

Binder for fully reproducible research in R (data, code, and computational environment).

Andrew Stewart
University of Manchester
@ajstewart_lang



Inspired by The Turing Way
<https://github.com/alan-turing-institute/the-turing-way>

Open and Reproducible Research

- Shared Data - we already know this is important for reproducibility.
- Shared Code - we already know this is important for reproducibility.
- Shared Computational environment - why is this important and how do we do it?

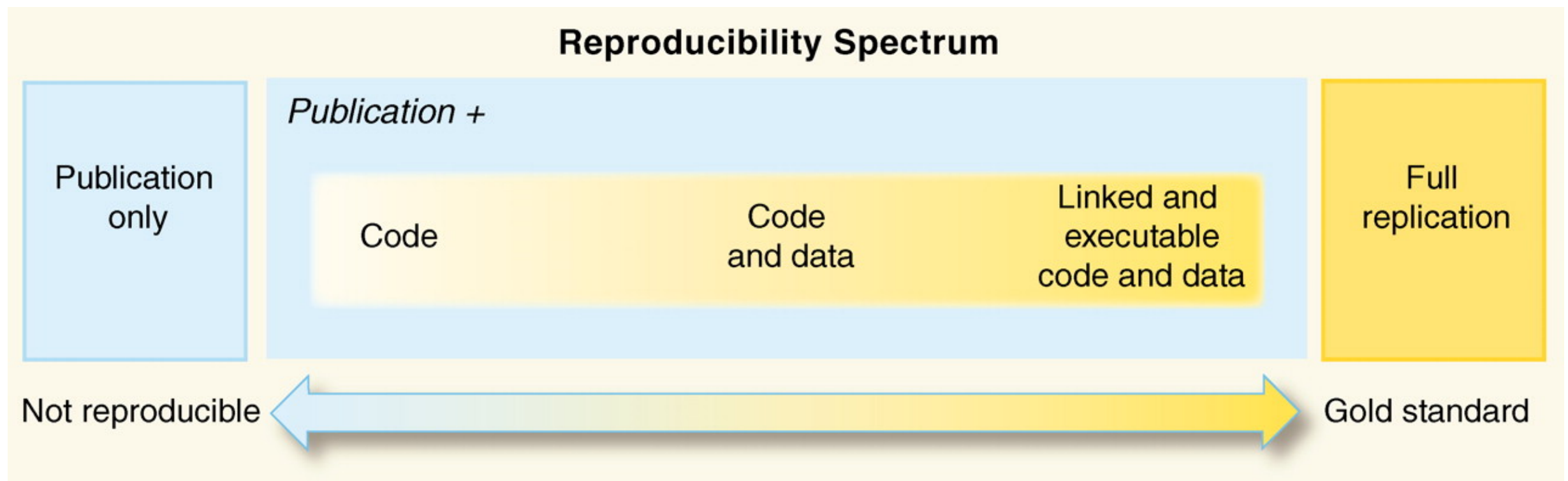
PERSPECTIVE

Reproducible Research in Computational Science

Roger D. Peng

+ See all authors and affiliations

Science 02 Dec 2011:
Vol. 334, Issue 6060, pp. 1226-1227
DOI: 10.1126/science.1213847



Why do we need to reproduce the computational environment?

- Quite often analysis code ‘breaks’ - often in one of two ways:
- Code that worked previously now doesn’t - maybe a function in an R package was updated (e.g., `lsmeans` became `emmeans` so old code using `lsmeans` wouldn’t now run).
- Code that worked previously still works - but produces a slightly different result or now throws a warning where it didn’t previously (e.g., convergence/singular fit warnings in `lme4` version 1.1-19 vs. version 1.1-20).

