Data Technologies

Computational tools for working with data

Martin Frigaard

2023-06-07

Table of contents

Welcome!

This book describes the essential tools and techniques for working with data, especially those 'behind-the-scenes' tasks that will take considerable time and effort but don't receive much attention.

Preface

This book is focused on computational tools for working with data. Data science is a rapidly evolving field, so in many ways, this text is an attempt to capture whatever 'best practices' can be codified or generalized.

The topics include coding languages, computational documents, data storage and file formats, version control systems, and general computer science topics.

This text emphasizes using tools that require typing code on a keyboard (rather than using a mouse to point and click).

This book was written in Quarto (learn more about Quarto books).

Introduction

> *This is a book created from markdown and executable code. See Knuth (1984) for additional discussion of literate programming.*

I've left the boilerplate text above because 1) everyone should know about Donald Knuth and his contributions to computational notebooks, and 2) I wanted to keep a working example for the `references.qmd` and `references.bib` files when I build the book.

Publishing this book (meta)

I'm going to include the steps for publishing this book *inside* the book so I can come back later and know what I've done (readers might also find this information useful).

This quarto book project has the following contents:

```
.
+-- _book
+-- _freeze
+-- _quarto.yml
+-- cover.png
+-- eRrata.Rproj
+-- fun-arguments.qmd
+-- fun-environments.qmd
+-- fun-names.qmd
+-- fun-structure.qmd
+-- index.html
+-- index.qmd
+-- intro.html
+-- intro.qmd
+-- intro.rmarkdown
+-- intro_files
+-- oop-basics.qmd
+-- oop-in-r.qmd
+-- preface.html
+-- preface.qmd
```

```
+-- references.bib
+-- references.qmd
+-- renv
+-- renv.lock
\-- site_libs
```

The GitHub repo for this book is here, and it's been set up to publish with GitHub pages.

Using Git

Basic knowledge of Git has become somewhat necessary when you decide to enter the R ecosystem (or other open-source projects). I won't be diving into Git commands here, but I *will* cover what commands I used to store and publish this book.

1. Create a repository on GitHub.com like this one: mjfrigaard/eRrata
2. Add and commit the files in the book repo:

```
$ git add -A
$ git commit -m "first commit"
```

3. Push the local repo to GitHub repo

```
$ git remote add origin git@github.com:mjfrigaard/eRrata.git
$ git branch -M main
$ git push -u origin main
```

4. Create a `gh-pages` branch

```
$ git checkout --orphan gh-pages
Switched to a new branch 'gh-pages'
$ git reset --hard
$ git commit --allow-empty -m "Initialising gh-pages branch"
Initialising gh-pages branch
$ git push origin gh-pages
Enumerating objects: 2, done.
Counting objects: 100% (2/2), done.
Writing objects: 100% (2/2), 176 bytes | 176.00 KiB/s, done.
Total 2 (delta 0), reused 0 (delta 0), pack-reused 0
remote:
remote: Create a pull request for 'gh-pages' on GitHub by visiting:
```

```
remote:      https://github.com/mjfrigaard/eRrata/pull/new/gh-pages
remote:
To github.com:mjfrigaard/eRrata.git
 * [new branch]      gh-pages -> gh-pages
```

5. Switch back to main branch

```
$ git checkout main
Switched to branch 'main'
Your branch is up to date with 'origin/main'.
```

6. Publish to `gh-pages` site

```
$ quarto publish gh-pages
? Update site at https://mjfrigaard.github.io/eRrata/? (Y/n)
```

```
? Update site at https://mjfrigaard.github.io/eRrata/? (Y/n) Yes
```

Part I

Object-oriented programming

Basics

Object-oriented programming is...

In R

R is a 'functional, object-oriented programming language.'

Part II

Functions

Names

Naming your functions should...

Arguments

Function arguments can be

Structure

Function structure refers to...

Environments

Inside your function...

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.