.r-project.org/
- REpidemics Consortium: https://www.repidemicsconsortium.org/projects/
- R4Epis: https://r4epis.netlify.app/
- PHAC R user group (note that this is available to PHAC employees only at this time and you will need to create an account or login with an existing account): https://message.gccollab.ca/channel/phac-r-user
- PopDataBC – Intro to R Studio for SAS Users: https://www.popdata.bc.ca/etu/online_courses/STAN104
- PopDataBC - Data Management and Cleaning for Analysis with R: https://www.popdata.bc.ca/etu/online_courses/STAN106
- A comprehensive introduction to handling date and time in R: https://blog.rsquaredacademy.com/handling-date-and-time-in-r/
- Regular tips, tricks, and demonstrations of new packages: https://www.r-bloggers.com/
- Stack overflow: https://stackoverflow.com/
- Coursera - Introduction to Statistics and Data Analysis in R: https://www.coursera.org/specializations/statistical-analysis-r-public-health?recoOrder=1&utm_medium=email&utm_source=recommendations&utm_campaign=btfwMI53EeuiLbV1NbTVAw

(Note: if experiencing issues accessing a link by clicking on it, please try to troubleshoot via copy-paste into your browser.)