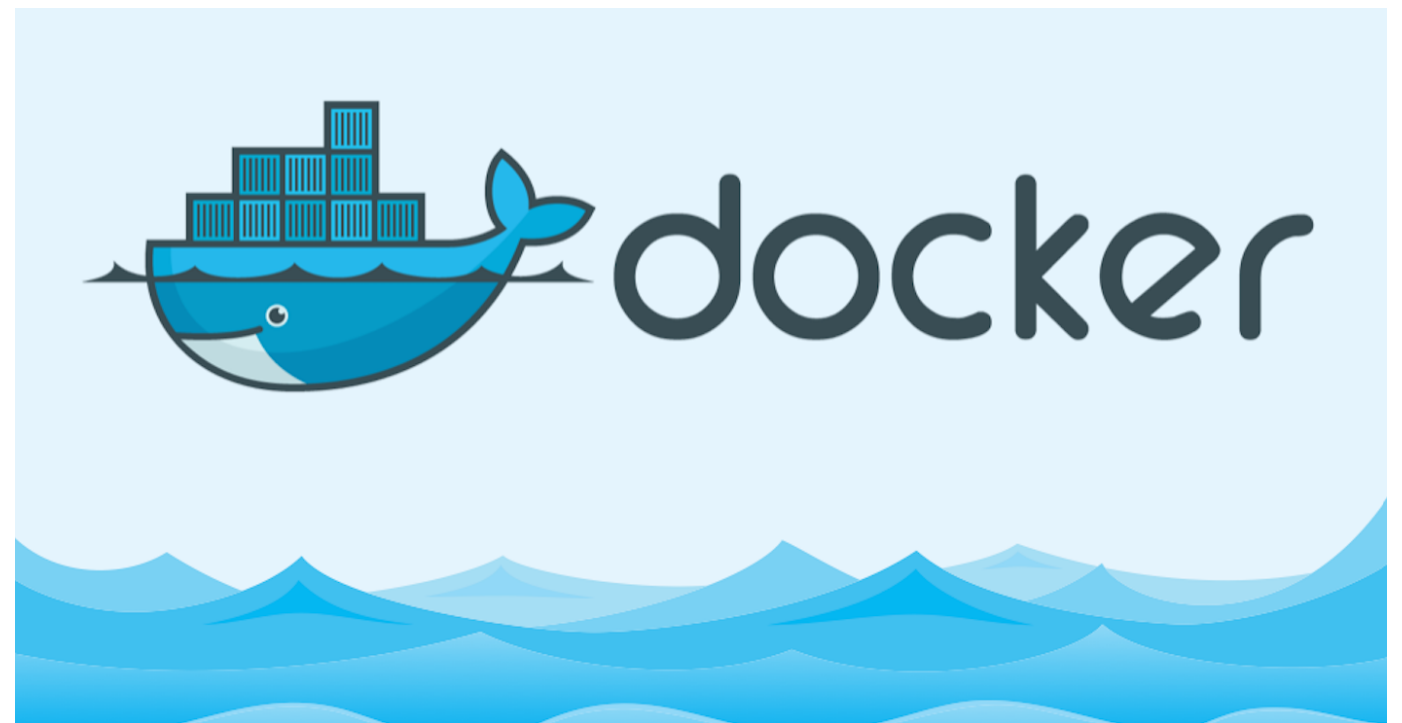
Capturing your local computational environment

- You need to capture the versions of the different R packages (plus their dependencies).
- May sound trivial but trying running some old R code and be amazed at how many things now don't work as they once did!

Docker for beginners

Docker packages your data, code and all its dependencies in the form called a docker container to ensure that your application works seamlessly in any environment.

When you run a docker container it's like running your analysis on a virtual computer that has the same configuration as our own one at the point in time when you ran the analysis.



<https://medium.com/the-andela-way/docker-for-beginners-61e8e0ce6a19>

So what's Binder?

- Binder is powered by BinderHub, which is an open-source tool that deploys the Binder service in the cloud.
- Binder works by pulling information from a repository that you set up on GitHub.
- Think of a repository as a folder containing your R code, your data, and a few other small bits and pieces - but it sits in the cloud rather than on your computer.

github.com/ajstewartlang/Turing_way2

Google Scholar Scopus jobs.ac.uk Apple BBC News Chester Weather The Telegraph The Grauniad The Independent Google Maps Chester Weather Station Favourites

