

**Introduction to R For Public Health Investigations**

Pre-Course Self-Study Module

Table of Contents

[Copyright statement  2](#_Toc184726834)

[Module description 3](#_Toc184726835)

[Learning objectives 3](#_Toc184726836)

[Overview 4](#_Toc184726837)

[Mandatory sea time: Essential pre-learning 5](#_Toc184726838)

[Swab the decks: Required self-study checklist 6](#_Toc184726839)

[There be treasure: Going further 12](#_Toc184726840)

[Appendix 1: Pre-course checklist 13](#_Toc184726841)

[Appendix 2: Self-assessment 14](#_Toc184726842)

[Pre-course self assessment 14](#_Toc184726843)

[Post-course self assessment 14](#_Toc184726844)

# Copyright statement

Developed by the Training and Development Unit, Regulatory, Operations and Emergency Management Branch, Public Health Agency of Canada

Version: December 2024

**Copyright**

This work is licensed under the Creative Commons Attribution-NonCommerical-ShareAlike 4.0 International License. To view a copy of this license, visit [Creative Commons — Attribution-NonCommercial-ShareAlike 4.0 International — CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).



**Notes**

As per this licence, you may not use the material for commercial purposes such as university courses or other paid events, workshops or conferences.

Our materials may be repurposed for other training (for example, targeted training in your organization) with attribution to the Training and Development Unit at the Public Health Agency of Canada as the original source. If you modify, add to or translate the material, please note it briefly when you attribute the Training and Development Unit as the original source.

If you share, re-use or re-purpose our training materials, please inform us via email at [ceptraining-formationcmu@phac-aspc.gc.ca](mailto:ceptraining-formationcmu@phac-aspc.gc.ca) so that we can track the reach of our training materials.

# Module description

**Introduction to R for Public Health Investigations** is a course designed to help applied public health practitioners and field epidemiologists learn how to use R for their work. The course gradually builds skills over time, focusing on basic R skills for outbreak investigations, public health surveillance, data management, and sharing information. This course is not meant to provide in-depth training in outbreak investigation, public health surveillance, scientific communication, biostatistics, or R in general. Instead, it aims to give epidemiologists and public health workers extra skills and resources for their work in the field.

The **pre-course self-study module** allows participants to review and get familiar with the course requirements at their own pace before the course starts. The module is divided into three sections: essential pre-learning, required self-learning, and optional content. The first two sections cover material related to the virtual classroom training, while the optional content is for those who want to learn more about R. Essential pre-learning items are for all learners to review, with self-assessment questions to help. These items are crucial for participating in the course.

**Required self-learning items** are especially important for those who:

* Have never used R or mainly use spreadsheet software like Excel instead of statistical software like SAS, STATA, or R for analysis.
* Have used R before and want to refresh their knowledge or learn more about specific topics.
* Are proficient in R and looking for a refresher or to learn more about specific topics.

Attached to this document is the **Pre-Course Checklist** for the four-day Introduction to R for Public Health Investigations course. Learners should review and complete each item listed.

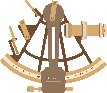
## Learning objectives

By completing the pre-course content, participants will be able to:

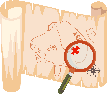
* **Install** R, RStudio, and any packages required for performing desired analyses
* **Discuss** fundamentals of data analysis in R, including efficient project setup and coding basics
* **Identify** available resources for troubleshooting errors in R code and learning more about R packages

# Overview

We have a lot of resources to share and don’t want you to feel overwhelmed. Throughout this document we will provide four different types of resources:

 **Mandatory sea time: Essential pre-learning**

These items are essential to your participation in Introduction to R for Public Health Investigations. If these topics are new to you, please make sure to review the items prior to joining the course.

 **Swab the decks: Required self-study**

Please complete prior to attending the virtual training sessions. We have included self-study exercises in this document. These exercises are provided for your use to assist in retention of content reviewed in self-study.

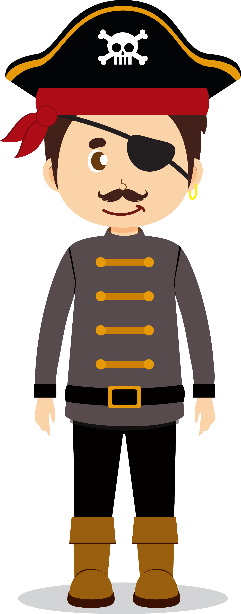
 **There be treasure: Going further**

Additional content provided as reference for self-learning and further development. Please note: this content is optional and not required to complete before participating in the course.

**X Appendix 1: Pre-course checklist**

We have appended a pre-course checklist to the end of this document. This checklist contains information to help you get ready for the course. We look forward to seeing you soon.

