Binderising your repo using `holepunch`

Andrew Stewart
University of Manchester

@ajstewart_lang

Software
Sustainability
Institute

Why?

- If you've been using Binder via the repo2docker route, you might have noticed that some Binders take quite a long time to build initially - oftentimes this happens when you're wanting to install the entire tidyverse or lots of packages with dependencies..

- By writing a Dockerfile, you're able to pull a pre-built Docker (Rocker) image into Binderhub so it will launch a lot more quickly. Typically this image will include the Tidyverse packages (and others) so things don't need to be built on-the-fly.

- More about Rocker here:

  https://www.rocker-project.org

How?

- The `holepunch::` package by Karthik Ram allows you to write a Dockerfile, and build your GitHub repo from within RStudio.

`https://github.com/karthik/holepunch`

- The Dockerfile will capture the date of the last time you updated any file of your project and then pull a pre-built Rocker image associated with that date into Binderhub when you launch Binder.

- You need to initialise the local R project folder with Git version control (or clone a repo from GitHub).

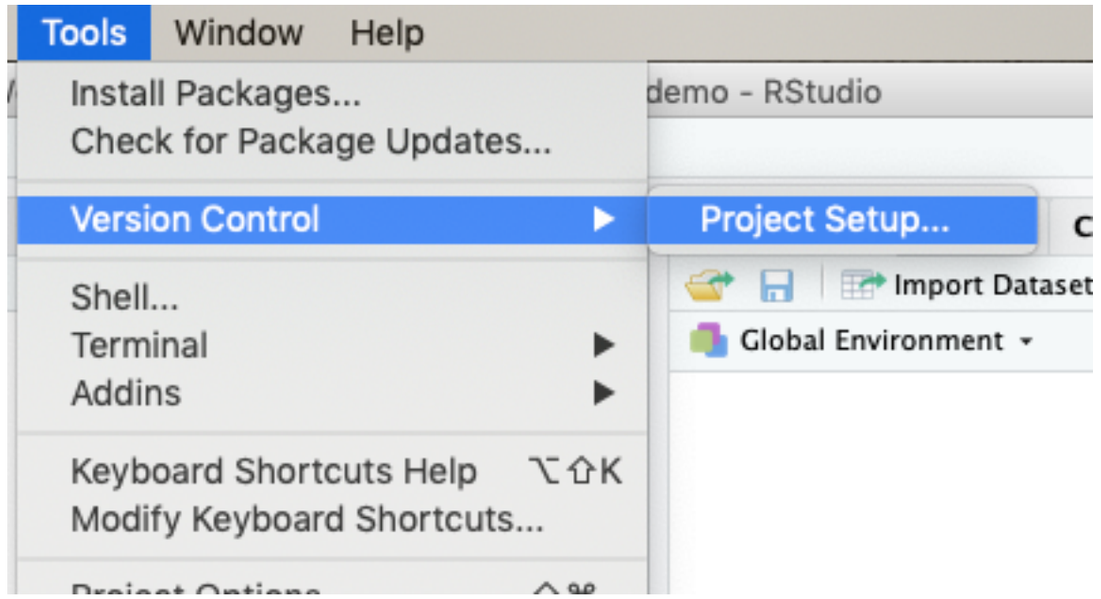
- First install the latest version of "holepunch" from GitHub - you may be prompted to update some other packages - please do so.

```
> remotes::install_github("karthik/holepunch")
```

