

# Slides

[http://ajstewartlang.github.io/Binder\\_slides.pdf](http://ajstewartlang.github.io/Binder_slides.pdf)

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

Andrew Stewart  
University of Manchester  
@ajstewart\_lang



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?

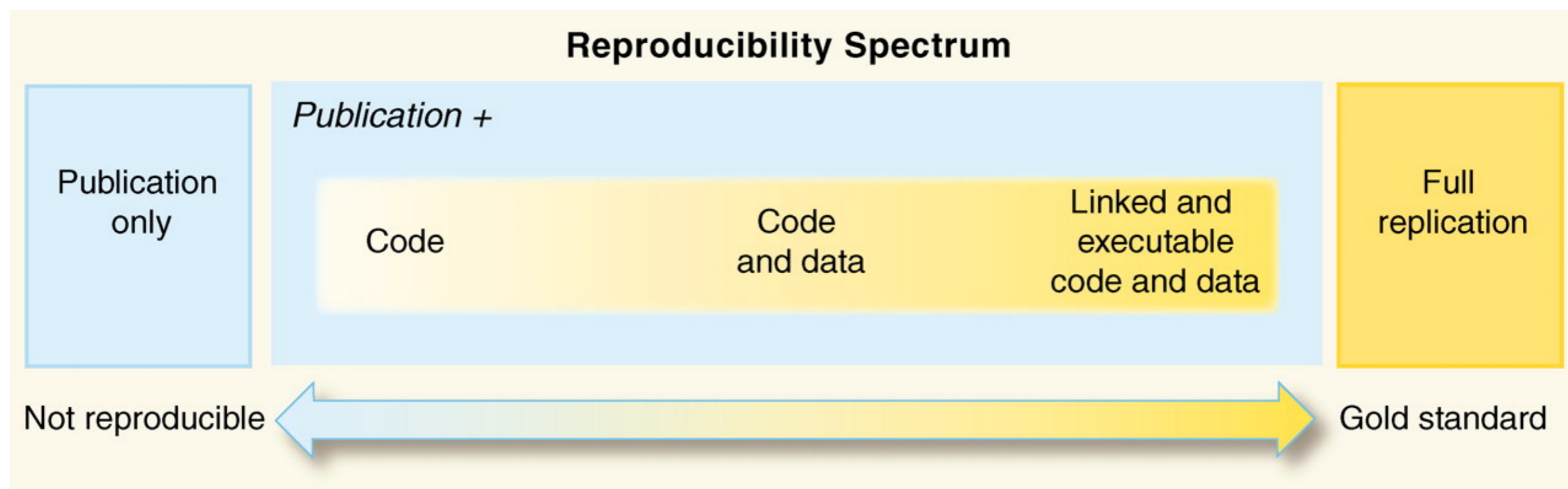
PERSPECTIVE

# 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.



# 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 a repository that you set up on GitHub into a Docker container.
- 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.