Search or jump to... Pull requests Issues Marketplace Explore

ajstewartlang / Turing_way2 Watch 0 Star 0 Fork 1

Code Issues 0 Pull requests 1 Projects 0 Wiki Insights Settings

markdown_for_Turing_Way Edit

Manage topics

6 commits 1 branch 0 releases 1 contributor

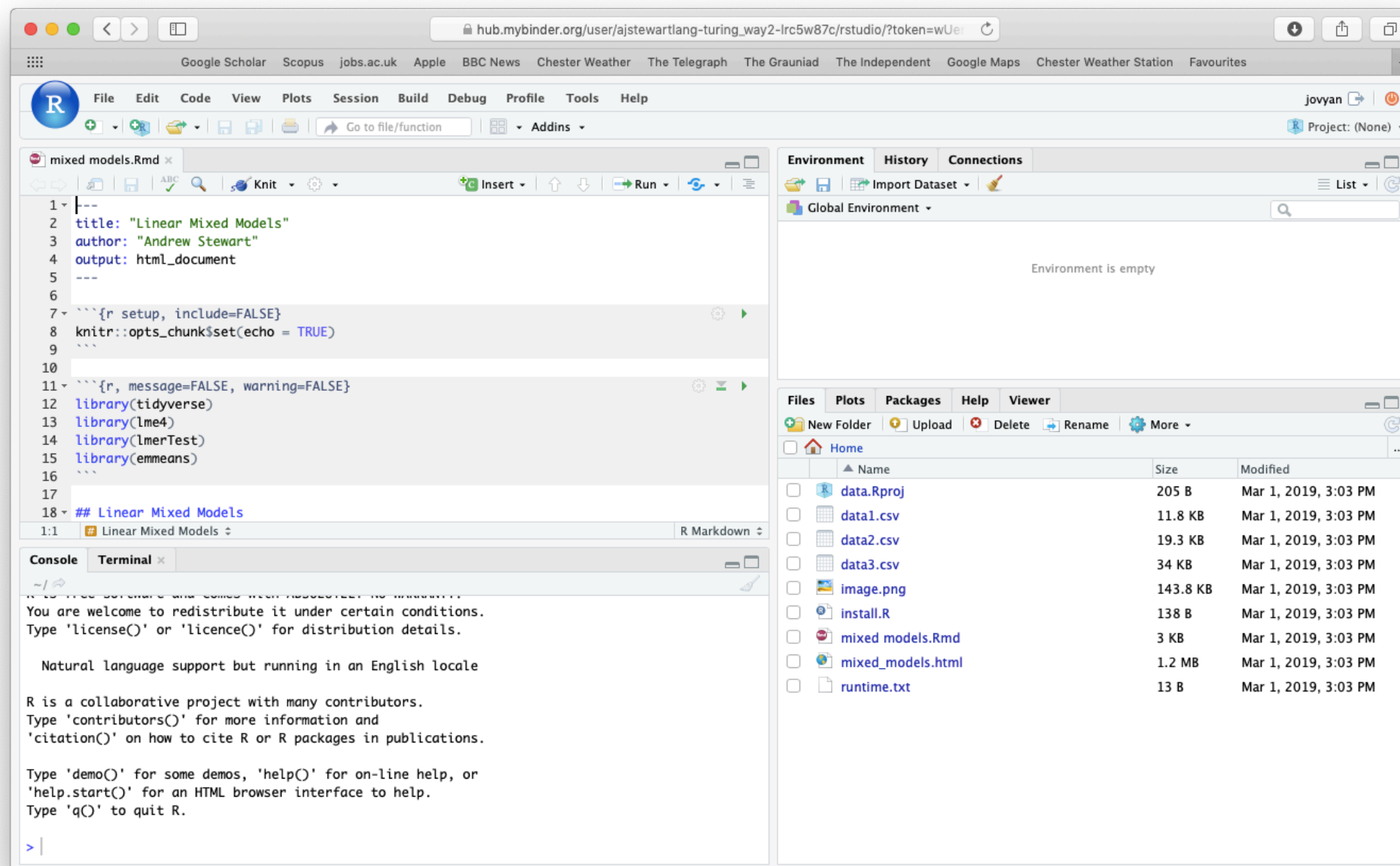
Branch: master New pull request Create new file Upload files Find file Clone or download

ajstewartlang Create install.R		Latest commit 36d3181 2 hours ago
.Rproj.user	commit	2 hours ago
data.Rproj	first commit	3 hours ago
data1.csv	first commit	3 hours ago
data2.csv	first commit	3 hours ago
data3.csv	first commit	3 hours ago
image.png	first commit	3 hours ago
install.R	Create install.R	2 hours ago
mixed models.Rmd	commit	2 hours ago
mixed_models.html	first commit	3 hours ago
runtime.txt	Update runtime.txt	2 hours ago