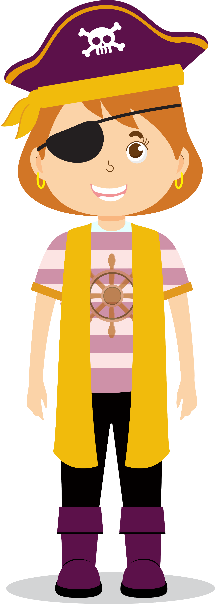
# Mandatory sea time: Essential pre-learning

Before starting an ‘Introduction to using aRrr for public health investigations’ course, you’re obviously going to need to have the proper software installed. Please complete the following tasks to make sure you’re up to date with all the tools you’ll need for the course.

We’ve provided links to resources which will walk you through basic information to get you up and running with R or to revisit if you need a refresher. The essential pre-learning items below will walk you through virtual training and technical requirements for starting Introduction to R for Public Health Investigations. See the descriptions and links below.

|  |  |
| --- | --- |
| Have you: | If not, please review: |
| Installed R and RStudio? | [The Epidemiologist R Handbook: 3.4 Installation](https://epirhandbook.com/en/new_pages/basics.html#installation) |
| Familiarized yourself with the RStudio interface? | [The Epidemiologist R Handbook: 3.5 RStudio](https://epirhandbook.com/en/new_pages/basics.html#rstudio) |
| Learned how to install R Packages? | [The Epidemiologist R Handbook: 3.7 Packages](https://www.epirhandbook.com/en/new_pages/basics.html#packages) |
| Ever imported data into R? | [The Epidemiologist R Handbook: 7 Import and export](https://www.epirhandbook.com/en/new_pages/importing.html) |
| Learned about R scripts? | [The Epidemiologist R Handbook: 3.8 Scripts](https://www.epirhandbook.com/en/new_pages/basics.html#scripts) |
| Learned about basic functions and arguments? | [The Epidemiologist R Handbook: 3.6 Functions](https://epirhandbook.com/en/new_pages/basics.html#functions) |
| Ever set up a working directory in R? | [The Epidemiologist R Handbook: 3.9 Working directory](https://www.epirhandbook.com/en/new_pages/basics.html#working-directory) |
| Encountered the here package? | [The Ultimate Guide to the here Package in R](https://rbasics.org/packages/here-package-in-r/) |
| Learned about tidyverse and what packages it includes? | [Tidyverse](https://www.tidyverse.org/) |
| Learned what an “object” is in R? | [The Epidemiologist R Handbook: 3.10 Objects](https://www.epirhandbook.com/en/new_pages/basics.html#objects) |
| Ever used the pipe operator (%>%) in R? | [The Epidemiologist R Handbook: 8.1 Piping (%>%)](https://www.epirhandbook.com/en/new_pages/basics.html#piping) |
| Learned about key operators and functions, including the %in% operator? | [The Epidemiologist R Handbook: 8.2 Key operators and functions](https://www.epirhandbook.com/en/new_pages/basics.html#in) |
| Learned about the difference between errors and warnings? | [The Epidemiologist R Handbook: 12.1 Errors & warnings](https://www.epirhandbook.com/en/new_pages/basics.html#errors-warnings) |
| Previously encountered cheat sheets for R? | [Posit Cheat sheets](https://rstudio.github.io/cheatsheets/) |

# Swab the decks: Required self-study checklist

This wouldn’t be much of a self-study module without providing some homework for you to study from. Please complete the following videos and tutorials that introduce the concepts that we’ll be applying to public health topics in the course. Review the resource if the topic is one that you are unfamiliar with or if you feel you need a refresher. At the end of this section we have included some self-study exercises. Please complete these exercises to consolidate and further develop your new skills and knowledge acquired through this self-study module, or to self-assess your pre-existing skills and knowledge.

### Data management

[The Epidemiologist R Handbook](https://www.epirhandbook.com/en/new_pages/cleaning.html): 8 Cleaning data and core functions

[The Epidemiologist R Handbook: 9 Working with dates](https://www.epirhandbook.com/en/new_pages/dates.html)

[The Epidemiologist R Handbook: 10 Characters and strings](https://www.epirhandbook.com/en/new_pages/characters_strings.html)

[The Epidemiologist R Handbook: 11 Factors](https://www.epirhandbook.com/en/new_pages/factors.html)

[The Epidemiologist R Handbook: 12 Pivoting data](https://www.epirhandbook.com/en/new_pages/pivoting.html)

[The Epidemiologist R Handbook: 13 Grouping data](https://www.epirhandbook.com/en/new_pages/grouping.html)

[The Epidemiologist R Handbook: 14 Joining data](https://www.epirhandbook.com/en/new_pages/joining_matching.html)

[The Epidemiologist R Handbook: 15 De-duplication](https://www.epirhandbook.com/en/new_pages/deduplication.html)

