

R 4 Epidemiology

2024-05-21

Table of contents

Welcome	3
Acknowledgements	3
Introduction	4
Goals	4
Text conventions used in this book	5
Other reading	5
Contributing	6
Typos	6
Issues	13
License Information	13
About the Authors	14
Brad Cannell	14
Melvin Livingston	15
I Coding Tools and Best Practices	16
1 Quarto Files	17
II Collaboration	18
2 Using git and GitHub	19
III References	20
References	21
Appendices	22
A Glossary	22

Welcome

Welcome to R for Epidemiology!

This electronic textbook was originally created to accompany the Introduction to R Programming for Epidemiologic Research course at the [University of Texas Health Science Center School of Public Health](#). However, we hope it will be useful to anyone who is interested in R, epidemiology, or human health and well-being.

Acknowledgements

This book is currently a work in progress (and probably always will be); however, there are already many people who have played an important role (some unknowingly) in helping develop it thus far. First, we'd like to offer our gratitude to all past, current, and future members of the R Core Team for maintaining this *amazing, free* software. We'd also like to express our gratitude to everyone at [Posit](#). You are also developing and *giving away* some amazing software. In particular, we'd like to acknowledge [Garrett Grolemund](#) and [Hadley Wickham](#). Both have had a huge impact on how we use and teach R. We'd also like to thank our students for all the feedback they've given us while taking our courses. In particular, we want to thank [Jared Wiegand](#) and Yiqun Wang for their many edits and suggestions.

This electronic textbook was created and published using [R](#), [RStudio](#), the [Quarto](#), and [GitHub](#).

Introduction

Goals

We’re going to start the introduction by writing down some basic goals that underlie the construction and content of this book. We’re writing this for you, the reader, but also to hold ourselves accountable as we write. So, feel free to read if you are interested or skip ahead if you aren’t.

The goals of this book are:

1. **To teach you how to use R and RStudio as tools for applied epidemiology.**¹
Our goal is not to teach you to be a computer scientist or an advanced R programmer. Therefore, some readers who are experienced programmers may catch some technical inaccuracies regarding what we consider to be the fine points of what R is doing “under the hood.”
2. **To make this writing as accessible and practically useful as possible without stripping out all of the complexity that makes doing epidemiology in real life a challenge.** In other words, We’re going to try to give you all the tools you need to *do* epidemiology in “real world” conditions (as opposed to ideal conditions) without providing a whole bunch of extraneous (often theoretical) stuff that detracts from *doing*. Having said that, we will strive to add links to the other (often theoretical) stuff for readers who are interested.
3. **To teach you to accomplish common *tasks*,** rather than teach you to use functions or families of functions. In many R courses and texts, there is a focus on learning all the things a function, or set of related functions, can do. It’s then up to you, the reader, to sift through all of these capabilities and decided which, if any, of the things that *can* be done will accomplish the tasks that you are *actually trying* to accomplish. Instead, we will strive to start with the end in mind. What is the task we are actually trying to accomplish? What are some functions/methods we could use to accomplish that task? What are the strengths and limitations of each?

¹In this case, “tools for applied epidemiology” means (1) understanding epidemiologic concepts; and (2) completing and interpreting epidemiologic analyses.

4. **To start each concept by showing you the end result** and then deconstruct how we arrived at that result, where possible. We find that it is easier for many people to understand new concepts when learning them as a component of a final product.
5. **To learn concepts with data** instead of (or alongside) mathematical formulas and text descriptions, where possible. We find that it is easier for many people to understand new concepts by seeing them in action.

Text conventions used in this book

- We will hyperlink many keywords or phrases to their [glossary](#) entry.
- Additionally, we may use **bold** face for a word or phrase that we want to call attention to, but it is not necessarily a keyword or phrase that we want to define in the glossary.
- **Highlighted inline code** is used to emphasize small sections of R code and program elements such as variable or function names.