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

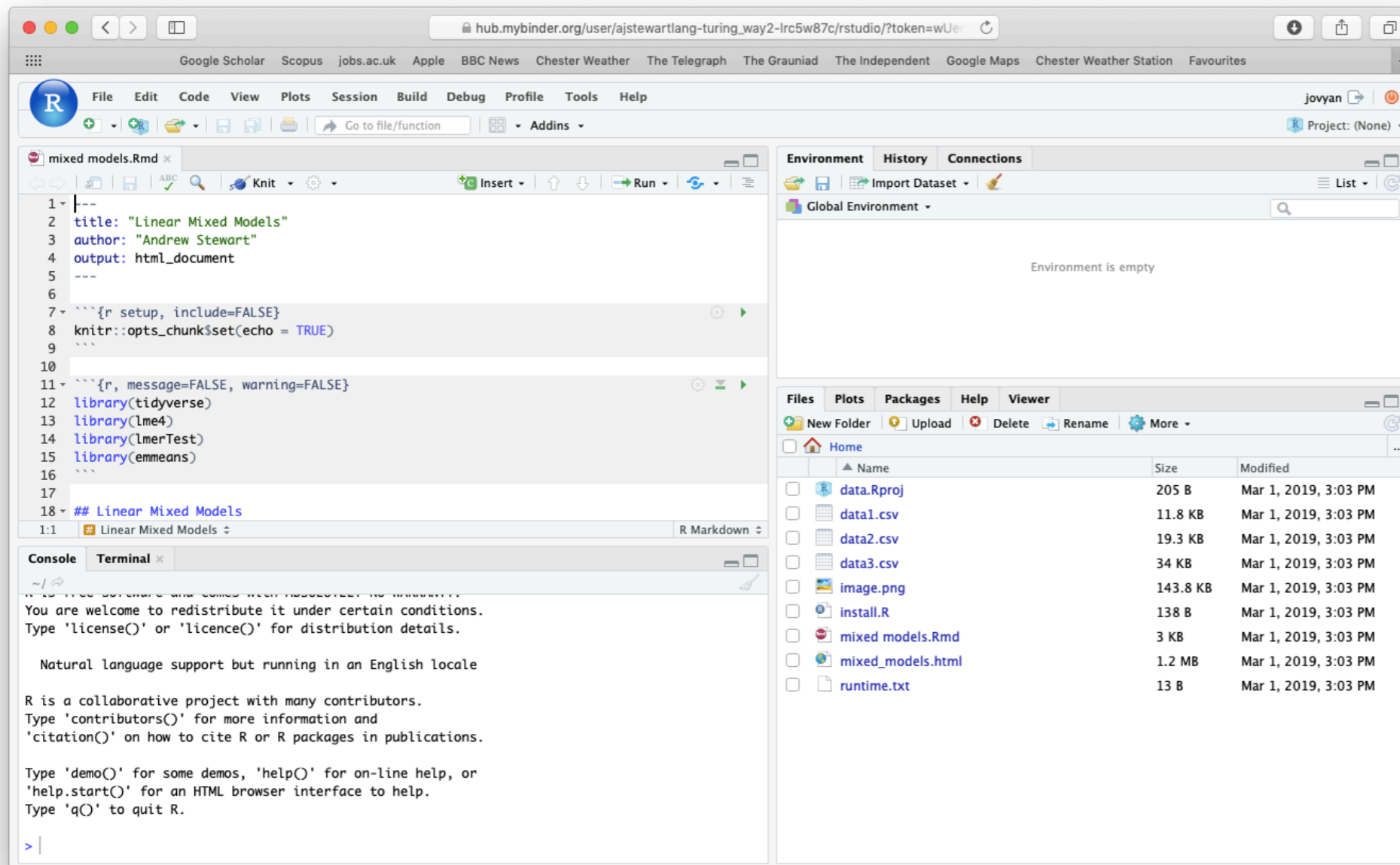
Branch: master New pull request Create new file Upload files Find file Clone or download

