

TDU.

Introduction to R For Public Health Investigations

Participant Guide



Public Health
Agency of Canada

Agence de la santé
publique du Canada

Canada

| | |
|---|----|
| Course Description | 2 |
| How the Course Works..... | 3 |
| Course Materials | 5 |
| Day 1 | 6 |
| Schedule | 6 |
| Prior to Joining on Day 1: Pre-Course Work | 7 |
| After Joining on Day 1: Practical Exercise(s) and Application | 10 |
| Day 2 | 13 |
| Schedule | 13 |
| Prior to Joining on Day 2: Pre-Course Work | 14 |
| After Joining on Day 2: Practical Exercise(s) and Application | 16 |
| Day 3 | 19 |
| Schedule | 19 |
| Prior to Joining on Day 3: Pre-Course Work | 20 |
| After Joining on Day 3: Practical Exercise(s) and Application | 21 |
| Day 4 | 23 |
| Schedule | 23 |
| Details | 23 |

COURSE DESCRIPTION

DESCRIPTION

Are you an epidemiologist who needs to improve or refresh your R skills? Or maybe you are an R enthusiast who started your day excited about the cool and seemingly endless ways you might use R today? Or perhaps you tried to use R a long time ago and have finally recovered from that difficult experience enough to want to try again? If so, we have a course for you!

Introduction to R for Public Health Investigations is a four-day course where our goal is to provide a relaxed and supportive environment for learning, practice, and discussion that is welcoming and engaging for both novice and intermediate R users alike.

Join us as we demystify the intricacies of R, become inspired by R's usefulness in efficient management of data projects and reporting processes, and find new ways to apply R to our day-to-day work in public health!

PREREQUISITES

Participants should have a background in epidemiology and have previously completed training in public health investigations and surveillance (or equivalent). Experience in R is not required; however, participants should have experience using a statistical or analytical software (i.e., SAS, SPSS, STATA – at minimum have a high proficiency using an analytical spreadsheet software like Excel). Participants with no prior experience in R or other statistical software may still participate in the course if they require the training for their role in public health response. Resources including learning pathways for beginner, novice and intermediate level R users will be provided as part of a pre-course self-directed study module several weeks ahead of the course to accommodate different learning needs.

LEARNING OBJECTIVES

By the end of this course, participants will be able to:

- Carry out data cleaning and processing, and descriptive epidemiological analyses (incl. commonly used data visualisations such as epidemic curves and social network diagrams);
- Create automated data products (e.g., epidemiologic summaries);
- Design and carry out a data collation plan that is consistent with proposed analysis plan (i.e., merging and appending data);
- Explain when it is most appropriate to program analyses and automate tasks using R; and
- Find possible solutions to R programming challenges.

Acknowledgements

This session relies on materials developed in collaboration with subject matter experts Emma Cumming (Canadian Public Health Service), Chris Mill (Canadian Public Health Service), and Najmus Saqib (Office of the Chief Data Officer). The Training and Development Unit is grateful for their significant contributions.

HOW THE COURSE WORKS

This course relies heavily on self-study activities: pre-learning, practical exercises, and the discussion board (Slack). The virtual classroom provides an opportunity to review, apply, and discuss self-study activities in more depth and debrief on after class activities. Completing self-study activities prior to joining the virtual session each day is critical to gaining the most from this learning experience. Support is available to learners outside the virtual classroom to ask questions or discuss the materials in more detail.

How do I know if I should make use of support outside the virtual classroom?

Though this component is not mandatory, the Intro to R is highly participatory and relies heavily on self-study. Reach out to your course facilitators on Slack when:

- *You have lingering or new questions regarding the material covered in videos or the virtual classroom; and/or*
 - *You are struggling with the practical exercise or are unclear as to what you are supposed to do.*
-

How the Virtual Classroom Works

Virtual classroom time takes place from 11:30 am for 3-4 hours each day. The classroom will open at 11:15 am ET and close at 3:30pm ET. Please join 15 minutes prior to the start of the virtual class. Details to join the virtual classroom can be found in the course communications sent to your email, or embedded in the Outlook calendar invite.

Note: Please complete the pre-course checklist provided in the pre-course self-study module prior to joining the virtual classroom on Day 1.