Other reading

If you are interested in R4Epi, you may also be interested in:

- [Hands-on Programming with R](#) by Garrett Golemund. This book is designed to provide a friendly introduction to the R language.
- [R for Data Science](#) by Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund. This book is designed to teach readers how to do data science with R.
- [Statistical Inference via Data Science: A ModernDive into R and the Tidyverse](#). This book is designed to be a gentle introduction to the practice of analyzing data and answering questions using data the way data scientists, statisticians, data journalists, and other researchers would.
- [Reproducible Research with R and RStudio](#) by Christopher Gandrud. This book gives you tools for data gathering, analysis, and presentation of results so that you can create dynamic and highly reproducible research.
- [Advanced R](#) by Hadley Wickham. This book is designed primarily for R users who want to improve their programming skills and understanding of the language.

Contributing

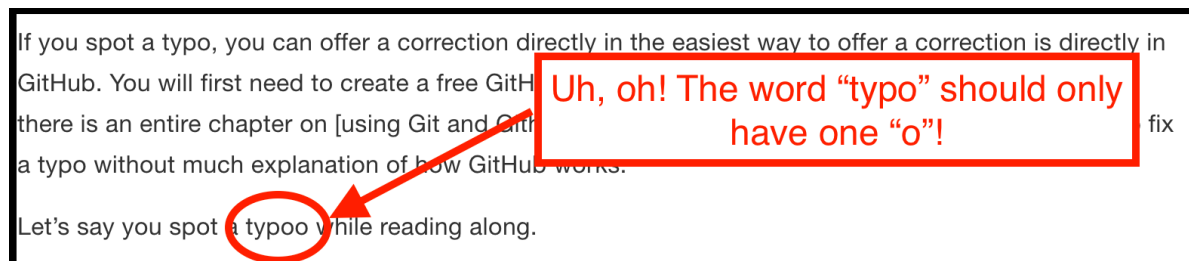
Over the years, we have learned so much from our students and colleagues, and we anticipate that there is much more we can learn from you – our readers. Therefore, we welcome and appreciate all constructive contributions to R4Epi!

Typos

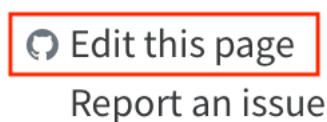
The easiest way for you to contribute is to help us clean up the little typos and grammatical errors that inevitably sneak into the text.

If you spot a typo, you can offer a correction directly in GitHub. You will first need to create a free GitHub account: [sign-up at github.com](https://github.com). Later in the book, we will cover using GitHub in greater depth (See Chapter 2). Here, we're just going to walk you through how to fix a typo without much explanation of how GitHub works.

Let's say you spot a typo while reading along.



Next, click the edit button in the toolbar as shown in the screenshot below.

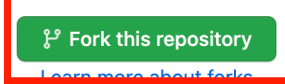


The first time you click the icon, you will be taken to the R4Epi repository on GitHub and asked to fork it. For our purposes, you can think of a GitHub repository as being similar to a shared folder on Dropbox or Google Drive.



You need to fork this repository to propose changes.

Sorry, you're not able to edit this repository directly — you need to fork it and propose your changes from there instead.



Fork the
Repository

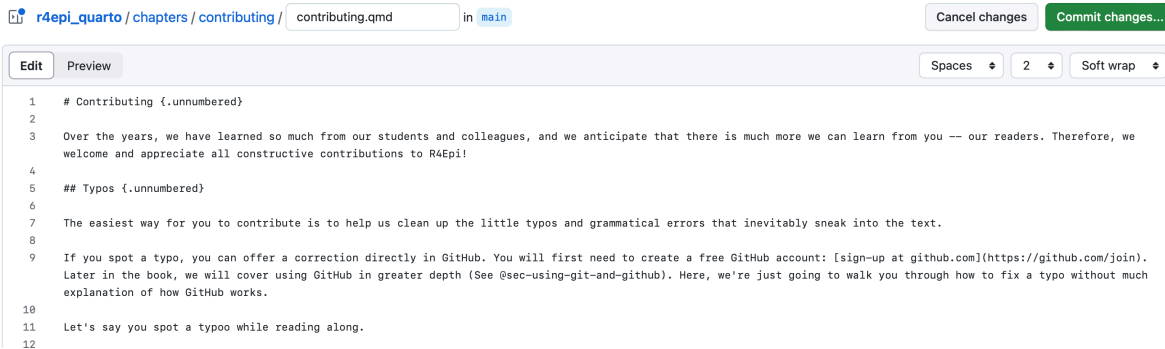
“Forking the repository” basically just means “make a copy of the repository” on your GitHub account. In other words, copy all of the files that make up the R4Epi textbook to your GitHub account. Then, you can fix the typos you found in your *copy* of the files that make up the book instead of directly editing the *actual* files that make up the book. This is a safeguard to prevent people from accidentally making changes that shouldn’t be made.