GitHub Desktop

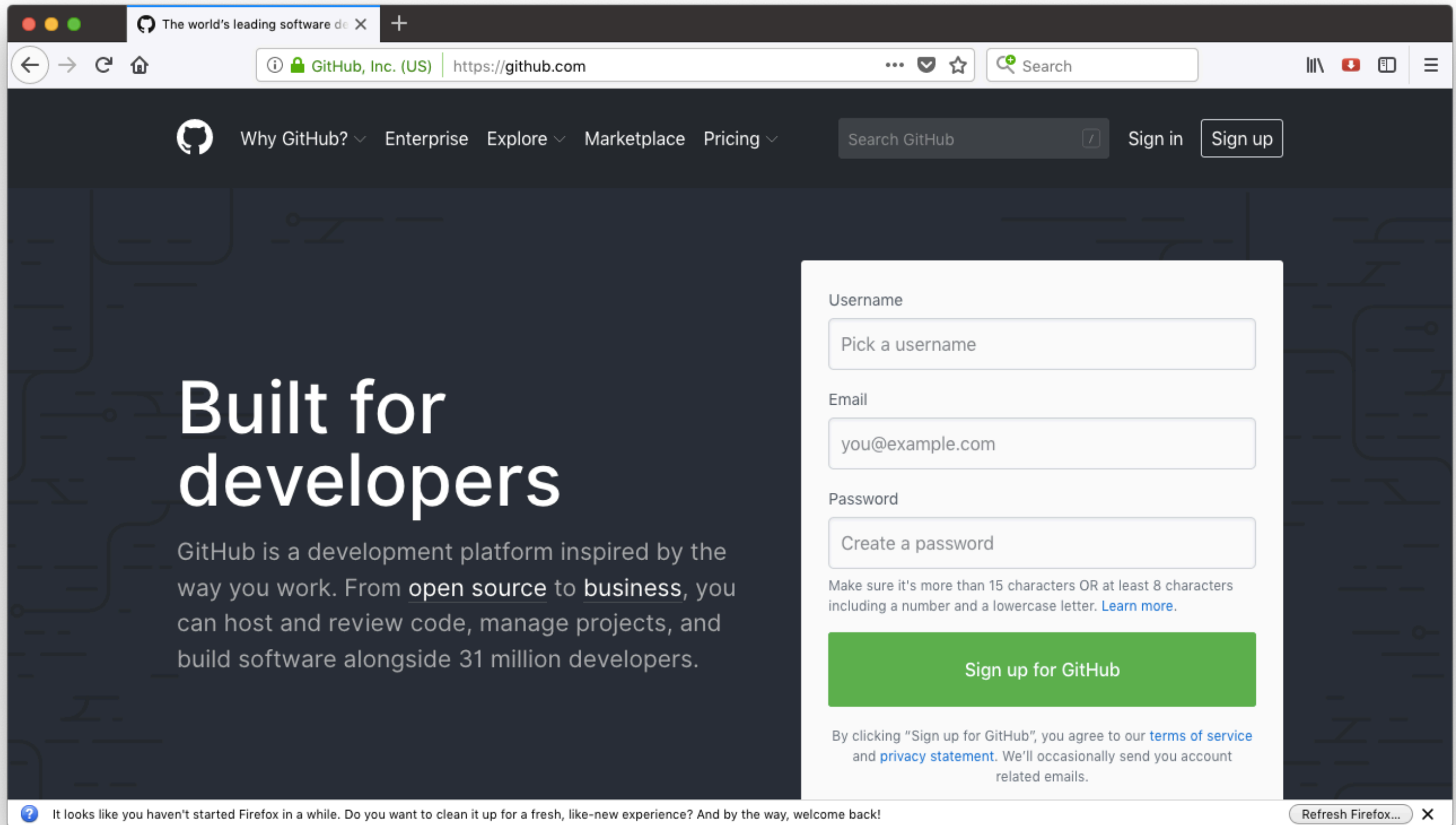
My R code and data files.

- When I link my GitHub repository to Binder and launch it I then get the following in my web browser.
- This is RStudio running the cloud using my code, my data and the appropriate versions of the packages that I was using when I did the analysis originally!



https://mybinder.org/v2/gh/ajstewartlang/Turing_way2/master?urlpath=rstudio

Step 1 - Set up a GitHub account



The screenshot shows the GitHub homepage in a web browser. The browser's address bar displays "https://github.com". The page features a dark blue header with the GitHub logo, navigation links ("Why GitHub?", "Enterprise", "Explore", "Marketplace", "Pricing"), a search bar, and "Sign in" and "Sign up" buttons. The main content area has a dark background with the text "Built for developers" and a description of GitHub as a development platform. A white sign-up form is overlaid on the right side of the page. The form includes fields for "Username", "Email", and "Password", each with a placeholder text. Below the password field is a note about password requirements and a link to "Learn more". A large green button labeled "Sign up for GitHub" is positioned below the form. At the bottom of the form, there is a disclaimer about agreeing to the terms of service and privacy statement. A Firefox notification bar is visible at the very bottom of the browser window.

