



Department  
for Education

# **DfE Statistics Development Team Workshops**

**Using git and GitHub (building a R-Shiny  
dashboard)**

# Contents

<b>Introduction</b>	<b>3</b>
GitHub versus Dev Ops	3
<b>Pre-workshop requirements</b>	<b>3</b>
<b>Setting up the repository</b>	<b>6</b>
Creating a new repository on GitHub	6
Cloning the repository to your local machine	7
Controlling packages with renv	10
Running the dashboard template locally	11
Summary	12
<b>Basics of git</b>	<b>13</b>
The git log	13
Adding, committing and pushing	13
Pulling from the remote repository	14
Summary of git basics	15
<b>Working collaboratively with git</b>	<b>16</b>
Branches and splitting tasks	16
Summary	18

## Introduction

We've prepared this walkthrough guide for statistics publication teams as an introduction on how to work collaboratively with git, using the creation of a data dashboard as a relevant context. The guide is intended to be step-by-step, building up from the very basics. The plan is to work through this in groups of 3 or so and with access to experienced git users to call on for support. If it starts too basic for your level, then just go through at your own/your group's pace as you see fit. By no means can we cover everything in this walkthrough, so please see it as a prompt to ask follow-up questions as you're working through on anything related to git, GitHub and Dev Ops.

For simplicity of access and accounts, we're focussing on GitHub rather than Dev Ops here, but much of the material is transferable.

## GitHub versus Dev Ops

GitHub and Dev Ops effectively provide the same service in terms of creating software via a git repository: they both act as the host for the remote repository, whilst offering important tools to manage bugs and issues, tasks, merging branches, deploying applications and so on.

Dev Ops is part of the Microsoft Azure platform and uses private DfE servers. This can allow you to connect or deploy your repository into wider Azure services. This includes SQL databases that you might already be storing data on as well as the DfE's implementation of rsconnect on DfE internal servers, which allows deployment of shiny apps for internal DfE use.

GitHub is hosted on external servers and therefore is more appropriate for making your code or application available for public access and use. For example, from a GitHub repository, you can deploy an R Shiny dashboard to shinyapps.io where members of the public may view and interact with your published data.

## Pre-workshop requirements

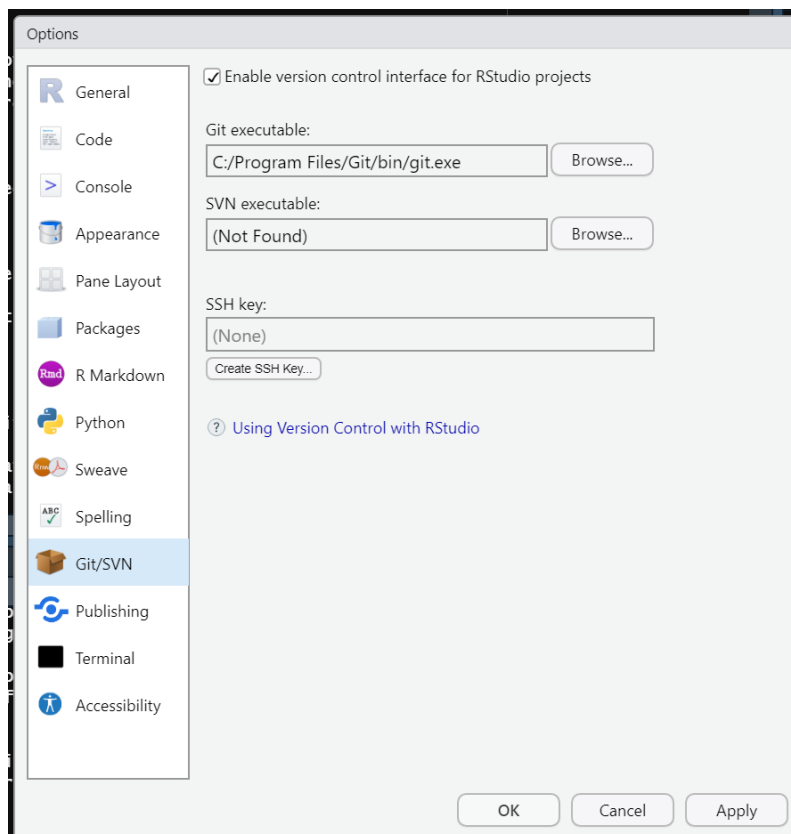
First of all, make sure to bring your laptop. This is going to be interactive and require you to do some coding.