Note

Forking the R4Epi repository does not cost any money or add any files to your computer.

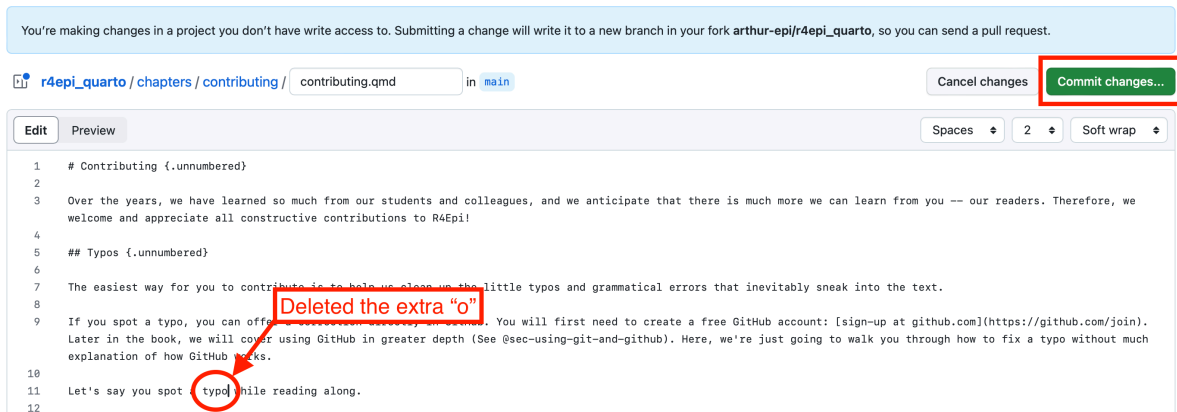
After you fork the repository, you will see a text editor on your screen.

You're making changes in a project you don't have write access to. Submitting a change will write it to a new branch in your fork `arthur-epi/r4epi_quarto`, so you can send a pull request.



The text editor will display the contents of the file used to make the chapter you were looking at when you clicked the `edit` button. In this example, it was a file named `contributing.qmd`. The `.qmd` file extension means that the file is a Quarto file. We will learn more about Quarto files in Chapter 1, but for now just know that Quarto files can be used to create web pages and other documents that contain a mix of R code, text, and images.

Next, scroll down through the text until you find the typo and fix it. In this case, line 11 contains the word “typoo”. To fix it, you just need to click in the editor window and begin typing. In this case, you would click next to the word “typoo” and delete the second “o”.



Now, the only thing left to do is propose your typo fix to the authors. To do so, click the green **Commit changes...** button on the right side of the screen above the text editor (surrounded with a red box in the screenshot above). When you click it, a new **Propose changes** box will appear on your screen. Type a brief (i.e., 72 characters or less) summary of the change you made in the **Commit message** box. There is also an **Extended description** box where you can add a more detailed description of what you did. In the screenshot below, shows an example commit message and extended description that will make it easy for the author to quickly figure out exactly what changes are being proposed.

Propose changes ×

Commit message

Fix a typo in contributing.qmd

Extended description

- Changed "typoo" to "typo" on line 11.