### Descriptive analysis and data visualization

[The Epidemiologist R Handbook: 17 Descriptive tables](https://www.epirhandbook.com/en/new_pages/tables_descriptive.html)

[The Epidemiologist R Handbook: 20 Missing data](https://www.epirhandbook.com/en/new_pages/missing_data.html)

[The Epidemiologist R Handbook: 29 Tables for presentation](https://www.epirhandbook.com/en/new_pages/tables_presentation.html#basic-flextable)

[The Epidemiologist R Handbook: 30 ggplot basics](https://www.epirhandbook.com/en/new_pages/ggplot_basics.html)

[The Epidemiologist R Handbook: 31 ggplot tips](https://www.epirhandbook.com/en/new_pages/ggplot_tips.html)

[The Epidemiologist R Handbook: 32 Epidemic curves](https://www.epirhandbook.com/en/new_pages/epicurves.html)

### Reporting

[The Epidemiologist R Handbook: 40 Reports with R Markdown](https://www.epirhandbook.com/en/new_pages/rmarkdown.html)

[The Epidemiologist R Handbook: 41 Organizing routine reports](https://www.epirhandbook.com/en/new_pages/reportfactory.html)

### Troubleshooting and errors

[Troubleshooting Tips](#_Knowledge_Boost:_Troubleshooting)

[Tutorial: Troubleshooting](https://ourcodingclub.github.io/tutorials/troubleshooting/)

[Video: Silly mistakes we all make in R/RStudio](https://www.youtube.com/watch?v=xQ9SJvuzg0A)

[Webpage: Common R Errors](https://www.programmingr.com/r-error-messages/)

[The Epidemiologist R Handbook: 47 Common errors](https://www.epirhandbook.com/en/new_pages/errors.html)

[The Epidemiologist R Handbook: 48 Getting help](https://www.epirhandbook.com/en/new_pages/help.html)

#### Knowledge Boost: 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
   1. 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/>

### Self-study exercises

After all these self-learning exercises, it only seems natural that you should test your knowledge at least a little bit. Please complete the following self-reflective and knowledge check questions to ensure you’re ready to start the course.

1. In your experience, how is the R programming language different from other statistical software you’ve used in your work? How is it similar?
2. What is one function for installing packages in R?
3. Using your answer from Question 2, please install the following packages:

here

tidyverse

scales

padr

writexl

fs

RColorBrewer

ggrepel

ggpubr

zoo

viridis

igraph

tidygraph

ggraph

flextable

viridis

incidence

officer

officedown

1. The R package “tidyverse” is unique, in that it is actually a collection of several packages that you will use frequently throughout this course (and hopefully in your career). In the space below, please list the different packages included in the “tidyverse”.

| “Core Packages” | “Non-core Packages” |
| --- | --- |
|  |  |

1. Did you encounter problems installing R packages? If yes, did you understand what the error message was telling you? If not, did you understand what the error message was telling you after you googled it?
2. List two different ways to look up help documentation for R packages and functions:
3. How do statistical software handle dates generally? Why is it useful for dates to be handled in this way?
4. What is R Markdown? Can you see any added value of the functionality provided by R Markdown in your day-to-day work? Why or why not?
5. What does the “gg” stand for in the package “ggplot”? Why do you think it’s called this?
6. What is the difference between long and wide format? Provide an example of when you could use each of these formats.

# **There be treasure: Going further**

A cat and dog with eye patch

The materials provided below are not in any way required for participation in the Introduction to R for Public Health Investigations course. These links are provided merely as a curated list to help you further your R knowledge based on your own interest level. Caution: using the following material may result in becoming an absolute swashbuckler in R! (Note: if experiencing issues accessing a link by clicking on it, please try to troubleshoot via copy-paste into your browser.)

#### R Learning Resources

* The Epidemiologist R Handbook: <https://epirhandbook.com/>
* R for Data Science (2nd Edition): <https://r4ds.hadley.nz/>
* R Shiny Tutorials: <https://shiny.rstudio.com/tutorial/>
* A comprehensive introduction to handling date and time in R: <https://blog.rsquaredacademy.com/handling-date-and-time-in-r/>
* STAT 545: <https://stat545.com/> developed at the University of British Columbia
* The R Graph Gallery – Help and inspiration for R charts: https://r-graph-gallery.com/
* Regular tips, tricks, and demonstrations of new packages: <https://www.r-bloggers.com/>
* R and SNA | Esthercita: <https://ekuki.netlify.app/courses/new_r_and_sna/>
* Reproducible Analytical Pipelines (RAP) Companion: <https://ukgovdatascience.github.io/rap_companion/>