Preferably before coming along, you'll need to go through the following list of things you'll need to make sure are set up on your DfE laptop:

- Create a GitHub account: <https://github.com/join>;
- Install git on your laptop: <https://git-scm.com/downloads>;
- Install R-Studio on your machine: Download **R for Windows (x64)** and **RStudio** from the Software Centre on your DfE laptop.

You'll also need to make sure that git is set up in the git/SVN pane of global options in R-Studio (found in the Tools drop down menu). Make sure the path to your git executable is entered in the git path box and git should automatically be integrated with R-Studio.

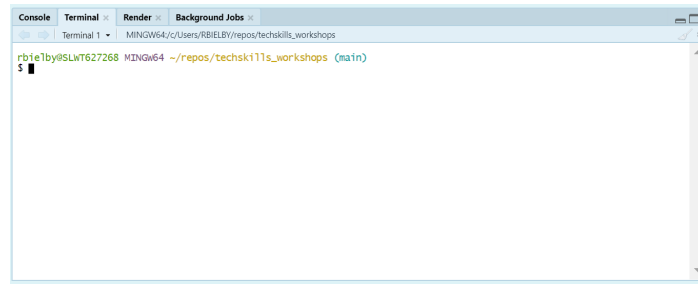
Figure 1: Enter the path to your git executable in the git path option box



Once you open a repository, you'll get an extra panel, named 'git', in the top right pane of R-Studio and you'll also be able to use git in the 'Terminal' tab at the bottom left (in the same place as the R console).

A useful thing here if you want to use git commands in the terminal is to switch the terminal from the default Windows Command Prompt to git BASH. You can do this in the Terminal tab of R-Studio's global options - just select git BASH from the 'New terminal opens with' pull down menu. Click apply and then select the Terminal tab (next to the Console tab), click 'Terminal 1' and then select 'New terminal' from the drop down menu.

Figure 2: The 'git BASH' terminal in R-Studio



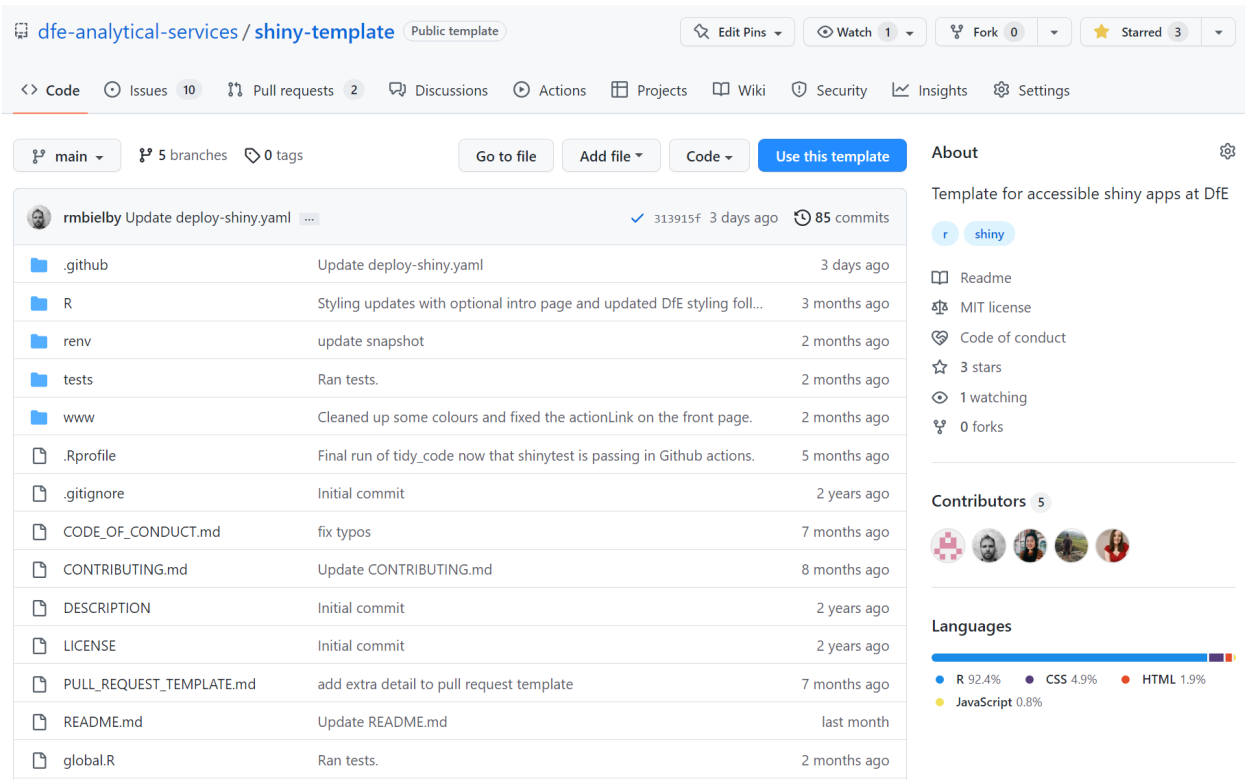
You should see something similar to the terminal screenshot.