File	Commit	Time
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



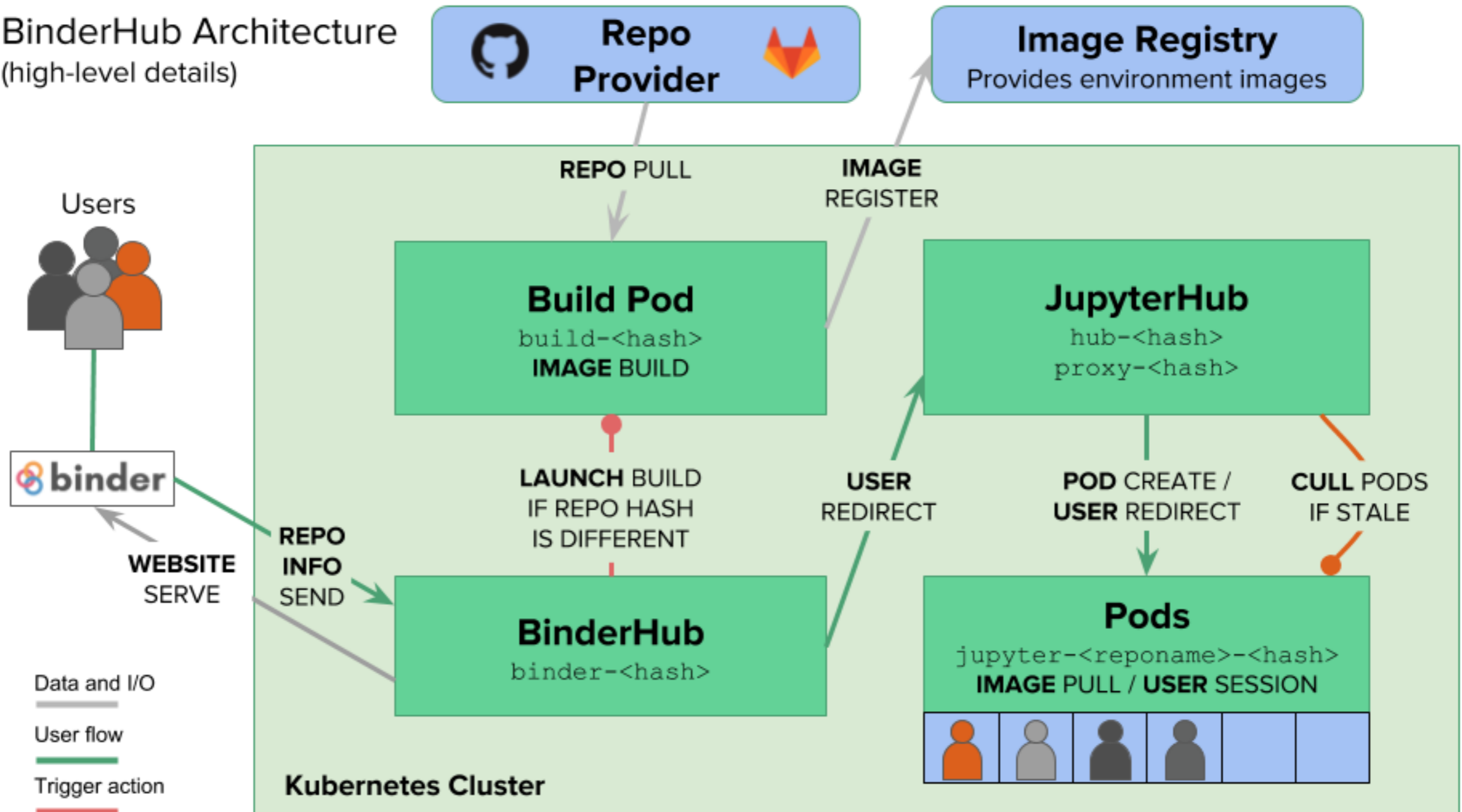
My R code and data files.

- 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](https://mybinder.org/v2/gh/ajstewartlang/Turing_way2/master?urlpath=rstudio)

# BinderHub Architecture (high-level details)



<https://binderhub.readthedocs.io/en/latest/index.html>

# Step 1 - Set up a GitHub account

The screenshot shows the GitHub homepage in a browser window. The browser's address bar displays "https://github.com". The navigation menu includes "Why GitHub?", "Enterprise", "Explore", "Marketplace", and "Pricing". A search bar and "Sign in" / "Sign up" buttons are visible. The main content area features the heading "Built for developers" and a sub-headline: "GitHub is a development platform inspired by the way you work. From open source to business, you can host and review code, manage projects, and build software alongside 31 million developers." A sign-up form is overlaid on the right side, containing fields for "Username" (with placeholder "Pick a username"), "Email" (with placeholder "you@example.com"), and "Password" (with placeholder "Create a password"). Below the password field, a note states: "Make sure it's more than 15 characters OR at least 8 characters including a number and a lowercase letter. Learn more." A green "Sign up for GitHub" button is positioned below the form. At the bottom of the form, a disclaimer reads: "By clicking 'Sign up for GitHub', you agree to our terms of service and privacy statement. We'll occasionally send you account related emails." A Firefox notification bar at the bottom of the browser window says: "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!" with a "Refresh Firefox..." button.

Why GitHub? Enterprise Explore Marketplace Pricing

Search GitHub Sign in Sign up

## Built for developers

GitHub is a development platform inspired by the way you work. From open source to business, you can host and review code, manage projects, and build software alongside 31 million developers.