Cancel Propose changes

Next, click the **Propose changes** button. That will take you to another screen where you will be able to create a pull request. This screen is kind of busy, but try not to let it overwhelm you.

Comparing changes

Choose two branches to see what's changed or to start a new pull request. If you need to, you can also [compare across forks](#) or [learn more about diff comparisons](#).

base repository: brad-cannell/r4epi_quarto base: main head repository: arthur-epi/r4epi_quarto compare: patch-1

✓ Able to merge. These branches can be automatically merged.

Discuss and review the changes in this comparison with others. [Learn about pull requests](#) **Create pull request**

1 commit 1 file changed 1 contributor

Commits on Dec 15, 2023

Fix a typo in contributing.qmd arthur-epi committed now

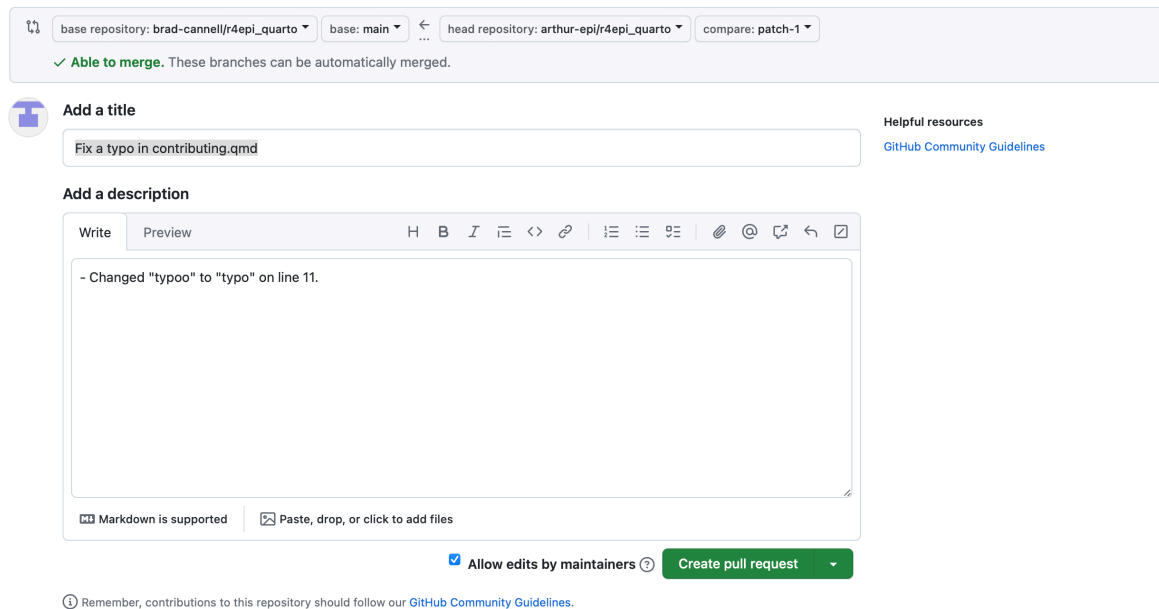
Showing 1 changed file with 2 additions and 2 deletions. Split Unified

```
@@ -8,7 +8,7 @@ The easiest way for you to contribute is to help us clean up the little typos an
8 8
9 9 If you spot a typo, you can offer a correction directly in GitHub. You will first need to create a free GitHub account: [sign-up at
github.com](https://github.com/join). Later in the book, we will cover using GitHub in greater depth (See @sec-using-git-and-github).
Here, we're just going to walk you through how to fix a typo without much explanation of how GitHub works.
10 10
11 11 - Let's say you spot a typoo while reading along.
11 11 + Let's say you spot a typo while reading along.
12 12
13 13 ```{r}
14 14 #| label: contributing_typo_on_screen
```

For now, we will focus on the three different sections of the screen that are highlighted with a red outline. We will start at the bottom and work our way up. The red box that is closest to the bottom of the screenshot shows us that the change that made was on line 11. The word “typoo” (highlighted in red) was replaced with the word “typo” (highlighted in green). The red box in the middle of the screenshot shows us the brief description that was written for our proposed change – “Fix a typo in contributing.qmd”. Finally, the red box closest to the top of the screenshot is surrounding the **Create pull request** button. You will click it to move on with your pull request.

