

Automating your R-based research project

Introduction to GitHub Actions

Ernest Guevarra

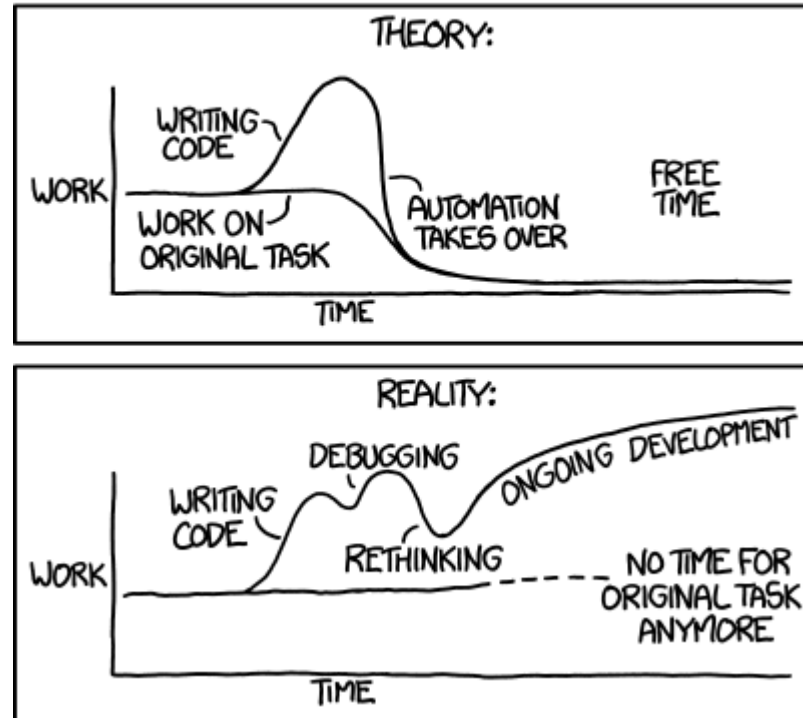
2022-03-13

Outline

- Overview of automation
- Introduction to GitHub Actions
- Demonstration of automation

Automation

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



'Automating' comes from the roots 'auto-' meaning 'self-', and 'mating', meaning 'screwing'

What can be automated?

- works best with a **project workflow** (sometimes called a **project pipeline**)
 - our recommended solution for this is the `{targets}` workflow/pipeline
- works best if project workflow is **portable**
 - our recommended solution for this is `{renv}`

Why automate?

- to reproduce previous results
- to recreate results deleted by *fat fingers*
- to rerun the workflow/pipeline with updated software
- to run the same workflow/pipeline on a new dataset

Example GitHub Actions automation

```
1 name: check assignment
2
3 on:
4   push:
5     branches: [main, master]
6   pull_request:
7     branches: [main, master]
8   workflow_dispatch:
9     branches:
10      - '*'
11
12 jobs:
13   check-assignment:
14     runs-on: ubuntu-latest
15     container: rocker/geospatial:4.1.2
16     steps:
17       - uses: actions/checkout@v2
18
19       - name: Install system dependencies
20         run: |
21           apt-get update && apt-get install -y --no-install-recommends \
22             git-crypt libxt6 libglpk-dev
23
24       - name: Install packages from renv.lock (with cache)
25         uses: r-lib/actions/setup-renv@master
26         with:
27           cache-version: 2
28
29       - name: Check that targets workflow has been enabled
30         run: |
31           testthat::test_that("Target script file exists",
32             testthat::expect_true(
33               list.files(pattern = "_targets.R") == "_targets.R"
34             )
35         )
36         shell: Rscript {0}
37
38       - name: Check that renv has been enabled/initialised
39         run: |
40           testthat::test_that("renv has been enabled", {
41             testthat::expect_true(
42               list.files(pattern = "renv.lock") == "renv.lock",
43             )
44             testthat::expect_true(
45               list.files(pattern = ".Rprofile", all.files = TRUE) == ".Rprofile"
46             )
47             testthat::expect_true(
48               any(list.dirs(recursive = FALSE) == "./renv")
49             )
50           })
51         shell: Rscript {0}
52
53       - name: Check that targets workflow runs without errors
54         run: |
55           targets::tar_make()
56         shell: Rscript {0}
```

- Example automation built-in to the latest assignment on creating targets workflow
- Can be found at [.github/workflows/check-assignment.yml](#)

Setting the trigger for the automation

```
name: check assignment

on:
  push:
    branches: [main, master]
  pull_request:
    branches: [main, master]
  workflow_dispatch:
    branches:
      - '*'
```


Defining jobs to be implemented in the workflow

```
jobs:
  check-assignment:
    runs-on: ubuntu-latest
    container: rocker/geospatial:4.1.2
    steps:
      - uses: actions/checkout@v2
```

Specifying steps in the workflow

Workflow step 1: reproduce the system environment

```
- name: Install system dependencies
  run: |
    apt-get update && apt-get install -y --no-install-recommends \
    git-crypt libxt6 libglpk-dev
```

Workflow step 2: Reproduce the R project-local environment

```
- name: Install packages from renv.lock (with cache)
  uses: r-lib/actions/setup-renv@master
  with:
    cache-version: 2
```

Specifying steps in the workflow

Workflow step 3: Perform tests/checks

```
- name: Check that targets workflow has been enabled
run: |
  testthat::test_that("Target script file exists",
    testthat::expect_true(
      list.files(pattern = "_targets.R") == "_targets.R"
    )
  )
shell: Rscript {0}

- name: Check that renv has been enabled/initialised
run: |
  testthat::test_that("renv has been enabled", {
    testthat::expect_true(
      list.files(pattern = "renv.lock") == "renv.lock",
    )
    testthat::expect_true(
      list.files(pattern = ".Rprofile", all.files = TRUE) == ".Rprofile"
    )
    testthat::expect_true(
      any(list.dirs(recursive = FALSE) == "./renv")
    )
  })
```

Specifying steps in the workflow

Workflow step 4: Run the targets workflow

```
- name: Check that targets workflow runs without errors
  run: |
    targets::tar_make()
  shell: Rscript {0}
```

Questions?

Thank you!

Slides can be viewed at <https://oxford-ihtm.io/open-reproducible-science/session11.html>

PDF version of slides can be downloaded at <https://oxford-ihtm.io/open-reproducible-science/pdf/session11-automating-r-projects.pdf>

R scripts for slides available [here](#)