**“Doctors as Data Scientists” – providers notes**

**This documents is in two parts and is for the providers of the “Doctors as data scientists” material developed by Tim Frayling during the academic year 2022-2023.**

**First**, it describes the set up of a teaching session in R studio cloud, now known as posit cloud. The teaching session was delivered to ~240 3rd year BMBS students in four batches of 30-60 students and over 3 lots of 2 hours sessions, delivered in September, December and March. This approach has the advantage that you do not need a computer lab , but can use any lecture room and students bring their laptops , or tablets. If it is possible to guide the students away from using tablets and to laptops that is much better, but maybe not possible for reasons of equity. More details are in the guide section of posit cloud but the below is specific to the teaching of ~240 3rd year undergraduate BMBS students, in four batches – a maximum of 50-60 in a room.

**Second** it gives some key pointers to providers on teaching in this format, as learnt the hard way from delivery to 240 University of Exeter 3rd year BMBS undergraduates in 2022-2023 (in four batches of 30-60 students per session).

**SECTION ONE: Teaching or shared working in R Studio Cloud**

*Overall comment on the functionality of R Studio cloud/posit cloud. Overall I found this an excellent format for introducing a large number of students to working with data – especially as the online option means you do not need to find a computer lab large enough. Files can be added to a base project that appears for all students and we had sufficient RAM to handle 60 students working with the same files at once. The only slight annoyance was that if the providers wanted to add new files to the project, the students did not see them but instead had to start a new project.*

Set up a personal R studio Cloud account. Follow the <https://rstudio.cloud/learn/guide> to get details of how to:

1. Invite people into a shared workspace e.g. tim has a workspace called “Doctors as Data Scientists” which contains a project “Doctors as Data Scientists project 1” and invited teaching assistants in:
2. Pick your plan (scroll down to “Teaching with Cloud” part of online guide. Options to give access to students and the teaching assistants:
   1. Individual Instructors plan with the following paying options:

Pay a metered, variable amount per month.

* You sign up for the [Instructor: Instructor Pays](https://rstudio.cloud/plans/instructor?option=instructor) plan and pay $15 at the beginning of each month.
* Your students only need to sign up for the [Cloud Free](https://rstudio.cloud/plans/free) plan in order to work in your course space.
* Within your course space, you and your students have access to all premium features associated with the Instructor plan.
* Each month, you will be charged based on the usage by you and your students within your course spaces during the prior month. You will be charged 10¢ per project hour over the 300 hours included with your plan.

If students only on for a few hours each per month, the above is cheapest e.g. $364 for 3850 person hours. (the first 300 hours come free) for 250 students x 15 hours plus 40 hours for instructor. **Set this up with procurement paying using their credit card. I did this with a MS teams call with procurement. Setting up for 280 students 15 hours each over 3 months.**

The other option is to pay a flat fee of $5 per student but given the $0.10c per student per hour costs that would only be cheaper if ALL students used more than 50 hours – extremely unlikely and too expensive anyway.

1. Follow guide to make a “base project” that contains all your critical files you want shared. Note when you set it to be a base project you cant see it but have to go to space settings page, click the gear and edit the project.
2. Make sure the base project is accessible to all by going into the project and clicking gear > access
3. Students will need to register with R Studio Cloud Free - username and password all that is required.

WORKING ON A TABLET – generally not advised, but some students do not have laptops

Note that if students are on a tablet they will have to turn off smart punctuation

Settings > general > keyboard on an ipad.

Otherwise “” come out in unrecognisable format

**SECTION 2. Important notes to providers on delivery**

Based on teaching up to 60 3rd year BMBS students per session x 4 batches:

1. A few slides at the start to try and explain the session (I used Doctors\_As\_Data\_Scientists\_BMBS\_ResearchSSU\_20221205) and what it is about, providing some headline grabbing examples, such as the ability to study all 65 million UK residents in primary care records for COVID. Also very helpful to have a junior doctor who has worked with R come along and explain their experience in person – we had a junior doctor who reminded/told students that they had to perform an audit as part of their F1 training, which involves data.
2. We started the second and third sessions with a reminder exercise involving writing a short script to return the attendance code for the session (our medical students have to sign in to all taught sessions with an attendance code and it is the first thing they wanted to know, so instead of telling them we asked them to :
   1. Use the <- operator to place the attendance code in an R object/vector called “attendance\_code”
   2. Use the <- operator to place a message such as “today’s attendance code is” in an R object/vector called “my\_message”
   3. Use the paste(,) function to paste the two objects together and return the complete message
3. Most students instinctively just wanted to go through the worksheet as fast as possible by typing in the commands in green and moving on to the next command, not learning along the way. Unless explicitly told, and often even when they were, they ignore the black text that explains what they are doing. The best way of addressing this problem I found was to stop every 15-20 minutes and inform the students that you would ask one of them at random to explain a command and what it is doing in ten minutes time. Then you and the teaching assistants spend that 10-15 minutes asking as many students as possible 1 to 1 or 1 to 2 to explain the command. For example, ask the students to explain exactly what is happening in:

My\_new\_data <- mydata[,]

head(testdata, n= 5)

testdata\_subset <- testdata[,c(1:3,6)]

testdata\_women <- filter(testdata, gender == "2")

1. Of the above we spent a lot of time asking and checking the students understood the concept of square brackets to select from a table.
2. Common causes of frustration and confusion were often caused by simple things:
   1. Not creating, saving and working from a file but instead working from the console
   2. Not running / executing a command from the file using “Run” , but typing it in and assuming it had worked.
   3. Spaces and case sensitive errors, and copying and pasting speech marks from the word document worksheet

Finally see if you can better some of the comments we received on asking the medical students if they found it fun:

“I don’t hate this as much as I thought I would”

“this is like crochet”

“its more fun than when a patient vomits on you in the ward rounds”.