Open a pull request

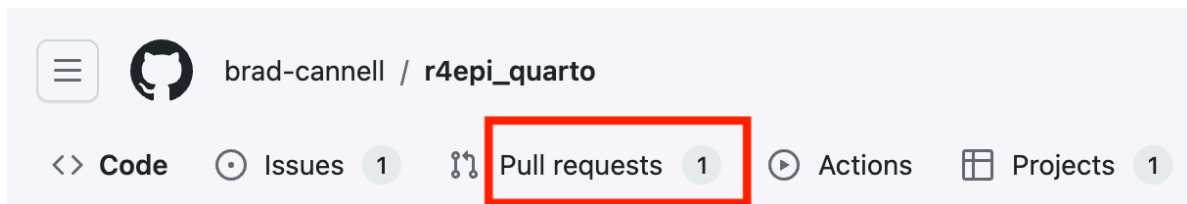
Create a new pull request by comparing changes across two branches. If you need to, you can also [compare across forks](#). [Learn more about diff comparisons here](#).



The screenshot shows the GitHub interface for creating a pull request. At the top, it displays the base repository as 'brad-cannell/r4epi_quarto' with the 'main' branch, and the head repository as 'arthur-epi/r4epi_quarto' with the 'patch-1' branch. A green checkmark indicates 'Able to merge'. Below this, there's a section to 'Add a title' with the text 'Fix a typo in contributing.qmd'. The 'Add a description' section has a text area containing '- Changed "typoo" to "typo" on line 11.' and a rich text editor toolbar. At the bottom, there's a checkbox for 'Allow edits by maintainers' and a green 'Create pull request' button. A footer note mentions the GitHub Community Guidelines.


After doing so, you will get one final chance to amend the description of your proposed changes. If you are happy with the commit message and description, then click the **Create pull request** button one more time. At this point, your job is done! It is now up to the authors to review the changes you've proposed and "pull" them into the file in their repository.


In case you are curious, here is what the process looks like on the authors' end. First, when we open the R4Epi repository page on GitHub, we will see that there is a new pull request.



When we open the pull request, we can see the proposed changes to the file.

Fix a typo in contributing.qmd #7

 Open

 arthur-epi wants to merge 1 commit into [brad-cannell:main](#) from [arthur-epi:patch-1](#) 

Conversation 0 Commits 1 Checks 0 Files changed 1

Changes from all commits File filter Conversations Jump to 0 / 1 files viewed Review in codespace Review changes

4 chapters/contributing/contributing.qmd

@@ -8,7 +8,7 @@ The easiest way for you to contribute is to help us clean up the little typos an

8 8

9 9 If you spot a typo, you can offer a correction directly in GitHub. You will first need to create a free GitHub account: [sign-up at github.com](https://github.com/join). Later in the book, we will

10 10 cover using GitHub in greater depth (See [@sec-using-git-and-github](#)). Here, we're just going to walk you through how to fix a typo without much explanation of how GitHub works.


11 10 - Let's say you spot a **typoo** while reading along.


11 11 + Let's say you spot a **typo** while reading along.

12 12



Then, all we have to do is click the **Merge pull request** button and the fixed file is “pulled in” to replace the file with the typo.

Fix a typo in contributing.qmd #7


 Open



 arthur-epi wants to merge 1 commit into [brad-cannell:main](#) from [arthur-epi:patch-1](#) 

Conversation 0 Commits 1 Checks 0 Files changed 1

 arthur-epi commented 3 minutes ago First-time contributor 


- Changed "typoo" to "typo" on line 11.






 Fix a typo in contributing.qmd 



Verified fd68f78

Add more commits by pushing to the [patch-1](#) branch on [arthur-epi/r4epi_quarto](#).