Username

Email

Password

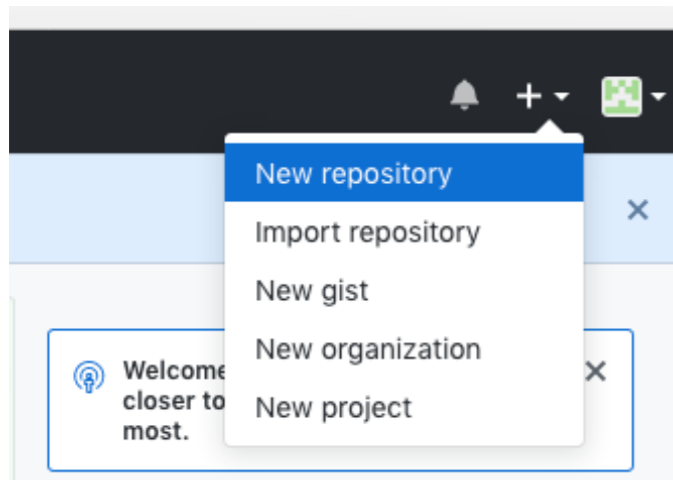
Make sure it's more than 15 characters OR at least 8 characters including a number and a lowercase letter. [Learn more.](#)

[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... X

# 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.

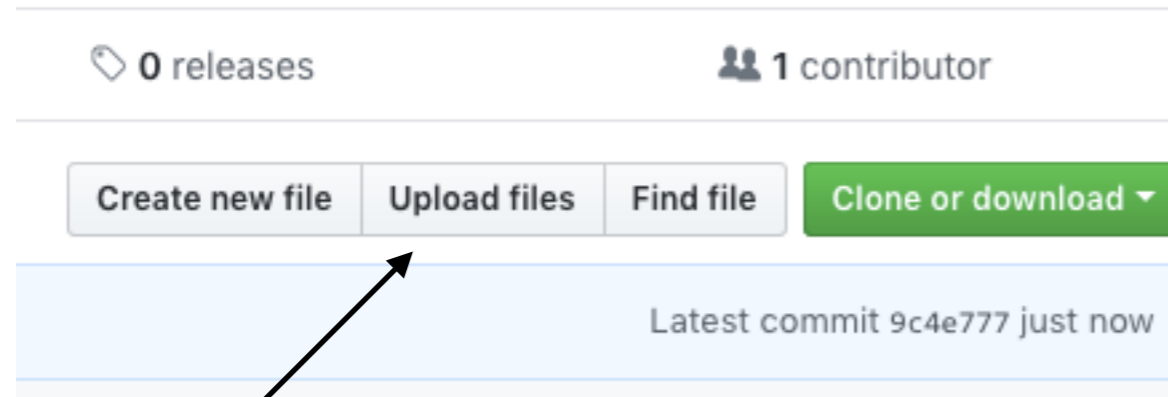
Add .gitignore: **None** ▾

Add a license: **None** ▾

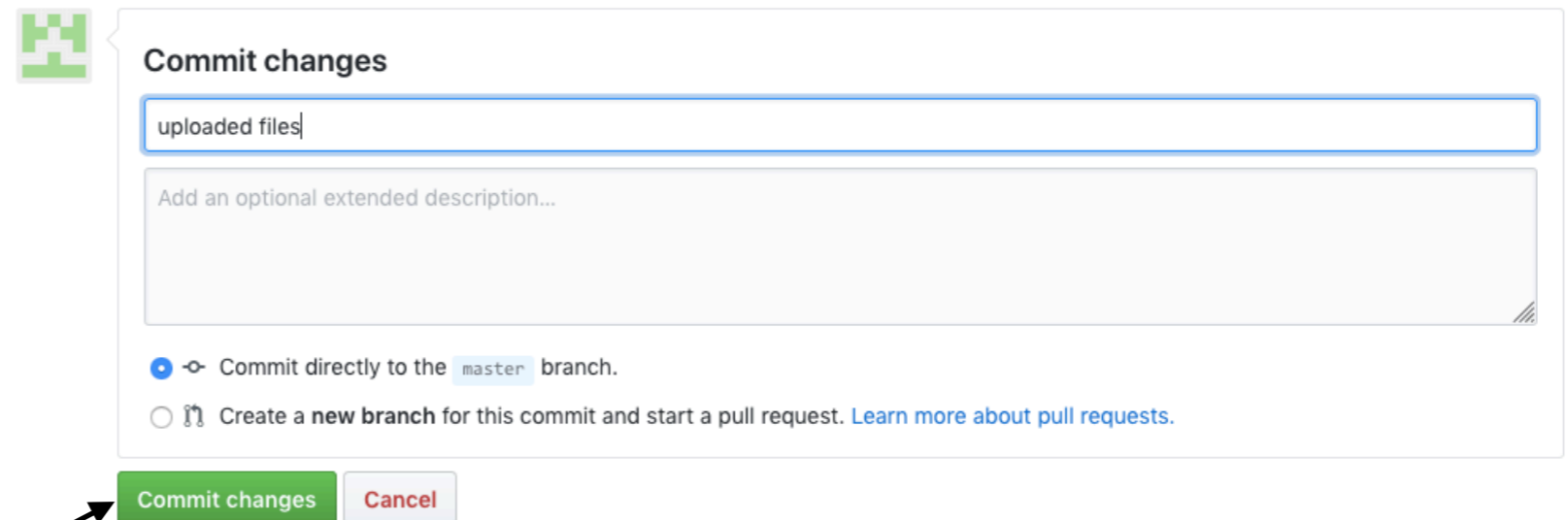