# Setting up the repository

## Creating a new repository on GitHub

At this point, we're ready to create a new repository. The context of this exercise is to create a dashboard, so let's get a head start on that by using the DfE R-Shiny template.

Figure 3: Click the use this template button to create the new repository



You can access the template here:

<https://github.com/dfе-analytical-services/shiny-template>

On that page, you'll see the main repository page. This contains a menu bar to navigate the range of GitHub features (e.g. Code, Issues, Pull Requests, Discussions, Actions and others); shortcuts to access different branches within the repository; some top-level summary information on the repository; a listing of the files and folders in the repository's root directory; and a markdown render of the repository's Readme file if one exists.

In the case of our template, you'll also see a button saying use this repo as a template. At this point, one (and only one!) of your group should click that button, which will take you to the create repository page. Here you have the option to create a copy of the template in

your own GitHub area as shown below. Give the new repository a name and a description and then click “Create repository from template”.

Figure 4: Put in a name for the new repo and a quick description and then click on Create repository from template

Create a new repository from shiny-template

The new repository will start with the same files and folders as [dfe-analytical-services/shiny-template](#).

Owner \* rmzielby / Repository name \* workshop\_gitdemoapp ✓

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

Description (optional)

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

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

☐ **Include all branches**  
Copy all branches from dfe-analytical-services/shiny-template and not just main.

You are creating a public repository in your personal account.

[Create repository from template](#)

[Terms](#) [Privacy](#) [Security](#) [Status](#) [Docs](#) [Contact GitHub](#) [Pricing](#) [API](#) [Training](#) [Blog](#) [About](#)

© 2022 GitHub, Inc.

One of you at least should now have their own repository produced from this starting template project. We need everyone in your team to be able to work on this project however, so now you’ll need to give access to your other team members. To do so, navigate to settings (at the far right of the menu bar on your repository page) and =

Now you’ll need to add your other team members as collaborators in order to allow them to contribute to the code development in this repository.

## Cloning the repository to your local machine

Cloning the repository refers to creating a copy of the remote repository (i.e. the copy on GitHub or Dev Ops) on the disk on your local machine (i.e. your DfE laptop). For an R project, there are two basic options to choose from for doing this:

- using the R-Studio new project wizard, or

- using `git BASH`.

We'd recommend trying the different options across your working group.

## Cloning in `git BASH`

You can open up a `git BASH` terminal, by typing `git BASH` in the Windows search bar and select `git BASH` when it comes up. With a terminal, you can interact with it just by typing, similar to working in the R console in RStudio. First let's make a directory in which to store our repositories:

```
mkdir repos
```

We can then move into the directory we just created using:

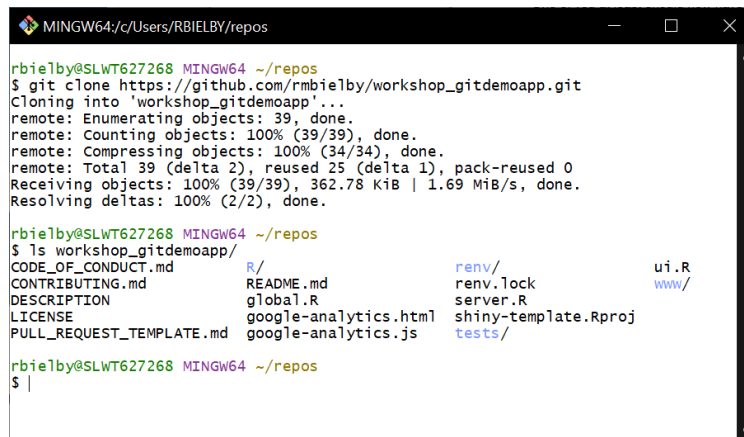
```
cd repos
```

Now grab the repo url and replace `<repo_url>` in the next command with the actual url:

```
git clone <repo_url>
```

You should get some messages letting you know git is connecting to the server and cloning the repository and it should look something like the figure below.