How the PHAC R Users Slack Workspace Works

In addition to the virtual classroom, learners are invited to join the Slack workspace used for this training (including facilitators, subject matter experts, and learners from other editions of this training). Within this workspace is a private channel for the course – here learners will post responses to discussion questions, or have additional opportunity to engage with facilitators,

subject matter experts, and peers. This private channel and its contents are visible only to facilitators, your peers in the course, and subject matter experts invited to participate in the session.

Note: If you are not able to access the PHAC R Users Slack workspace and/or are not able to see the private channel associated with your training (e.g., denoted by a lock symbol and date of when you're taking the course) by end of day on the Friday prior to the first day of the course, please email ceptraining-formationcmu@phac-aspc.gc.ca

How Support Outside of the Virtual Classroom Works

This is optional. Reach out to your course facilitators and subject matter experts by posting your question to the Slack channel associated with this training event and someone will respond to your question as soon as possible. If more intensive or one-on-one support is needed, please directly reach out to your course facilitators. Note, on day 4 of the course there will be a block of time set aside for practice and troubleshooting the practical exercises associated with days 1-3 of the course.

COURSE MATERIALS

Where a weblink is not provided in this document, all course materials and resources are located in GitHub. All materials for required self-learning will be posted one week prior to the start of course. Please email ceptraining-formationcmu@phac-aspc.gc.ca at your earliest convenience should you have any troubles accessing the course materials via the link below.

TDU GitHub repository: <https://github.com/hc-sc/tdu-intror>

Instructions to download materials from GitHub: https://github.com/hc-sc/tdu-intror/blob/main/Instructions_Download_Course_Materials.pdf

DAY 1

SCHEDULE

| Session | Activity | Description |
|--|---------------------------|---|
| Pre-Learning (set aside 30 minutes) | Foundational Concepts | 10 Considerations for Data Analysis (Handout) |
| | Foundational Concepts | Day1_Ben Introduces R.mp4 (Video) |
| | Foundational Concepts | Day1_Joanne walks us through the basics.mp4 (Video) |
| Virtual Classroom | Welcome and Introductions | Introduction to the Course |
| | Foundational Concepts | 10 Considerations for Data Analysis |
| | Break | Heave Away |
| | Foundational Concepts | R Syntax and Data Storage |
| | Activity | Let's Review – Cross Word Puzzle |
| | Break | Trivia: Famous Shipwrecks |
| | Foundational Concepts | R Syntax, Indexing, and Considerations for Programming |
| | Demo | Functions to Explore Your Data |
| | Break | Trivia: Public Health at Sea |
| | Demo | Here Package |
| | Wrap-Up | Wrap-Up Day 1 |
| | Demo | Getting Started with Exercise 1 |
| Practical Exercise(s) and Application | Self-Study Exercise | Exercise 1 |
| Support on Slack | Optional | For learners to engage more deeply with materials or ask questions relating to the self-study exercise (optional) |

PRIOR TO JOINING ON DAY 1: PRE-COURSE WORK

Pre-Learning

These resources are required pre-learning for Day 1 of the Introduction to R for Public Health Investigations course. All materials (unless a weblink has been provided in this guide) are located in your course materials from GitHub in the “Pre-Learning” folder.

10 Considerations for Data Analysis

This handout provides an overview of the 10 considerations for data analysis that we will revisit throughout the training. While it is best to use your judgement to decide how much time you will need to review this resource, we estimate that you will need **5-10 minutes**.

Link - English:

https://drive.google.com/file/d/1mTV52q3PZLMEVcbQdV22QgogTAHa51uS/view?usp=share_link

Link – French:

https://drive.google.com/file/d/1e8PUpkRzbicmS4WoDyU75lrDI8lpt8Gv/view?usp=share_link
[SJ(1)]

Introducing R

These videos provide an overview of why R is a useful software for epidemiologists to work with, as well as an overview of the RStudio interface.