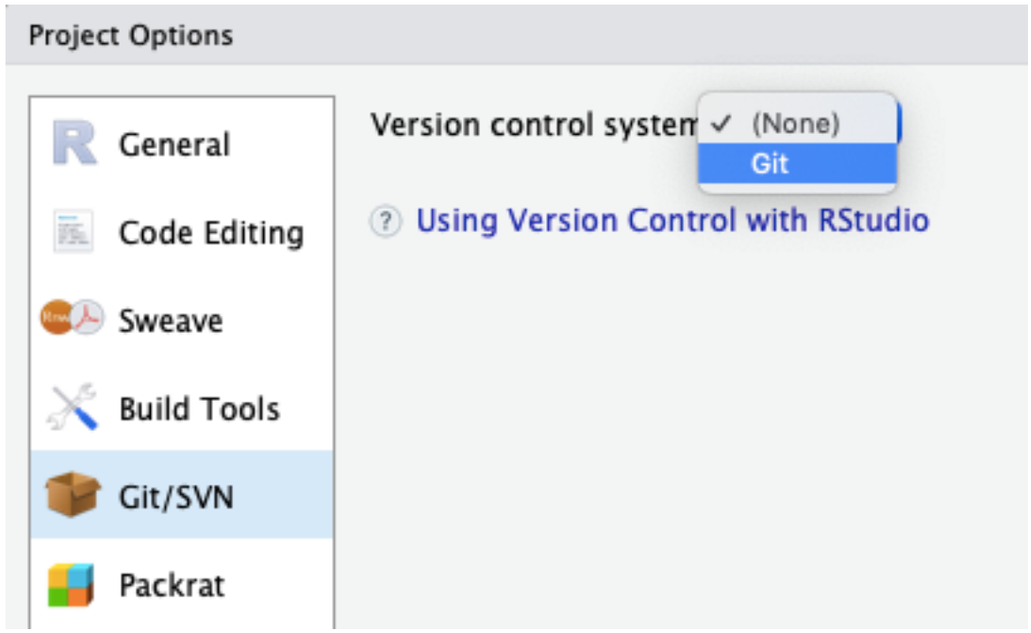
```
> library(holepunch)
```

- You can either clone a pre-existing repo from GitHub, or create a new R Project and turn that folder into a git version controlled repo - in which case...

First we need to ensure our folder associated with a project is a repository with git version control.

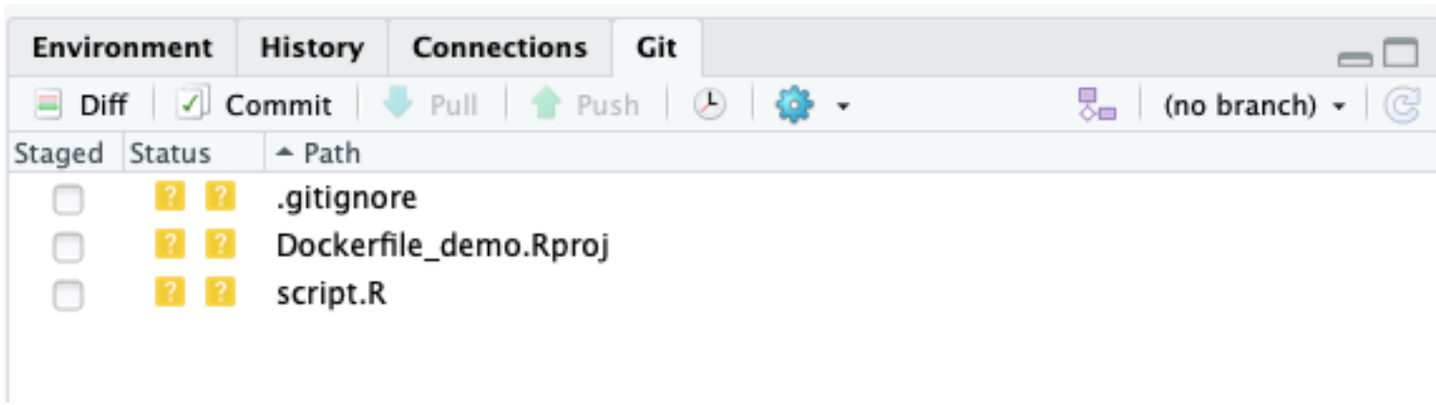


Select Git - you will need to restart your R session at this point.