Create repository

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



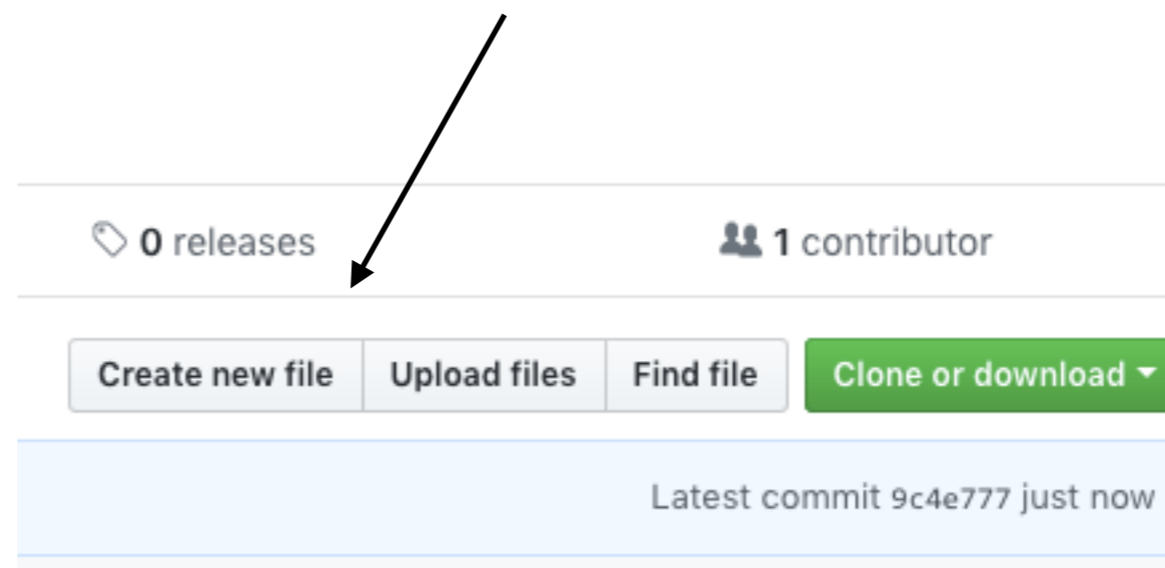
Click here to upload



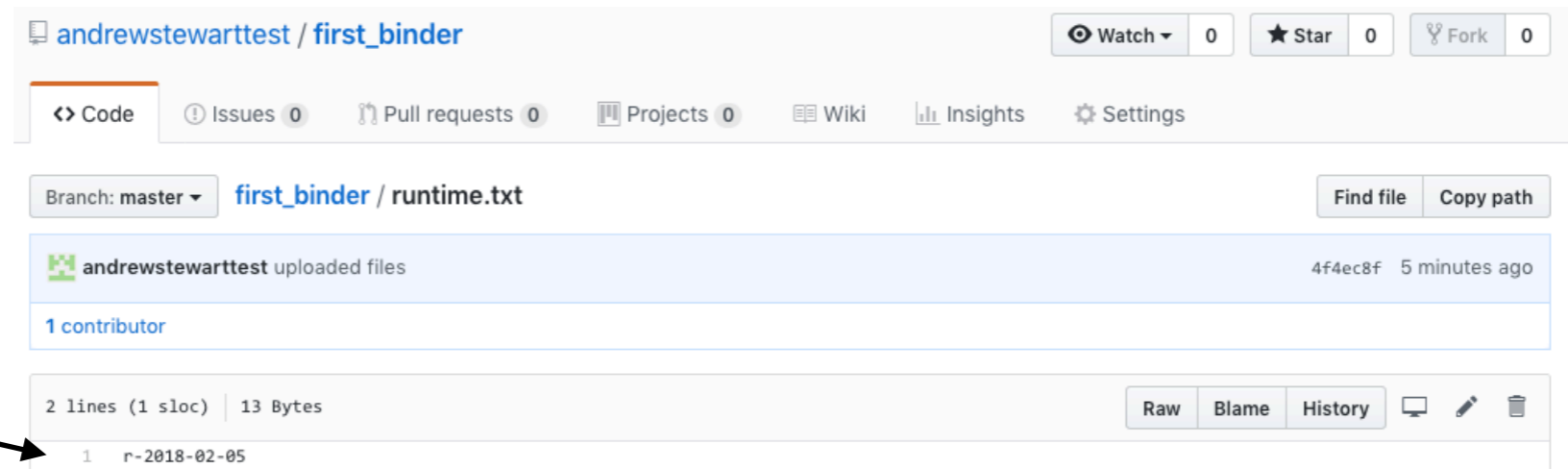
Click here to Commit

# 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”



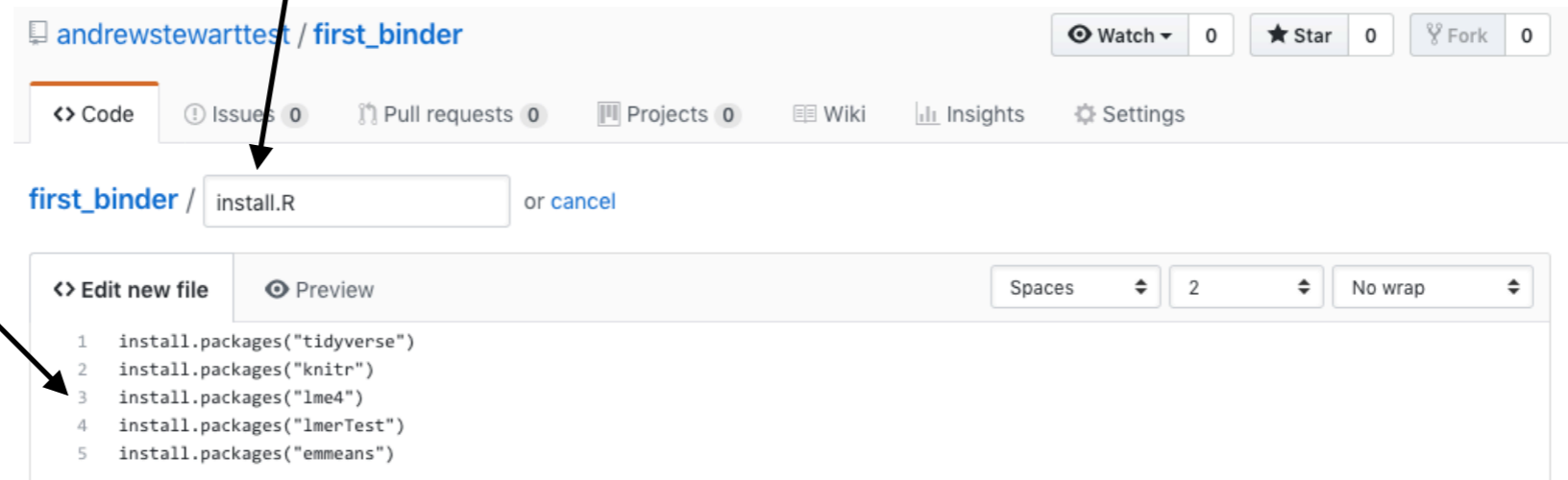
In the runtime.txt file type the date you want in the format r-YYYY-MM-DD



The screenshot shows the GitHub interface for a repository named 'first\_binder' by 'andrewstewarttest'. The file 'runtime.txt' is selected, showing its content: '1 r-2018-02-05'. The interface includes navigation tabs for Code, Issues, Pull requests, Projects, Wiki, Insights, and Settings. A commit history section shows the file was uploaded by 'andrewstewarttest' 5 minutes ago. The file details indicate it has 2 lines (1 sloc) and is 13 bytes in size.