Username

Pick a username

Email

you@example.com

Password

Create a password

Make sure it's more than 15 characters OR at least 8 characters including a number and a lowercase letter. [Learn more.](#)

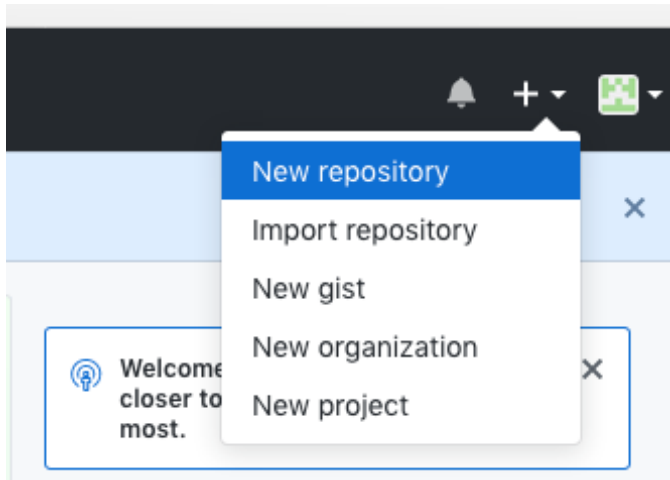
[Sign up for GitHub](#)

By clicking "Sign up for GitHub", you agree to our [terms of service](#) and [privacy statement](#). We'll occasionally send you account related emails.

It looks like you haven't started Firefox in a while. Do you want to clean it up for a fresh, like-new experience? And by the way, welcome back!

Refresh Firefox...

Step 2 - Create a new repository



Give it a name,
set it to public
and tick
“Initialise this
repository with a
README”.

Create a new repository

A repository contains all project files, including the revision history.

Owner

 andrewstewarttest ▾


Repository name *

first_binder ✓

Great repository names are short and memorable. Need inspiration? How about [probable-funicular?](#)

Description (optional)


☒  **Public**
Anyone can see this repository. You choose who can commit.

☐  **Private**
You choose who can see and commit to this repository.

☒ **Initialize this repository with a README**

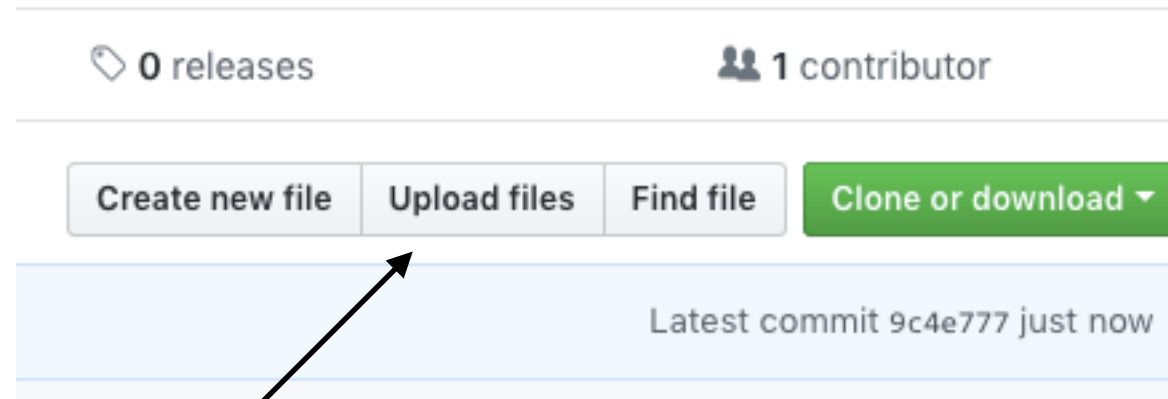
This will let you immediately clone the repository to your computer. Skip this step if you're importing an existing repository.