- Day1_Ben Introduces R.mp4 (**Duration: 10 mins**)
Link: https://drive.google.com/file/d/1FTWh9t23QAv4-9BlpzbURuepQ309c0Wi/view?usp=share_link
- Day1_Joanne walks us through the basics.mp4 (**Duration 10 mins**)
Link: https://drive.google.com/file/d/1-iq xv11-Ay17GEvtZKYdmS4nkhDbrUR6/view?usp=share_link

Intro to R: Pre-Course Self-Study Module

These resources were shared as part of the Intro to R Pre-Course Self-Study Module in advance of this course. The materials review should have been completed prior to joining. Should you need a refresher, here is a list of learning resources pertinent to Day 1 of the Introduction to R for Public Health Investigations course.



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:

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=CNo4rlg4nU8>

Technical Check: 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 aid them in installing R and RStudio on their computer themselves.

<https://techvidvan.com/tutorials/install-r/#install-r-windows>

Technical Check: 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>

Technical Check: 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/>

Technical Check: 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>

Tidy data and `tidyr`

Duration: **18 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 second in a series of four, reviewing concepts such as tidy data and the `tidyr` package. <https://youtu.be/1ELALQIO-yM>

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 the pre-course self study module!

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>

AFTER JOINING ON DAY 1: PRACTICAL EXERCISE(S) AND APPLICATION

For after joining the virtual classroom.

Optional Resource

Duration: **15 mins**

A recording of an earlier version of the here package demonstration:
https://drive.google.com/file/d/1RB0Rb1Pk3w4kEUv-IU9-GWF-v3Br/Viu/view?usp=share_link

Exercise 1

Note: there is dedicated course time for practice with exercises 1, 2, or 3 (your choice) on day 4 of the course. Course facilitators will be available on Slack and Zoom for your questions.

Duration: **180 mins**

Get as far as you can with this exercise within 2 hours and 10 minutes (maximum). Don't worry if you need extra time. The learning curve for R is steep and learners will benefit most from time spent practicing. We will be scheduling a follow-up webinar to debrief on the exercises several weeks out from the end of this course (date to be announced). Prior to that debrief, reach out to your facilitators on Slack or by email if you require assistance with the course material.

We recommend:

- Novice users (Boatswains): Instead of writing out the code from the images provided, follow along the workbook with the R script for the exercise. Prioritise running the code piece by piece to understand what a chunk of code does, and what specific functions are doing rather than being able to write code and troubleshoot.
- Beginner/Intermediate users (First Mates): R code is provided as a screen capture image in the workbook. You should have sufficient understanding of coding to get a general sense of what the code is doing by reading it (with the assistance of the help documentation and a few Google searches as needed). It is our intention to have you write the code out from the workbook as you progress through the scenario.

- Advanced users (Master Mariners): We encourage you to try writing your own code where you like and contrast it with the code used for the exercise, and to help your peers as questions arise.

Scenario

It is the beginning of December, 2020. You are placed in the Office of the Provincial Health Officer at the British Columbia Ministry of Health. You and your public health colleagues are busy responding to the COVID-19 pandemic and developing an understanding of the occurrence of COVID-19 in the province. One of the items requested by the Provincial Health Officer for her situational awareness is a regularly updated analysis of COVID-19 in other Canadian jurisdictions. It is your job to produce the figures and share them with the team compiling information for the Provincial Health Officer.

Materials

All materials are located on GitHub - <https://github.com/hc-sc/tdu-intror/tree/main/Day1/PracticalExercise>.

For this exercise you will need:

- An internet connection
- Intro to R Day 1 Workbook
- covid19.csv
- 01 - define_paths_Final.R
- 02 - load_libraries_Final.R
- 03 - import_data_Final.R
- 04.1 - clean_population_denominator_Final.R
- 04.2 - clean_phac_case_data_Final.R
- 05.1 - plot_epi_curve_Final.R

Exercise 1: What learners will do (conceptual)

- In the context of the scenario (CMOH wants a repeat analysis from PHAC open access COVID-19 linelist):
 - Manipulating dates
 - Reshaping data
 - Creating new variables
 - Creating the plot
 - Automating the results

Exercise 1: What learners will do (technical)

- Prepping to work in R
 - Load installed packages
 - Import data
 - Open scripts

- Cleaning data
- Creating new variable
- Summarising data
- Plotting data (epidemic curve)
- Automating the analysis