Figure 5: Cloning a repository in `git BASH`



```
MINGW64~/c/Users/RBIEBY/repos
rbieby@SLWT627268 MINGW64 ~/repos
$ git clone https://github.com/rmbielby/workshop_gitdemoapp.git
Cloning into 'workshop_gitdemoapp'...
remote: Enumerating objects: 39, done.
remote: Counting objects: 100% (39/39), done.
remote: Compressing objects: 100% (34/34), done.
remote: Total 39 (delta 2), reused 25 (delta 1), pack-reused 0
Receiving objects: 100% (39/39), 362.78 KiB | 1.69 MiB/s, done.
Resolving deltas: 100% (2/2), done.

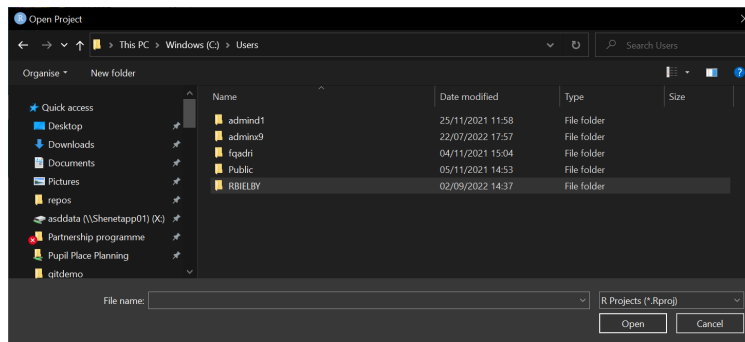
rbieby@SLWT627268 MINGW64 ~/repos
$ ls workshop_gitdemoapp/
CODE_OF_CONDUCT.md  README.md      renv/          ui.R
CONTRIBUTING.md    global.R       renv.lock      www/
DESCRIPTION         google-analytics.html  server.R
LICENSE             google-analytics.js  shiny-template.Rproj
PULL_REQUEST_TEMPLATE.md
```

If all went well, you'll now have a complete copy of the repository on your laptop. To open the repository in RStudio, start up RStudio and select Open project. In the file explorer window that opens up, type `C:\Users\` and hit enter (see the screenshot below) and then open up your home folder.

Then navigate into `repos` and the repository folder. The full path should be something along the lines of:



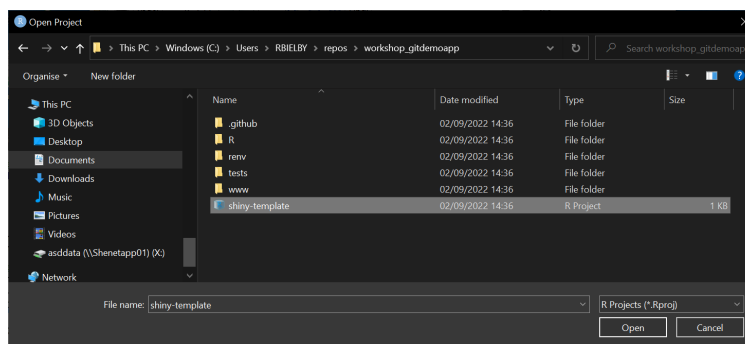
Figure 6: Open a cloned project in RStudio



This PC > Windows (C:) > Users > <USERNAME> > repos > <REPONAME>

Select the R project file and select open.