Add .gitignore: **None** ▾

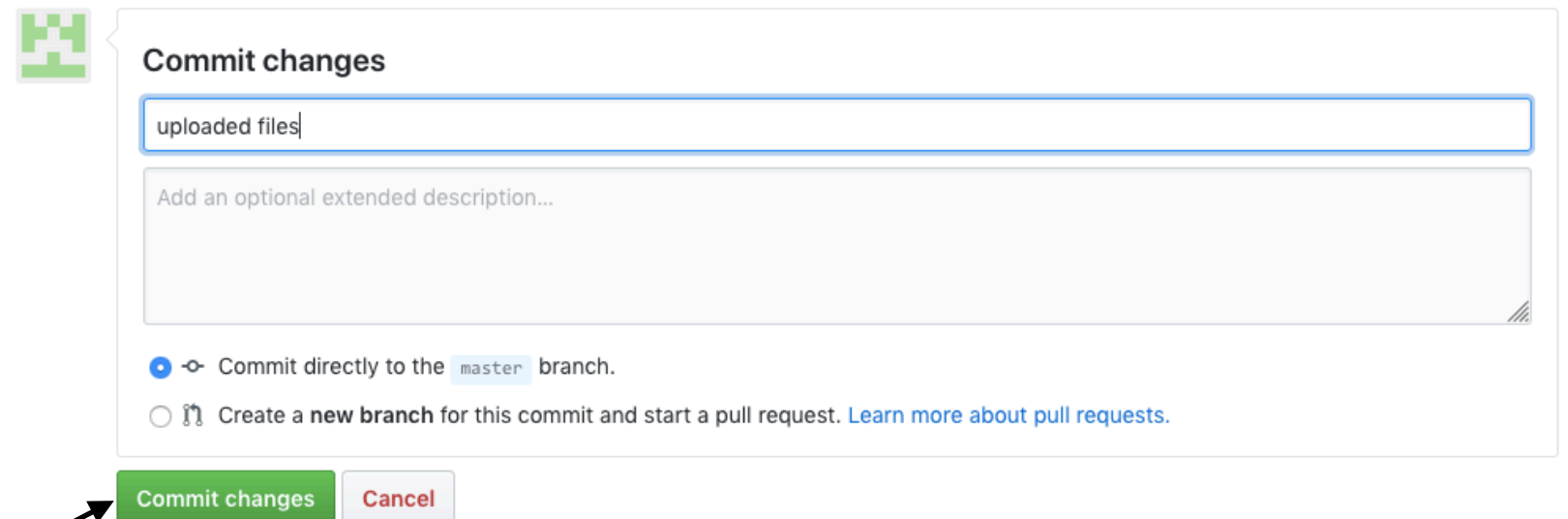
Add a license: **None** ▾ 

Create repository

Step 3 - Upload your R script and data and make your first “Commit”



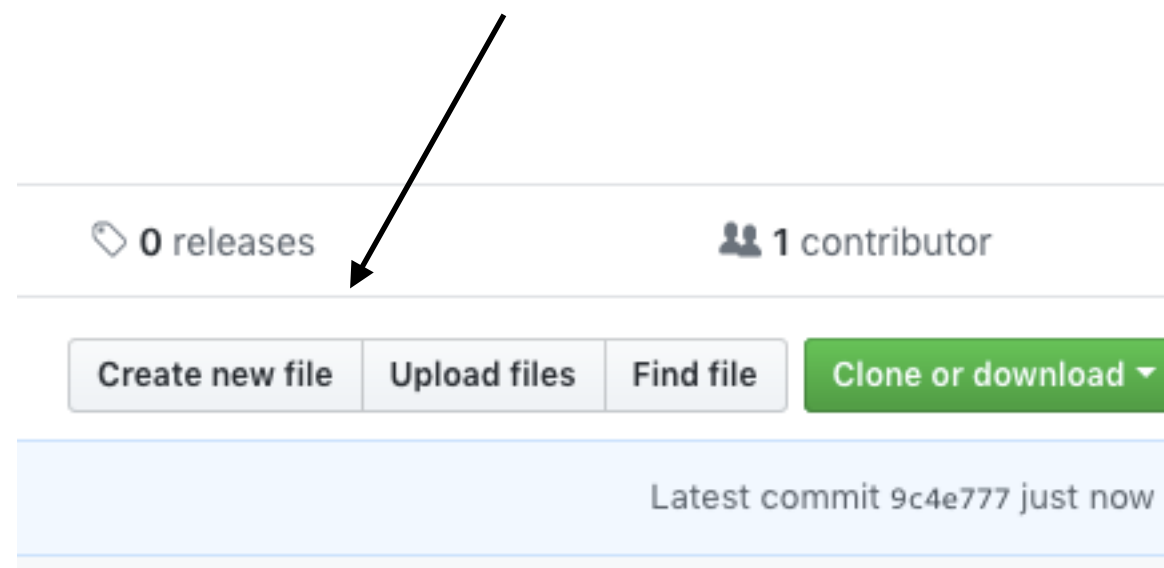
Click here to upload



Click here to Commit

Step 3 - Upload your R script and data and make your first “Commit”

- We need two other files at this point - one is called “runtime.txt” and contains the date of R and its associated packages that you want to simulate.
- The other is called “install.R” and contains the list of R packages that need to be installed in order for your script to run.
- To create a new file select “Create new file”



In the
runtime.txt file
type the date
you want in the
format r-YYYY-
MM-DD

andrewstewarttest / first_binder

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master first_binder / runtime.txt Find file Copy path

andrewstewarttest uploaded files 4f4ec8f 5 minutes ago

1 contributor

2 lines (1 sloc) | 13 Bytes Raw Blame History

```
1 r-2018-02-05
```

Name your file

List your
packages like
this in the
install.R file

andrewstewarttest / first_binder

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

first_binder / install.R or cancel

Edit new file Preview Spaces 2 No wrap

```
1 install.packages("tidyverse")
2 install.packages("knitr")
3 install.packages("lme4")
4 install.packages("lmerTest")
5 install.packages("emmeans")
```