DAY 2

SCHEDULE

| Session | Activity | Description |
|--|-----------------------|---|
| Pre-Learning (set aside 30 minutes) | Context | TB 101 (Handout) |
| | Foundational Concepts | Flextable (Webpage) |
| | Foundational Concepts | Tidygraph (Webpage) |
| Virtual Classroom | Debrief | Debrief Day 1/Exercise 1 |
| | Foundational Concepts | Reporting and Visuals |
| | Break | Trivia: Know your knots! |
| | Foundational Concepts | Tables and Sourcing Code |
| | Activity | Sourcing Code |
| | Break | Paper Flotilla |
| | Foundational Concepts | Troubleshooting |
| | Activity | Decode the Message |
| | Demo | CanMaps |
| | Break | Beach Party! |
| | Wrap-Up | Wrap-Up Day 2 |
| | Demo | Getting Started With Exercise 2 |
| Practical Exercise(s) and Application | Self-Study Exercise | Exercise 2 |
| Support on Slack | Optional | For learners to engage more deeply with materials or ask questions relating to the self-study exercise (optional) |

PRIOR TO JOINING ON DAY 2: PRE-COURSE WORK

Pre-Learning

These resources are required pre-learning for Day 2 of the Introduction to R for Public Health Investigations course. All materials (unless a weblink has been provided in this guide) are located in your course materials from GitHub in the “Pre-Learning” folder.

Context: Tuberculosis

This paper provides a primer on tuberculosis (TB): what it is, how it is diagnosed, and some high level points relevant for transmission and contact tracing. While it is best to use your judgement to decide how much time you will need to review this resource, we estimate that you will need **10 minutes**.

Link: <http://www.letstalktb.org/wp-content/uploads/2015/11/nature-tb-review-2016.pdf>

Tidygraph

This webpage reviews network analysis using tidygraph and ggraph packages. While it is best to use your judgement to decide how much time you will need to review this resource, we estimate that you will need **10 minutes**.

Link: https://www.shirin-glander.de/2018/03/got_network/

Flextable

This webpage provides an overview of how to use the flextable package to create tables for reporting summarised data. While it is best to use your judgement to decide how much time you will need to review this resource, we estimate that you will need **10 minutes**.

Link: <https://davidgohel.github.io/flextable/>

Intro to R: Pre-Course Self-Study Module

These resources were shared as part of the Pre-Course Self-Study Module in advance of the course. The materials review should have been completed prior to joining the course. Should you need a refresher, here is a list of learning resources pertinent to Day 2 of the Introduction to R for Public Health Investigations course



Swab The Decks: Foundational Concepts

Data Manipulation Tools and Dplyr

Duration: **20 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 third in a series of four, reviewing concepts such as the dplyr package, key functions for data manipulation, and how to chain statements together with the pipe operator. https://www.youtube.com/watch?v=Zc_ufg4uW4U

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:

stackoverflow.com > questions > ho... ▼ Traduire cette page

How to change legend title in ggplot - Stack Overflow

14 août 2016 — If you were using a different type of plot where the code was `ggplot(data, colour= cond)`, then you could use `+labs(colour= "Title Name")`. In ...

12 réponses

| | | |
|--|------------|----------------|
| Editing legend (text) labels in ggplot | 2 réponses | 20 juill. 2016 |
| How can I change the title of a ggplot2 legend? | 4 réponses | 27 sept. 2016 |
| How to change labels (legends) in ggplot ? | 2 réponses | 10 août 2020 |
| Changing legend title ggplot | 1 réponse | 9 janv. 2020 |

Autres résultats pour stackoverflow.com

cookbook-r.com > Graphs > Legend... ▼ Traduire cette page

Legends (ggplot2) - Cookbook for R

Aller à **Changing** the factor in the data frame — There are two ways of **changing** the **legend title** and labels. The first way is to tell the scale to use have a ...

[Solution · Changing the order of items ...](#) · [Modifying the text of legend ...](#)

- 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/>

AFTER JOINING ON DAY 2: PRACTICAL EXERCISE(S) AND APPLICATION

For after joining the virtual classroom.