Figure 7: Open a cloned project in RStudio



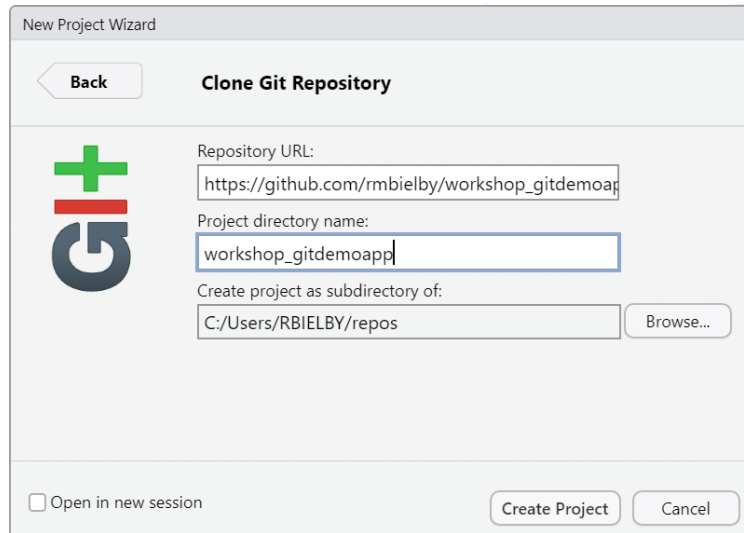
## Cloning using the RStudio wizard

If that looks like a bit too much text based effort, RStudio offers a way to clone a repository with it's New project wizard. To do this navigate the menu bar to **File > New Project...**, select **Version Control** and then Git. This opens up a dialogue box to enter the repository url and select where to save it. As with the git BASH version, copy and paste your remote repo URL here and set a directory where you want it saved on your laptop.

## A note on local repository clones and OneDrive

Note that saving a repository within your OneDrive folder structure can cause some awkward issues. If you use git to perform version control on a repository saved within a OneDrive folder, you may start receiving warning messages that large numbers of files have been removed from OneDrive. In addition, it can put a heavy burden on your internet connection as OneDrive tries to keep up with

Figure 8: Clone a project using the RStudio git wizard



changes to the files managed by git. Best practice therefore is to store your repositories somewhere outside your OneDrive file structure. We recommend creating a `repos` directory within your base User directory (i.e. `C:\Users\<USERNAME>\repos\`. Windows sometimes tries to make it awkward for you to navigate to places on your laptop outside of the OneDrive folders, so a useful tip is to add your `repos` folder to your Quick access list in File Explorer.

## Controlling packages with `renv`

Now that you've each got a clone of the repository, it's useful to understand a little bit about environment control. Using R as the specific example, any app or pipeline that you develop will have packages that it depends on. If someone wants to use your code, they need to know what packages are involved, whilst it can also be helpful for them to be able to use the same versions of those packages in order to recreate exactly what you intended the code to be able to do.

To manage this, we can use an R package called `renv`. This manages the R environment for you, helping you keep track of your repositories required R packages and their versions. In a brand new empty project, you'd initiate `renv` with the command `renv::init()`. In this case, the template already has `renv` initiated. So instead we can tell `renv` to check the necessary packages and install them for us. You can do this by entering the following command in the R console:

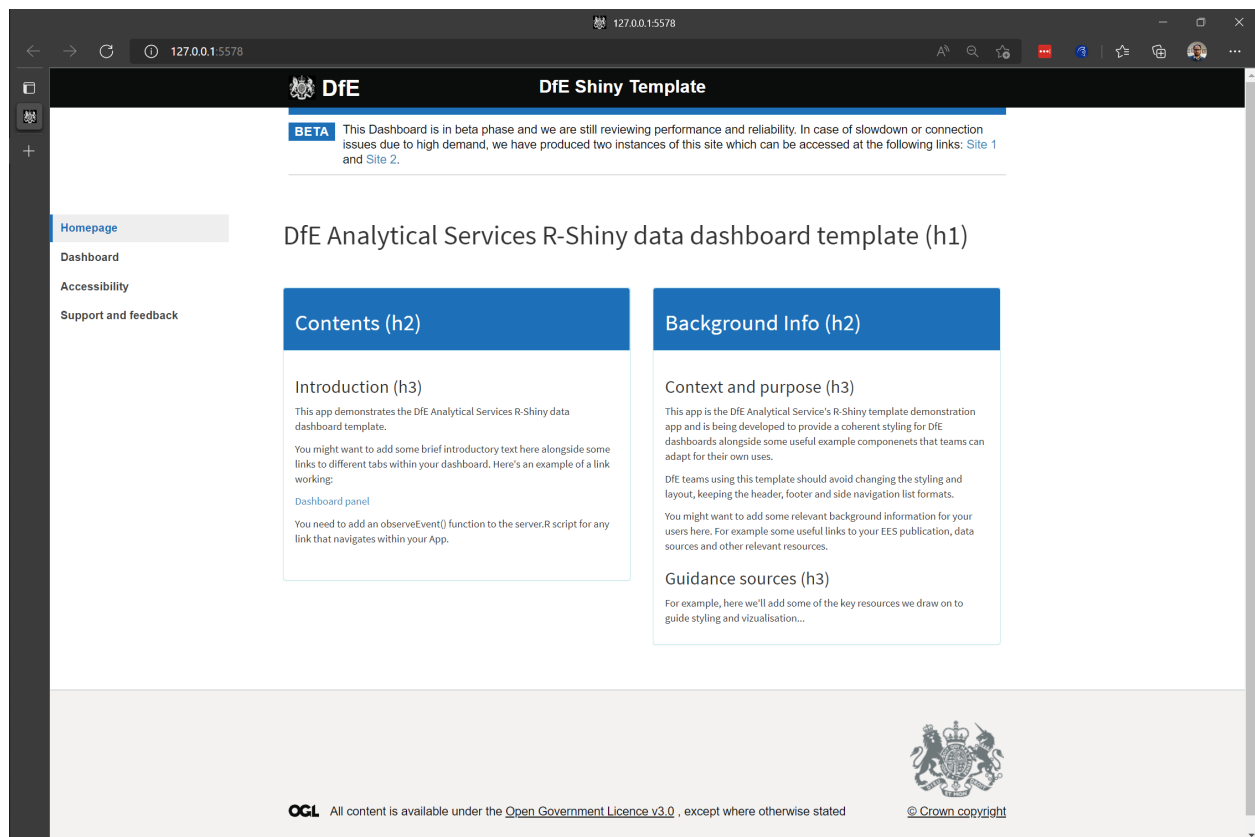
```
renv::restore()
```

This should go through `renv`'s list of the required packages and install them all in one go (note that you'll need to be connected to the internet for it to retrieve any packages that aren't already installed). Once you've done this, the contents of the repository should run as intended. If at some point you install an extra package for the repository, you can add this to `renv`'s package list by entering the command `renv::snapshot()` into the R console.

## Running the dashboard template locally

Once you've cloned the package and run `renv::restore`, you should then be able to run the Shiny dashboard template. To do this, either a) enter `shiny::runApp()` in the R console or b) open up the file **ui.R** and then click **run App** to the top right of the editor pane. The dashboard should then load either in an RStudio window or in your default web browser (depending on your chosen settings in RStudio).

Figure 9: The DfE R-Shiny template loaded locally from RStudio



Although this workshop is primarily a chance to practice using git, you'll need a little understanding of the components of an R-Shiny app as well. A single app generally

consists of 3 key scripts: *global.R*, *ui.R* and *server.R*. Each of these fulfills a specific purpose as follows:

- *global.R*: contains variables and functions required by the rest of the app (e.g. reading in data, connecting to a database);
- *ui.R*: contains code to create the app's user interface (e.g. arranging the layout, placement of charts, tables and text);
- *server.R*: contains the functions that make the app responsive (e.g. rendering tables and charts and updating dynamic text and input fields).

We'll explain more as needed, but for now that should give you the basic overview.

## Summary

In this section, we've covered creating a repository via a template on GitHub, cloning it to your local drive (using both the BASH terminal and the RStudio wizard) and getting the necessary packages installed in one go with `renv::restore()` via the R console.

In the next section, we'll cover some of the basics of using git to log changes to your code and sync them between the remote repo and local copies.

## Basics of git

We'll now take a look at updating repositories with your working. To do so, we'll follow some example first steps you might take with developing the template dashboard app to use your own data. The steps we'll follow will be to create a new branch, add the data in, commit it to the github log and then push it to the remote (GitHub) repository.

## The git log

In order to move quickly between different versions of files and code, git is built around indexing and a log file that track the changes in a repository. To view the log of any repository, we can simply go into that repository and run the command `git log` from the BASH terminal. For example, if you run it in the repository you've created from the template, you'll get something along the lines of the following:

```
rbielby@SLWT627268 MINGW64 ~/repos/workshop_gitdemoapp (main)
$ git log
commit 467df8800a108c0e1c98cb8db5ff0f377db8ffbb (HEAD -> main, origin/main, origin/HEAD)
Author: Rich Bielby <richbielby@e.email>
Date:   Fri Aug 19 17:27:36 2022 +0100

    Initial commit
```

The log shows all “commits” that have been made to the repository. We'll go into making commits in the next section.

## Adding, committing and pushing

First of all, one of your group should make a folder called **data** in your local repository. To do this, you can either go to the BASH terminal and type the command `mkdir data` (making sure you're in your local repositories root directory) or click the add folder button in the Files tab in the bottom right panel of RStudio.

For some example data, we'll use a file from a publication on Explore Education Statistics. Go to [the progression to higher education or training data catalogue](#) and download the **Progression to higher education and training - Local authority level (csv, 4 Mb)** file and extract the data csv file (*l4\_tidy\_2020\_all\_la.csv*) into the data folder you just created.