Name your file

List your packages like this in the install.R file



The screenshot shows the GitHub interface for editing a new file named 'install.R' in the 'first\_binder' repository. The file content is as follows:

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

The editor interface includes tabs for 'Edit new file' and 'Preview', along with settings for 'Spaces' (set to 2) and 'No wrap'.

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. The form is titled "Build and launch a repository". It has three main sections: 1. "GitHub repository name or URL" with a text input field containing "https://github.com/ajstewartlang/my-first-binder" and a "GitHub" dropdown menu. 2. "Git branch, tag, or commit" with a text input field containing "Git branch, tag, or commit", and "URL to open (optional)" with a text input field containing "rstudio" and a "URL" dropdown menu. 3. A "launch" button. Below the form, there is a section titled "Copy the URL below and share your Binder with others:" with a text input field containing the URL "https://mybinder.org/v2/gh/ajstewartlang/my-first-binder/master?urlpath=rstudio" and a clipboard icon.

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.

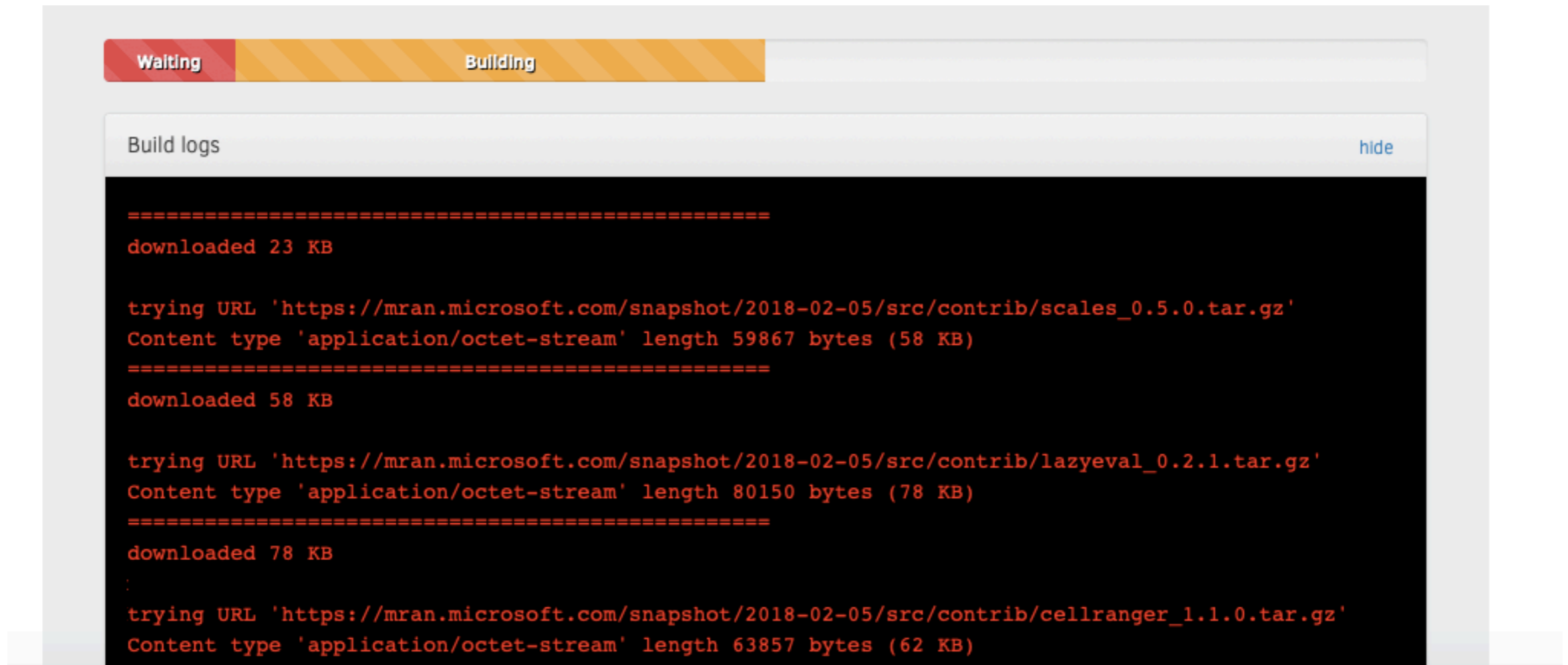
Copy the text below, then paste into your README to show a binder badge: 

```
m [![Binder](https://mybinder.org/badge_logo.svg)](https://mybinder.org/v2/gh/ajstewartlang/Binder_demo/master?urlpath=rstudio)
```

```
.rst .. image:: https://mybinder.org/badge_logo.svg
      :target: https://mybinder.org/v2/gh/ajstewartlang/Binder_demo/master?urlpath=rstudio
```