Optional Resource

As referred to during the presentation on Day 2, this is a handy document to refer back to when working with R Markdown: [Link](#)^[SJ(2)]

Exercise 2

Note: there is dedicated course time for practice with exercises 1, 2, or 3 (your choice) on day 4 of the course. Course facilitators will be available on Slack and Zoom for your questions.

Duration: 210 mins

Get as far as you can with this exercise within 3 hours (maximum). Don't worry if you need extra time. The learning curve for R is steep and learners will benefit most from time spent practicing.

We will be scheduling a follow-up webinar to debrief on the exercises several weeks out from the end of this course (date to be announced). Prior to that debrief, reach out to your facilitators on Slack or by email if you require assistance with the course material.

We recommend:

- Novice users (Boatswains): Instead of writing out the code from the images provided, follow along the workbook with the R script for the exercise. Prioritise running the code piece by piece to understand what a chunk of code does, and what specific functions are doing rather than being able to write code and troubleshoot.
- Beginner/Intermediate users (First Mates): R code is provided as a screen capture image in the workbook. You should have sufficient understanding of coding to get a general sense of what the code is doing by reading it (with the assistance of the help documentation and a few Google searches as needed). It is our intention to have you write the code out from the workbook as you progress through the scenario.
- Advanced users (Master Mariners): We encourage you to try writing your own code where you like and contrast it with the code used for the exercise, and to help your peers as questions arise.

Scenario

You've been sent on a mobilization to a TB outbreak, affecting members of a First Nation both on and off-reserve. The affected reserve (population 2000) is in a remote northern area, with the nearest town being 25 km away (mixed settler and First Nations population, population 7500). TB cases off-reserve are all linked to a rooming house in the nearby town. You've been provided with data by the local health authorities, and need to clean it for analysis. The site wishes you to record all your code and steps in an R Markdown file, so they can repeat the analysis once you leave, if necessary. Your R Markdown report will be "rendered" or exported into a Word document.

Materials

All materials are located on Github: <https://github.com/hc-sc/tdu-intror/tree/main/Day2/PracticalExercise>

For this exercise you will need:

- An internet connection
- Intro to R Day 2 Workbook
- tb_cases.xlsx
- tb_contacts.xlsx
- 01.1 - define_paths_final.R
- 01.2 - load_libraries_final.R
- 01.3 - load_data_final.R
- 02.1 - clean_data_final.R

- 03.1 - plot_case_time_final.R
- 03.2 - plot_contact_demographics_final.R
- 03.3 - tab_case_site_final.R
- 04.1 - plot_sna_final.R
- 05.1 - plot_epicurve_final.R
- 05.2 - tab_location_final.R
- Day2_final.Rmd

DAY 3

SCHEDULE

| Session | Activity | Description |
|--|-----------------------|---|
| Pre-Learning (set aside 10 minutes) | Context | Canadian Chronic Disease Indicators (Webpage) |
| Virtual Classroom | Debrief | Debrief Day 2/Exercise 2 |
| | Foundational Concepts | Appending Data |
| | Break | Trivia: Sailing the Seven Seas |
| | Activity | Navigating Tricky Waters |
| | Foundational Concepts | Data Joins |
| | Break | Des(s)ert Island |
| | Activity | Join Types |
| | Demo | Markdown and HTML Notebooks |
| | Break | Stretch Break |
| | Wrap-Up | Wrap-Up Day 3 |
| | Demo | Getting Started with Exercise 3 |
| Practical Exercise(s) and Application | Self-Study Exercise | Exercise 3 |
| Support on Slack | Optional | For learners to engage more deeply with materials or ask questions relating to the self-study exercise (optional) |

PRIOR TO JOINING ON DAY 3: PRE-COURSE WORK

Course Pre-Learning

These resources are required pre-learning for Day 3 of the Introduction to R for Public Health Investigations course. All materials (unless a weblink has been provided in this guide) are located on Github: <https://github.com/hc-sc/tdu-intror/tree/main/PreLearning>

Context: Canadian Chronic Disease Indicators

The Canadian Chronic Disease Indicators is a resource regarding the burden of chronic disease and associated determinants. Drawing on sources of data that include but are not limited to the Canadian Community Health Survey and Canadian Health Measures Survey, Canadian Chronic Diseases Surveillance System, Discharge Abstract Database, Labour Force Survey, statistics for each indicator are presented along with breakdowns by various determinants. While it is best to use your judgement to decide how much time you will need to review this resource, we estimate that you will need **10 minutes**.