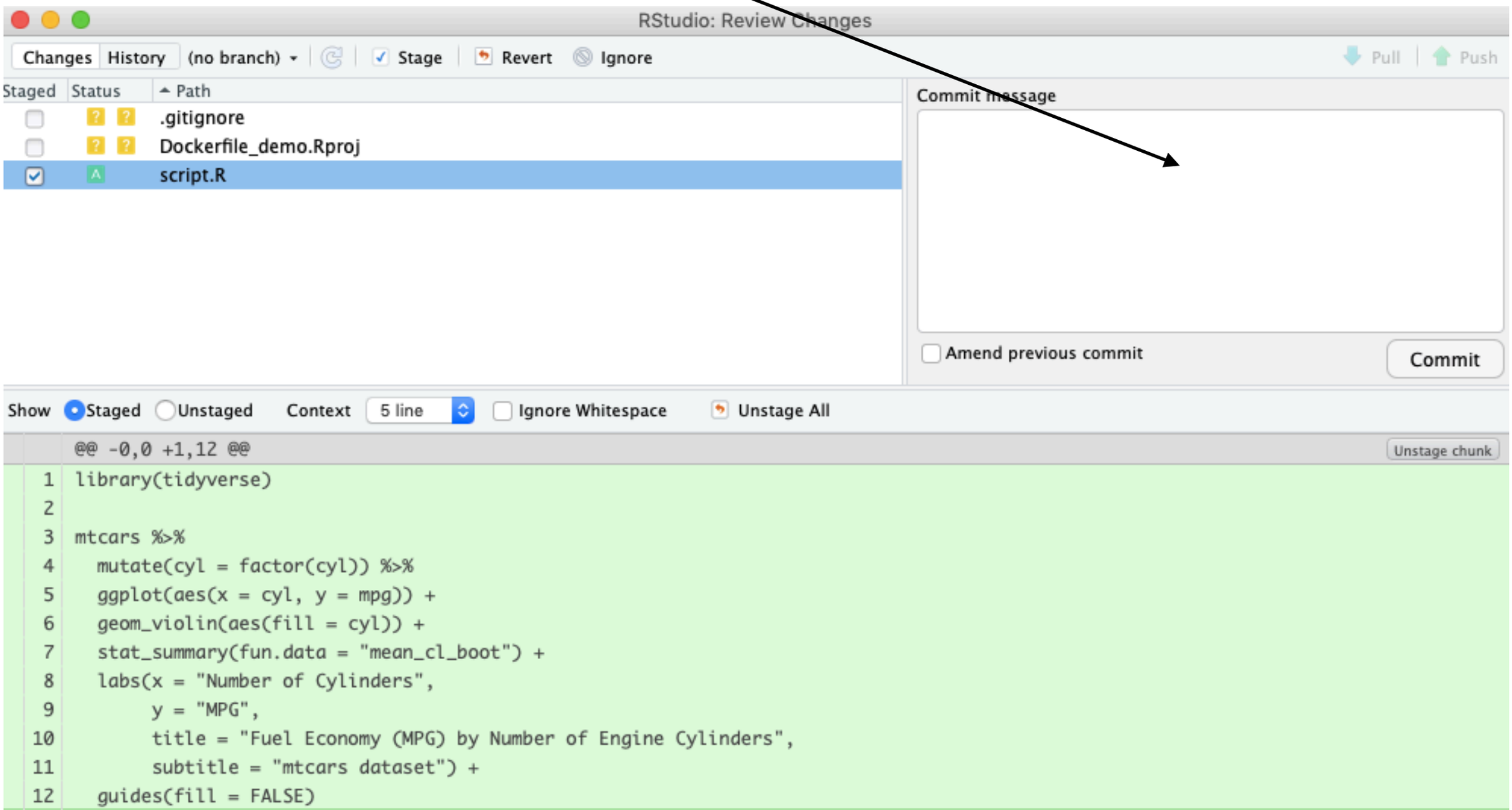


When you restart, you'll see you now have a new Git tab in your Environment window.

If you click on the Tab you'll see the contents of your folder.



In this example, I've selected the script.R file, saved it (you can't Commit without saving first), and then clicked on Commit - the following window now appears. Write a meaningful Commit message in here and the click on the 'Commit' button underneath.



- Once you have a git repository set up locally, push it to GitHub either within RStudio or via GitHub desktop (or the command line).

- Alternatively, you can set up a repo on GitHub and then clone it locally - but you'll still need to push changes to GitHub.

- You can then write your Dockerile via the console in RStudio.

Step 1 - write a Dockerfile

```
> write_dockerfile(maintainer = "your_name",
r_date = "2019-06-27")
```

- If you leave out the date, `holepunch` will create a Dockerfile associated with the date you last changed your repo. It uses the version of R and packages on MRAN associated with the date you specify (or the last change date if you don't specify an actual date).

Step 2 - generate a binder badge

```
> generate_badge()
```

- this will generate the code you need to paste in your repo README that will launch Binder upon clicking.

Step 3 - build your binder

```
> build_binder()
```

- will start building your Binder in the background - this will still be much quicker that building from scratch as the Dockerfile will pull a Rocker image and associated R packages for the date you specified during `write_dockerfile()`

Any caveats?

- `holepunch::` is very much still in development but Karthik responds super quickly to issues, enhancement suggestions, and bug reports - and it will be on CRAN (and therefore more stable) sooner rather than later...

- Great `rstudio::conf 2019` video of Karthik talking about reproducibility in general and `holepunch::`

https://resources.rstudio.com/rstudio-conf-2019/a-guide-to-modern-reproducible-data-science-with-r