- Paste this code into your GitHub repo README.md - you'll then be able to click on the 'launch binder' button in your repository to launch the actual binder once it has been built - makes it easy for others to go from you GitHub repo to your code running in Binder.

# Once you click 'Launch'...



The screenshot shows a build progress interface. At the top, there is a progress bar with two segments: 'Waiting' (red) and 'Building' (orange). Below the progress bar is a 'Build logs' window with a 'hide' button in the top right corner. The logs window contains the following text:

```
=====
downloaded 23 KB

trying URL 'https://mran.microsoft.com/snapshot/2018-02-05/src/contrib/scales_0.5.0.tar.gz'
Content type 'application/octet-stream' length 59867 bytes (58 KB)
=====
downloaded 58 KB

trying URL 'https://mran.microsoft.com/snapshot/2018-02-05/src/contrib/lazyeval_0.2.1.tar.gz'
Content type 'application/octet-stream' length 80150 bytes (78 KB)
=====
downloaded 78 KB

trying URL 'https://mran.microsoft.com/snapshot/2018-02-05/src/contrib/cellranger_1.1.0.tar.gz'
Content type 'application/octet-stream' length 63857 bytes (62 KB)
```

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

- If Binder can find an image that you've built previously, it will simply launch that.
- If you've made changes to your GitHub repo, it will rebuild the Docker image and create a new Binder.
- Either way, once Binder launches you get the following in your browser (even on mobile devices so you can even R away on your phone)...

# And then...

The screenshot displays the RStudio web interface in a browser. The browser address bar shows the URL: `hub.mybinder.org/user/andrewstewarttest-first_binder-z4kwp6gl/rstudio/?token=`. The RStudio interface includes a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help) and a toolbar with icons for file operations and execution. The main editor window shows the following R Markdown code:

```
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
```

The console window at the bottom left shows the R startup message:

```
~/ |
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.

> |
```

The file browser on the right side of the interface shows a list of files in 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

# A few other things...

- Installing the entire Tidyverse in a Binder can take a long time - better to install only the packages you use (e.g., ggplot2, dplyr, readr etc.) - this will also ensure the packages are consistent with the date in your runtime.txt file.
- Even with just a couple of packages it can take ~15 minutes or so for your Binder to be built.
- Some R packages need system-level packages to also be installed - you can do that via an additional apt.txt file which lists those packages - this is used by apt-install to install those packages from the Ubuntu apt repository.

# A few other things...

- At the moment, you can't change the version of R that runs on Binder (currently set to 3.4.4.) so need to go down the Rocker route but be aware that you may not get the right version of the packages that you want...
- You can close your laptop if Binder is taking too long - the image and your Binder will continue to be built in the Cloud. And it's always a good excuse for another coffee...

# For Ultimate Reproducibility

- Make sure you have updated all your packages before you run your script.
- Build your Binder and specify the day you ran your analysis in the runtime.txt file
- Patience while your Binder builds...



# Slides

[http://ajstewartlang.github.io/Binder\\_slides.pdf](http://ajstewartlang.github.io/Binder_slides.pdf)

**My step-by-step guide (including an example R script if you don't have one)**

<https://hackmd.io/gO3cehAVSpuCB2EDEvepzg?view>