Introduction to Colabs, Binder and R

Mark Bell

28 June 2022

THE NATIONAL ARCHIVES

Opening in Google Colabs

There are 3 notebooks which are each linked to on the following page:

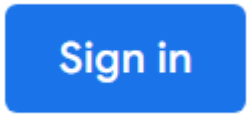
https://github.com/nationalarchives/I-CeM-workshop-2022/blob/main/NotebookLinks.md

Notebook 1: Loading and Viewing I-CeM records 

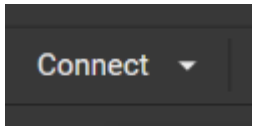
Click the “Open in Colab” button to open a notebook (Hint: right click – open in new tab, to keep Links document open)

THE NATIONAL ARCHIVES

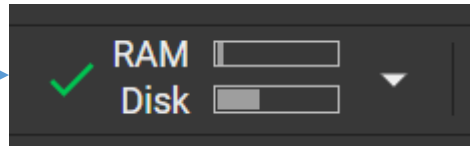
Getting started with Google Colab

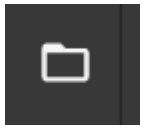


Sign in to your Google account, if not already

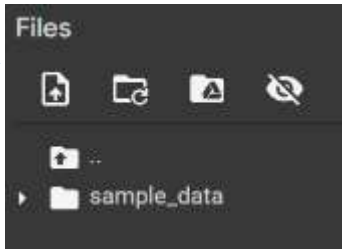


Click the Connect button → Which turns into this

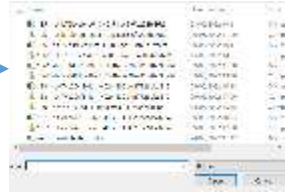


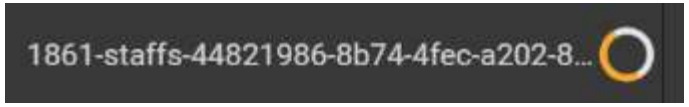


Click the Files button → Which opens this



Click Upload button → Choose I-CeM download file

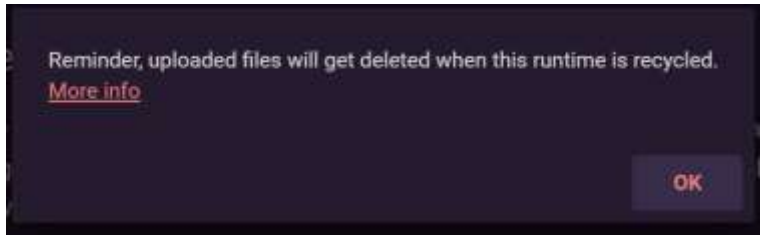




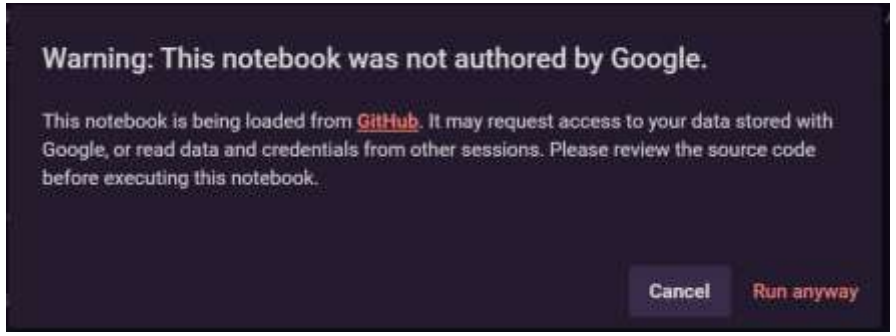
Wait for this to disappear

THE NATIONAL ARCHIVES

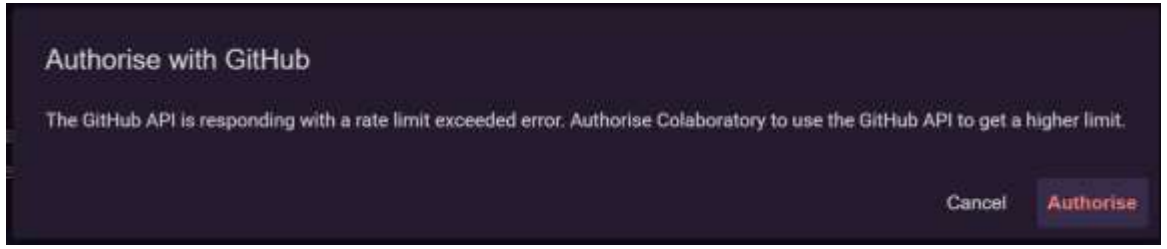
Google Colab messages



Click OK



Click Run anyway



Click Cancel

THE NATIONAL ARCHIVES

Opening in RStudio on MyBinder

The notebooks can also be hosted on the MyBinder service. This link will open the service in Rstudio mode:

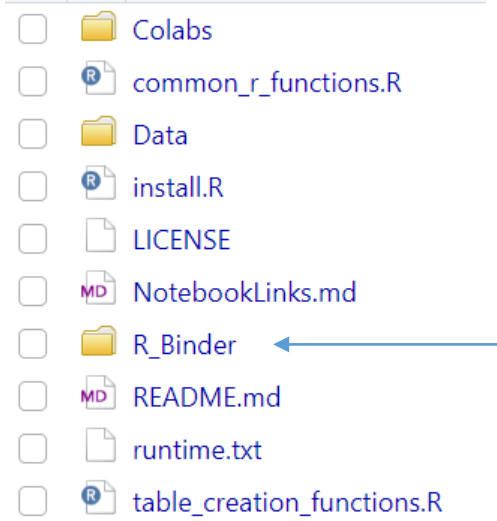
RStudio hosted on Binder

launch binder

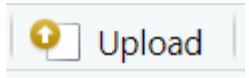
Wait for a cloud machine to start up:

THE NATIONAL ARCHIVES

Getting started with RStudio



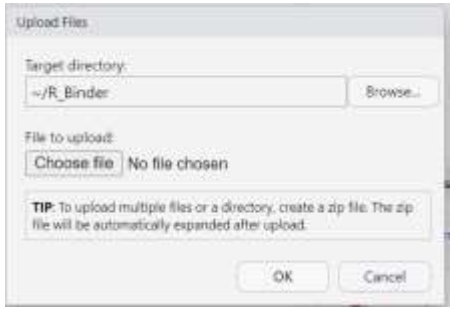
Click on R_Binder to open that folder



Click to upload a data file

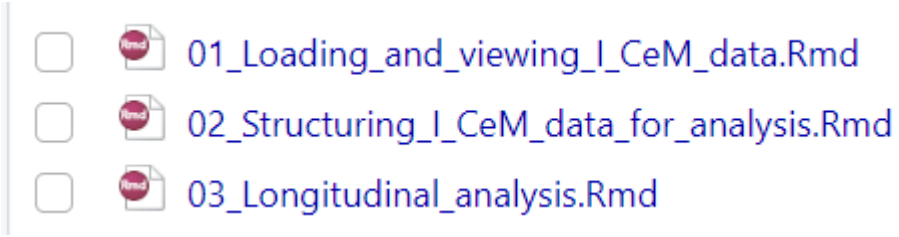
Click choose file to browse to I-CeM file

Then click OK



THE NATIONAL ARCHIVES

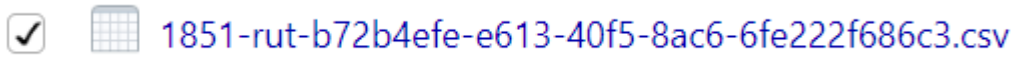
Initialising a notebook

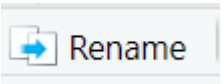


Click on Notebook name to open notebook

Add the I-CeM download file where indicated – 1 & 2 only
(instructions below for copying filename)

```
Replace the text INSERT FILE NAME HERE with the name of your csv file
```{r}
SOURCE_DATA_FILE <- "INSERT FILE NAME HERE"
```
```

Select data file 

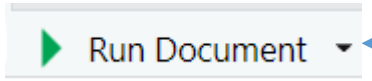
and 

Is easiest way to copy file name

Don't forget the .csv part, and to press Cancel after copying

THE NATIONAL ARCHIVES

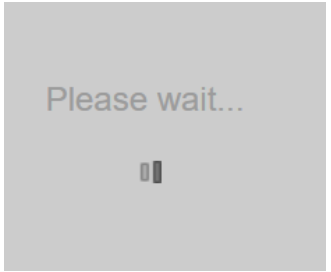
Running a notebook



If name starts ‘Shiny’: Click to run the notebook using specified data file

If name starts ‘Static’: Click to run the notebook using specified data file

New window opens, be patient while it loads.



THE NATIONAL ARCHIVES

A short guide to dplyr

dplyr is a package of functions for processing tabular data

dplyr works as a pipeline of functions which can be chained ad infinitum

Anatomy of a dplyr command:

"Output data" <- "Input data" %>% "Function 1" %>% "Function 2" %>% ...

Optionally write result to a variable using the <- operator

The table of data that is to be processed

Passes data from one command to the next

A data manipulation function

Example dplyr functions

select(Parish, MalePop, FemalePop, Population)	Returns only the named columns
filter(Age >= 18)	Returns only rows meeting the criteria
mutate(Occode = as.numeric(Occode))	Update the values in a column
mutate(AgeBand = as.integer(Age/5))	Create a column based on another
group_by(Sex, AgeBand) **summarise**(count = n())	Count records in groups
inner_join(occode_lookup, by = 'Occode')	Join another table to data. Specify column(s) to join on.

THE NATIONAL ARCHIVES

Tidy Data: pivoting

pivot_wider( )

Year	Sex	Count
1851	M	103
1851	F	100
1881	M	150
1881	F	152
1891	M	130
1891	F	135

Sex	1851	1881	1891
F	100	152	135
M	103	150	130

pivot_longer( )

THE NATIONAL ARCHIVES