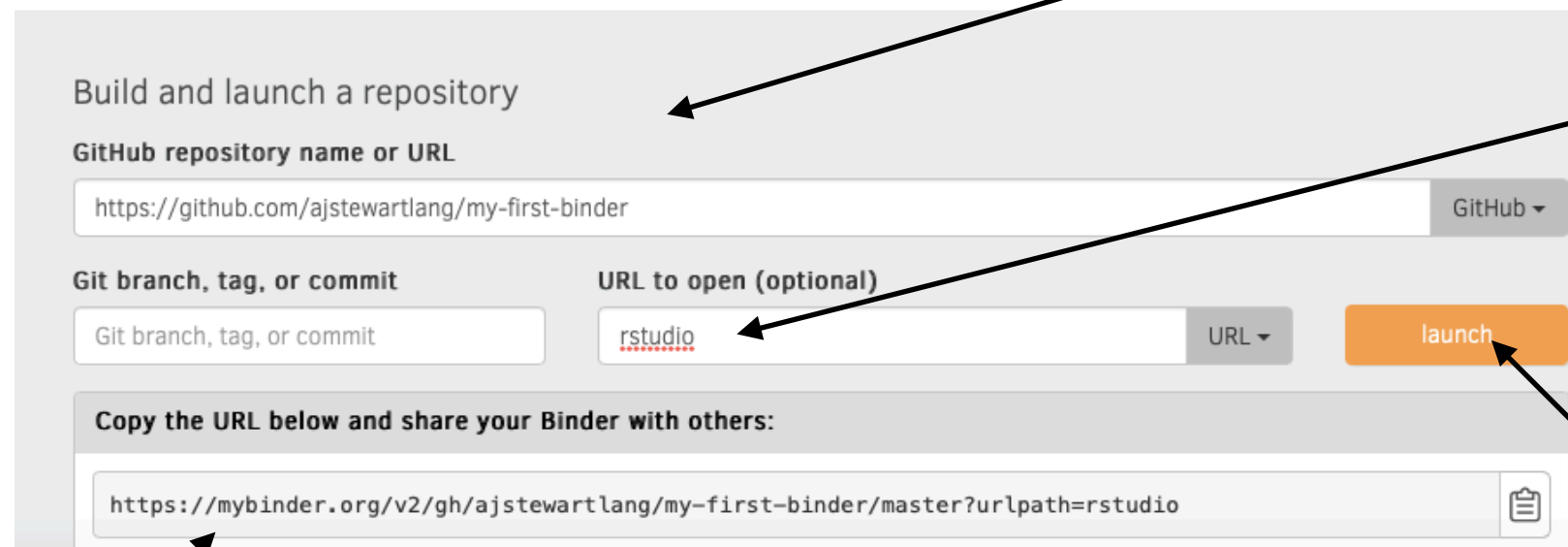
Don't forget to click "Commit" after you've created each file!

Step 5 - Now we need to link our repo to Binder (mybinder.org)



Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

A screenshot of the Binder web interface. It shows a form titled "Build and launch a repository". The first field is "GitHub repository name or URL" with the value "https://github.com/ajstewartlang/my-first-binder" and a "GitHub" dropdown. Below this are two fields: "Git branch, tag, or commit" (empty) and "URL to open (optional)" (containing "rstudio" with a red error message). To the right of the second field is a "URL" dropdown and an orange "launch" button. At the bottom, there is a section "Copy the URL below and share your Binder with others:" containing a text box with the URL "https://mybinder.org/v2/gh/ajstewartlang/my-first-binder/master?urlpath=rstudio" and a clipboard icon. Arrows from the numbered instructions point to the repository URL field, the "rstudio" input, the "launch" button, and the final URL output.