#### R Communities

* Applied Epi - Elevating frontline epidemiology: <https://www.appliedepi.org/>
* RainbowR: <https://rainbowr.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>
* Comprehensive R Archive Network: <https://cran.r-project.org/>
* Regular tips, tricks, and demonstrations of new packages: <https://www.r-bloggers.com/>
* Stack overflow: <https://stackoverflow.com/>

# Appendix 1: Pre-course checklist

Introduction to R for Public Health Investigations

|  |  |
| --- | --- |
| **Actions to be completed prior to joining the virtual classroom on Day 1** | **Completed** |
| Items to complete 1 to 3 weeks before the course |  |
| Complete the knowledge check and associated pre-learning in the Introduction to R for Public Health Investigations Self-Study Module | ☐ |
| Complete the pre-course self-assessment appended to the Introduction to R for Public Health Investigations Self-Study Module | ☐ |
| Items to complete 3 to 6 days before the course |  |
| Advise the TDU if you are unable to access the Participant Guide located on the file share platform for this course. Note that the link will follow in an email and course content will be provided one week prior to the course. | ☐ |
| Ensure that R and R Studio are installed on your computer. | ☐ |
| Test virtual classroom (Zoom) audio and video on the computer you will be using for the duration of the course: <https://support.zoom.us/hc/en-us/articles/115002262083-Joining-a-test-meeting> | ☐ |
| Advise supervisor and coworkers that you will be unavailable for the duration of the course. | ☐ |
| Items to complete 1 to 3 days prior the course |  |
| Ensure that you have reviewed pre-learning for Day 1 (refer to the Introduction to R for Public Health Investigations Participant Guide). | ☐ |
| Download materials for the practical exercise on Day 1 and advise the TDU of any issues. | ☐ |
| Advise your supervisor and coworkers that you will be unavailable for the duration of each virtual classroom session (incl. ~15 minutes before and after). | ☐ |
| Locate the Zoom meeting details for the virtual classroom and office hours, and ensure that they will be easy to find during the course. | ☐ |
| OPTIONAL: Print course materials for note taking as needed. | ☐ |
| Steps to complete 15 minutes prior to joining each virtual session |  |
| Turn off all telephone notifications, and set wireless devices to silent. | ☐ |
| Prepare your computer by closing all non-required applications (incl. email and instant messaging). | ☐ |
| Access the virtual classroom 15 minutes prior to each session to resolve potential issues and get ready on time. | ☐ |

Please contact the Training and Development Unit at [ceptraining-formationcmu@phac-aspc.gc.ca](mailto:phac.ceprtraining-formationcmiu.aspc@phac-aspc.gc.ca) for assistance.

# Appendix 2: Self-assessment

This self-assessment is intended to be a self-reflective exercise for you to complete ahead of and following the course. This activity is optional. However, we recommend that field epidemiologists in the Canadian Field Epidemiology Program complete this activity and use it to inform discussions with their program in relation to remaining individual gaps and needs as they relate to technical skill proficiency.

## Pre-course self assessment

Before joining the Introduction to R for Public Health Investigations course, how confident do you feel developing R code by yourself for the following tasks (1= not confident; 10 = very confident)?

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Skill | Not confident | | | | | Very confident | | | | |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Setting a working directory |  |  |  |  |  |  |  |  |  |  |
| Reshaping data |  |  |  |  |  |  |  |  |  |  |
| Creating new variables |  |  |  |  |  |  |  |  |  |  |
| Identifying missing data |  |  |  |  |  |  |  |  |  |  |
| Descriptive epidemiology |  |  |  |  |  |  |  |  |  |  |
| Creating an epidemic curve |  |  |  |  |  |  |  |  |  |  |
| Importing a dataset |  |  |  |  |  |  |  |  |  |  |
| Merging and appending data |  |  |  |  |  |  |  |  |  |  |
| Automating a report |  |  |  |  |  |  |  |  |  |  |

## Post-course self assessment

After completing the Introduction to R for Public Health Investigations course, how confident do you feel developing R code by yourself for the following tasks (1= not confident; 10 = very confident)?

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Skill | Not confident | | | | | Very confident | | | | |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Setting a working directory |  |  |  |  |  |  |  |  |  |  |
| Reshaping data |  |  |  |  |  |  |  |  |  |  |
| Creating new variables |  |  |  |  |  |  |  |  |  |  |
| Identifying missing data |  |  |  |  |  |  |  |  |  |  |
| Descriptive epidemiology |  |  |  |  |  |  |  |  |  |  |
| Creating an epidemic curve |  |  |  |  |  |  |  |  |  |  |
| Importing a dataset |  |  |  |  |  |  |  |  |  |  |
| Merging and appending data |  |  |  |  |  |  |  |  |  |  |
| Automating a report |  |  |  |  |  |  |  |  |  |  |