 **This branch has not been deployed**
No deployments

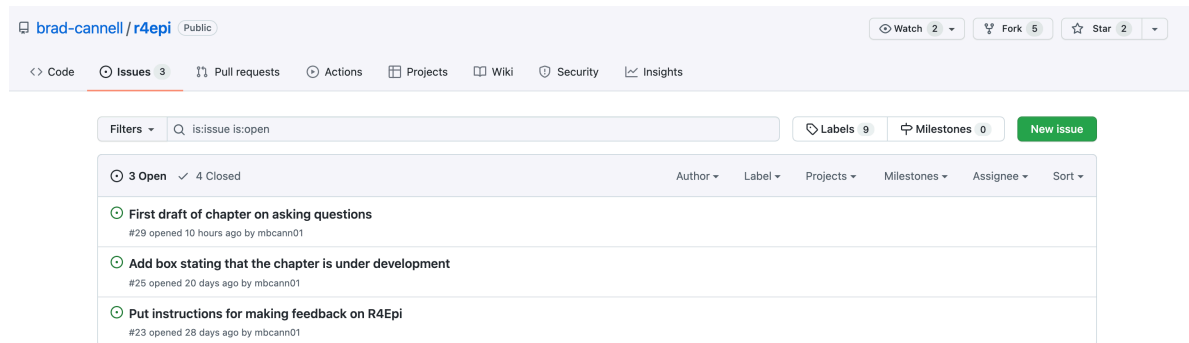
 **Require approval from specific reviewers before merging**
[Branch protection rules](#) ensure specific people approve pull requests before they're merged. Add rule 

 **This branch has no conflicts with the base branch**
Merging can be performed automatically.

 Merge pull request  You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

Issues

There may be times when you see a problem that you don't know how to fix, but you still want to make the authors aware of. In that case, you can create an [issue](#) in the R4Epi repository. To do so, navigate to the issue tracker using this link: <https://github.com/brad-cannell/r4epi/issues>.



Once there, you can check to see if someone has already raised the issue you are concerned about. If not, you can click the green “New issue” button to raise it yourself.

Please note that R4Epi uses a [Contributor Code of Conduct](#). By contributing to this book, you agree to abide by its terms.

License Information

This book was created by Brad Cannell and is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

About the Authors

Brad Cannell

Michael (Brad) Cannell, PhD, MPH

Associate Professor

Elder Mistreatment Lead, UTHealth Institute of Aging

Director, Research Informatics Core, Cizik Nursing Research Institute

UTHealth Houston

McGovern Medical School

Joan and Stanford Alexander Division of Geriatric & Palliative Medicine

www.bradcannell.com

Dr. Cannell received his PhD in Epidemiology, and Graduate Certificate in Gerontology, in 2013 from the University of Florida. He received his MPH with a concentration in Epidemiology from the University of Louisville in 2009, and his BA in Political Science and Marketing from the University of North Texas in 2005. During his doctoral studies, he was a Graduate Research Assistant for the Florida Office on Disability and Health, an affiliated scholar with the Claude D. Pepper Older Americans Independence Center, and a student-inducted member of the Delta Omega Honorary Society in Public Health. In 2016, Dr. Cannell received a Graduate Certificate in Predictive Analytics from the University of Maryland University College, and a Certificate in Big Data and Social Analytics from the Massachusetts Institute of Technology.

He previously held professional staff positions in the Louisville Metro Health Department and the Northern Kentucky Independent District Health Department. He spent three years as a project epidemiologist for the Florida Office on Disability and Health at the University of Florida. He also served as an Environmental Science Officer in the United States Army Reserves from 2009 to 2013.

Dr. Cannell's research is broadly focused on healthy aging and health-related quality of life. Specifically, he has published research focusing on preservation of physical and cognitive function, living and aging with disability, and understanding and preventing elder mistreatment. Additionally, he has a strong background and training in epidemiologic methods and predictive analytics. He has been principal or co-investigator on multiple trials and observational studies in community and healthcare settings. He is currently the principal investigator on multiple data-driven federally funded projects that utilize technological solutions to public health issues in novel ways.