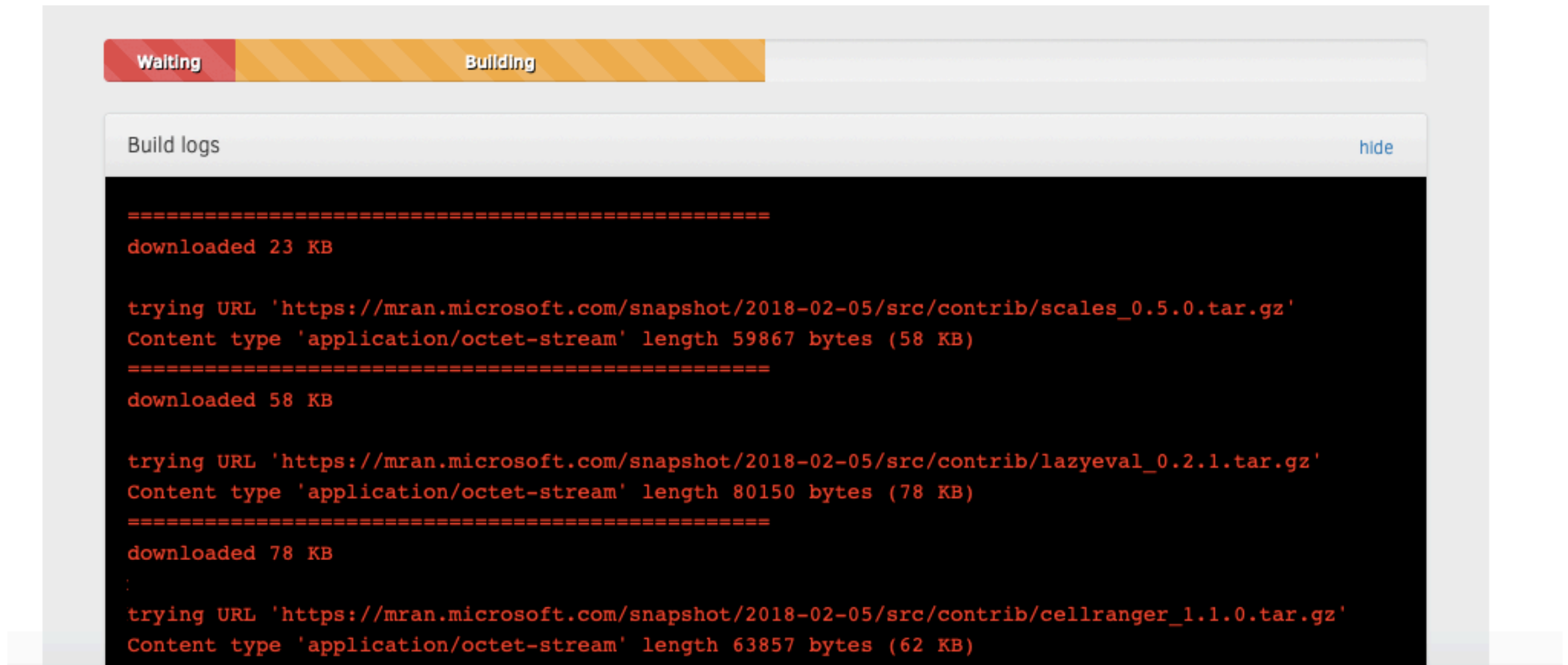
1. Paste the link to your repo here.

2. Type rstudio here and select "URL"

3. Then click on "launch"

4. This is the URL to share with others.

And wait...



You can check the progress of the build by clicking on the “Build logs” bar.

And wait...

And then...

The screenshot shows the RStudio web interface running in a browser. The browser's address bar displays the URL: `hub.mybinder.org/user/andrewstewarttest-first_binder-z4kwp6gl/rstudio/?token=`. The browser's top bar includes search engines like Google Scholar, Scopus, and various news sources.

The RStudio interface has a menu bar with options: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar.

The main editor window displays a file named `mixed models.Rmd`. The code is as follows:

```
1 ---
2 title: "Linear Mixed Models"
3 author: "Andrew Stewart"
4 output: html_document
5 ---
6
7 ```{r setup, include=FALSE}
8 knitr::opts_chunk$set(echo = TRUE)
9 ```
10
11 ```{r, message=FALSE, warning=FALSE}
12 library(tidyverse)
13 library(lme4)
14 library(lmerTest)
15 library(emmeans)
16 ```
17
18 ## Linear Mixed Models
```

Below the editor is a console window showing the R startup message:

```
~/ |
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

On the right side of the interface, there are three panels: Environment, History, and Connections. The Environment panel shows "Global Environment" and "Environment is empty". Below these panels is a file explorer showing the contents of the "Home" directory:

Name	Size	Modified
data.Rproj	205 B	Mar 1, 2019, 5:39 PM
data1.csv	11.8 KB	Mar 1, 2019, 5:39 PM
data2.csv	19.3 KB	Mar 1, 2019, 5:39 PM
data3.csv	34 KB	Mar 1, 2019, 5:39 PM
image.png	143.8 KB	Mar 1, 2019, 5:39 PM
install.R	138 B	Mar 1, 2019, 5:39 PM
mixed models.Rmd	3 KB	Mar 1, 2019, 5:39 PM
mixed_models.html	1.2 MB	Mar 1, 2019, 5:39 PM
README.md	14 B	Mar 1, 2019, 5:39 PM
runtime.txt	13 B	Mar 1, 2019, 5:39 PM