Now to add this to the git log. Run the following commands:

```
git add .
```

This searches the repo for any files that have been modified since the last commit and creates a log of the changes.

```
git commit -m "Added data file into repository."
```

This adds an entry on to the log, updating it with the file changes that you've just made. Note that the text after the `-m` is a comment used to describe the changes to make it easier for someone looking back from the log to see what changes have happened. Those are the two key steps for tracking changes to the files and folders in your repository. Say we decide we're not happy with the data filename, maybe we think the filename should be more informative for users. We'd now change the filename as normal, and then re-run the commands `git add .` and `git commit -m "Made data filename more informative."` (updating the commit comment to describe the changes that we've made). If we now run `git log` again, we get something along the lines of:

```
rbielby@SLWT627268 MINGW64 ~/repos/workshop_gitdemoapp (main)
$ git log
commit a57f66898df18eb43a685a414b9ab805301a6c01 (HEAD -> main)
Author: Rich Bielby <richard.bielby@education.gov.uk>
Date: Tue Sep 6 18:30:30 2022 +0100

    Renamed data file.

commit 8d4ef984664a23d64fe0fbfa296729af0c23094c
Author: Rich Bielby <richard.bielby@education.gov.uk>
Date: Tue Sep 6 18:28:44 2022 +0100

    Added data file into repository.

commit 467df8800a108c0e1c98cb8db5ff0f377db8ffbb (origin/main, origin/HEAD)
Author: Rich Bielby <richbielby@e.email>
Date: Fri Aug 19 17:27:36 2022 +0100

    Initial commit
```

Here we can see, in reverse order, the commits that have been made, who made them, when they made them, and the messages that have been recorded with them.

Finally, it's important to note that what we've done so far is only being applied to the local copy of the repository (i.e. the copy on your laptop). To apply your changes to the remote repository (i.e. on GitHub or Dev Ops), you need to "push" the changes. This can be done a couple of different ways: a) in the terminal type `git push` or b) on the toolbar in the git tab in RStudio press the green up button! Once you've done this, open up a browser and go to your remote repository on GitHub and you should now see the data file stored there.

## Pulling from the remote repository

Now that you've made changes, the rest of your team can update their own local copies of the repository with your updates by pulling from the remote. Similarly to pushing, they can do this by either a) typing `git pull` in the BASH terminal or b) pressing the down arrow in the toolbar of the git tab in RStudio.

## Summary of git basics

We've quickly tried out a quick cycle of adding and committing, which is used to log changes into the local repository and then we've pushed and pulled to and from the remote repository and local copies on different laptops. The table below gives a summary of the relevant commands in the BASH terminal and the corresponding buttons in the RStudio git panel.

Process	git BASH	RStudio git panel
Add	<code>git add .</code>	Stage using tickbox next to each modified file.
Commit	<code>git commit -m "Commit message."</code>	"Commit" button in toolbar.
Push	<code>git push</code>	Green up arrow in toolbar.
Pull	<code>git pull</code>	Blue down arrow in toolbar.
View the log	<code>git log</code>	Clock icon in toolbar.

## Working collaboratively with git

Git only really makes proper sense once multiple people start working on a project collaboratively. Solo working, git is useful for version control and syncing your work to a remote repository site like Dev Ops and GitHub, but doesn't necessarily offer all that much more beyond that. Once we start working collaboratively however, the benefits of using git (alongside GitHub or Dev Ops) become more apparent. We'll now look further into this with some worked examples.

## Branches and splitting tasks

### Task management

One useful management tool that we can use from GutHub is the Issues tab. Here we can create individual tasks, assign them to team member and then create new **branches** from those tasks. You can think of **branches** as self contained copies of the repository that complementary or conflicting differences with all other **branches** in the repository. These allow you to work on different multiple tasks on your code independently of any other changes you might be making. Bringing these different **branches** or tasks together is then managed using **merges** or **pull requests** (PRs).

We'll demonstrate this by performing 3 related tasks on 3 different branches. Split the following 3 tasks between your group and each follow the relevant instructions:

- Reading in the data
- Creating a chart
- Creating a table

Once you've decided who's doing what, each of you should jump to the relevant section below. And remember that you're not working in independent silos here, what you do can impact what other people are doing so communication needs to happen along the way.