Link: <https://health-infobase.canada.ca/ccdi/>

Intro to R: Pre-Course Self-Study Module

Link: <https://bit.ly/3EC4bK3>

These resources were shared as part of the Pre-Course Self-Study Module in advance of the course. The self-assessment and materials review should have been completed prior to joining the course. Should you need a refresher, here is a list of learning resources pertinent to Day 3 of the Introduction to R for Public Health Investigations course



Swab The Decks: Foundational Concepts

Data Manipulation Tools and Dplyr

Duration: **20 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 third in a series of four, reviewing concepts such as the `dplyr` package, key functions for data manipulation, and how to chain statements together with the pipe operator. https://www.youtube.com/watch?v=Zc_ufg4uW4U

R Markdown

Duration: **7 minutes**

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

AFTER JOINING ON DAY 3: PRACTICAL EXERCISE(S) AND APPLICATION

For after joining the virtual classroom.

Exercise 3

Note: there is dedicated course time for practice with exercises 1, 2, or 3 (your choice) on day 4 of the course. Course facilitators will be available on Slack and Zoom for your questions.

Duration: 210 mins

Get as far as you can with this exercise within 3 hours (maximum). Don't worry if you need extra time. The learning curve for R is steep and learners will benefit most from time spent practicing. We will be scheduling a follow-up webinar to debrief on the exercises several weeks out from the end of this course (date to be announced). Prior to that debrief, reach out to your facilitators on Slack or by email if you require assistance with the course material.

We recommend:

- Novice users (Boatswains): Instead of writing out the code from the images provided, follow along the workbook with the R script for the exercise. Prioritise running the code piece by piece to understand what a chunk of code does, and what specific functions are doing rather than being able to write code and troubleshoot.
- Beginner/Intermediate users (First Mates): R code is provided as a screen capture image in the workbook. You should have sufficient understanding of coding to get a general sense of what the code is doing by reading it (with the assistance of the help documentation and a few Google searches as needed). It is our intention to have you write the code out from the workbook as you progress through the scenario.
- Advanced users (Master Mariners): We encourage you to try writing your own code where you like and contrast it with the code used for the exercise, and to help your peers as questions arise.

Scenario

In this session you will apply your R and data management skills to prepare and collate data from multiple datasets, apply a case identification algorithm, and conduct a basic descriptive epidemiological analysis. You have been provided with five datasets which are mock extracts from administrative datasets (physician billings, discharges from hospital, and provincial health insurance client roster) and have been simplified for training purposes. These datasets must be combined in order to efficiently apply administrative case definitions (algorithms) so that you can describe the occurrence of health service use for chronic conditions among residents of a small province.

Materials

All materials are located on Github: <https://github.com/hc-sc/tdu-intror/tree/main/Day3/PracticalExercise>

For this exercise you will need:

- An internet connection
- Intro to R Day 3 Workbook
- Dad1213.csv
- Dad1314.csv
- Msp1213.csv
- Msp1314.csv
- Reg1314.csv
- Day3_Exercise_script_commented.R
- Day3_Exercise_script_commented.Rmd

DAY 4

SCHEDULE

| Session | Activity | Description |
|-------------------|-----------------------------|------------------------------------|
| Virtual Classroom | Context | Welcome to Day 4 |
| | Context | Course recap |
| | Activity | Jeopardy |
| | Break | |
| | Inside the Analyst's Studio | Panel discussion |
| | Supported Self-Study | Practice Time: Exercise 1, 2, or 3 |
| | Pearls of Wisdom | Reflection activity |
| | Wrap-Up | Course Wrap-Up |

DETAILS

Note: there is no pre-learning for day 4. Day 4 will prioritise time for practice with exercises 1, 2, or 3 (your choice). Course facilitators will be available in Zoom and Slack to answer your questions. Participants are welcome to drop off of Zoom during this time, or stay and we can create small groups for individuals who would like to work together. **Please return to the virtual classroom by 3:10 PM ET.**