Contact

Connect with Dr. Cannell and follow his work.



Melvin Livingston

Melvin (Doug) Livingston, PhD

Research Associate Professor

Department of Behavioral, Social, and Health Education Sciences

Emory University Woodruff Health Sciences Center

Rollins School of Public Health

[Dr. Livingston's Faculty Profile](#)

Dr. Livingston is a methodologist with expertise in the application of quasi-experimental design principals to the evaluation for both community interventions and state policies. He has particular expertise in time series modeling, mixed effects modeling, econometric methods, and power analysis. As part of his work involving community trials, he has been the statistician on the long term follow-up study of a school based cluster randomized trial in low-income communities with a focus on explaining the etiology of risky alcohol, drug, and sexual behaviors. Additionally, he was the statistician for a longitudinal study examining the etiology of alcohol use among racially diverse and economically disadvantaged urban youth, and co-investigator for a NIAAA- and NIDA-funded trial to prevent alcohol use and alcohol-related problems among youth living in high-risk, low-income communities within the Cherokee Nation. Prevention work at the community level led him to an interest in the impact of state and federal socioeconomic policies on health outcomes. He is a Co-Investigator of a 50-state, 30-year study of effects of state-level economic and education policies on a diverse set of public health outcomes, explicitly examining differential effects across disadvantaged subgroups of the population.

His current research interests center around the application of quasi-experimental design and econometric methods to the evaluation of the health effects of state and federal policy.

Contact

Connect with Dr. Livingston and follow his work.



Part I

Coding Tools and Best Practices

1 Quarto Files

Part II

Collaboration

2 Using git and GitHub

Part III

References

References

1. GitHub. About repositories. Published online December 2023.
2. GitHub. *About Issues*. Github; 2024.
3. R Development Core Team. *An Introduction to R.*; 2020.
4. RStudio. R markdown. Published online 2020.
5. RStudio. RStudio. Published online 2020.

A Glossary

Console. Coming soon.

Data frame. For our purposes, data frames are just R's term for data set or data table. Data frames are made up of columns (variables) and rows (observations). In R, all columns of a data frame must have the same length.

Functions. Coming soon.

- **Arguments:** Arguments always go *inside* the parentheses of a function and give the function the information it needs to give us the result we want.
- **Pass:** In programming lingo, you *pass* a value to a function argument. For example, in the function call `seq(from = 2, to = 100, by = 2)` we could say that we passed a value of 2 to the `from` argument, we passed a value of 100 to the `to` argument, and we passed a value of 2 to the `by` argument.
- **Returns:** Instead of saying, “the `seq()` function *gives us* a sequence of numbers...” we could say, “the `seq()` function *returns* us a sequence of numbers...” In programming lingo, functions *return* one or more results.

Repository GitHub's documentation says “a repository contains all of your code, your files, and each file's revision history. You can discuss and manage your work within the repository.”¹ A repository can exist *locally* as a set of files on your computer. A repository can also exist *remotely* as a set of files on a server somewhere, for example, on GitHub.

Global environment. Coming soon.

Issue (GitHub) GitHub's documentation says issues are “items you can create in a repository to plan, discuss and track work. Issues are simple to create and flexible to suit a variety of scenarios. You can use issues to track work, give or receive feedback, collaborate on ideas or tasks, and efficiently communicate with others.”²

Objects. Coming soon.

R. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly and has been extended by a large collection of packages. However, most programs written in R are essentially ephemeral, written for a single piece of data analysis.³

R markdown documents. R markdown documents are text files that can be used to clean and analyze your data interactively as well as share your final results in many different formats (e.g., Microsoft Word, PDF, and even websites). R markdown documents weave together R code, narrative text, and multimedia content together into a polished final product.⁴

RStudio. RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, Red Hat/CentOS, and SUSE Linux).⁵