### Task 1a - Reading in the data

This task underpins the other two. You need to read in the data file that we've already downloaded. Start by creating a new branch: go to the BASH terminal, and type:

```
git checkout -b featReadData
```

This creates and moves you into a new branch called featReadData. Anything you do here



has no affect on the contents of the main branch until you explicitly tell git to merge your changes. Note that this only creates the branch in your local copy of the repository, for it to be added to the remote, you'll need to push:

```
git push
```

Now to write the code in this branch that's going to read in the data. We'll read it in to a data frame and the first task is to choose a name for the data frame and let the rest of the team know, as they'll need that info (in the example below, we're calling it `dfProgressHE`). If you've got the R expertise to do this already, then feel free to go for it, otherwise just follow the following steps below.

Open up the script `R/read_data.R` and add the following code:

```
funcReadData <- function(){  
  read.csv("data/progress_he_la_2020.csv", stringsasfactors=FALSE)  
}
```

You've made a substantive change, so let's run a quick add, commit and push in the BASH terminal:

```
git add .  
git commit "Added function to read data in data."  
git push
```

We've made a function to read in the data, but we're not running that function anywhere. So now open up *global.R* and add the following line at the end of the script:

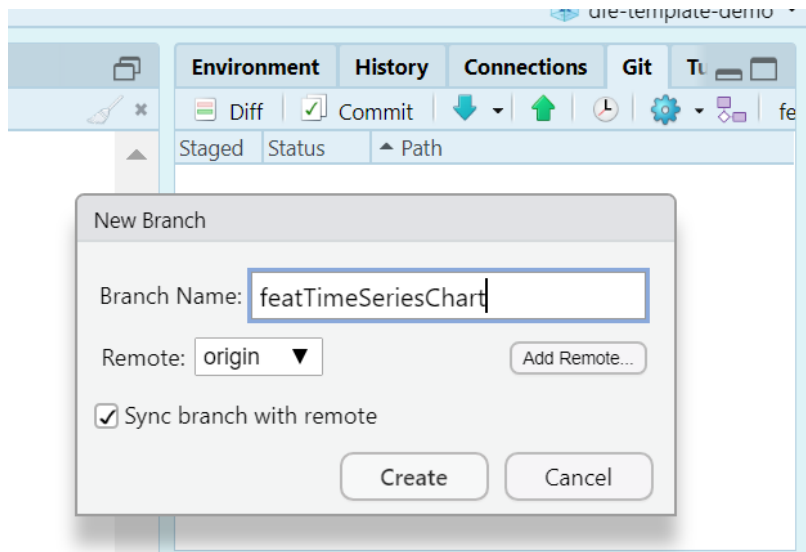
```
dfProgressHE <- featReadData()
```

Give it a test, by sourcing *global.R* and checking the contents of `dfProgressHE` (e.g. by typing `dfProgressHE` in the R console). Once you're happy, then run another add/commit/push cycle and flag to your team that you've finished your task and the data read in is ready to use.

## Task 1b - Creating a chart

There's already an example chart in the template, so let's re-wire that as the basis for our demo chart here. Let's start by creating a new branch in RStudio. With your repo open (and pulled with the latest changes) in RStudio, click the new branch button in the git panel toolbar. This will open a dialogue box as shown here in which you can enter a new branch name. Enter a name, make sure syncing to the remote is selected and press *Create*.

Figure 10: The new branch dialogue box in RStudio is opened using the purple new branch icon in the git toolbar.



**NOTE**

When your team mate from task 1 has completed their task, you'll want to merge their changes into your own branch, so that you have access to the data that's being read in:

```
git merge featReadData
```

**Task 1c - Creating a table**

**Pull requests (PRs) and merging branches**

**Summary**

Process	git BASH	RStudio git panel
Create new branch	<code>git checkout -b branch_name</code>	Click the new branch symbol

Process	git BASH	RStudio git panel
Switch to existing branch	<code>git checkout branch_name</code>	
Merge	<code>git merge branch_name</code>	Not available in RStudio - merge by creating a pull request in GitHub/Dev Ops



# Department for Education

© Crown copyright 2022

This publication (not including logos) is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

To view this licence:

visit [www.nationalarchives.gov.uk/doc/open-government-licence/version/3](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3)

email [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

write to Information Policy Team, The National Archives, Kew, London, TW9 4DU

About this publication:

enquiries [www.education.gov.uk/contactus](http://www.education.gov.uk/contactus)

download [www.gov.uk/government/publications](http://www.gov.uk/government/publications)



Follow us on Twitter:  
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook: [face-  
book.com/educationgovuk](https://facebook.com/educationgovuk)