

Analytical Platform and related tools training

MoJ coding training leads (the Analytical Platform team is not responsible for the content on this page)

2022-09-01

Contents

1	Summary of available resources	5
1.1	The Analytical Platform	5
1.2	R mentoring	5
1.3	Internal Training Group materials	6
1.4	Coffee and Coding and Bitesize sessions	6
1.5	Analytical Function training opportunities	6
1.6	Datacamp	6
1.7	Other assistance	7
2	Analytical Platform	9
2.1	Introduction	9
2.2	Summary of key terms	9
2.3	Getting set up	10
2.4	Managing data	10
3	Internal Training Group materials	13
3.1	R Training	13
3.2	GitHub Training	14
3.3	SQL Training	14
3.4	Python Training	14
3.5	Get involved!	14

Chapter 1

Summary of available resources

1.1 The Analytical Platform

The Analytical Platform is a data analysis environment, enabling the use of modern open source tools such as R and Python, and holding key datasets for MoJ analysts.

To learn more about the Analytical Platform and to get up and running, go to the Analytical Platform chapter. More extensive information is provided by the Analytical Platform user guidance.

1.2 R mentoring

It is recommended that all who are new to R or Data and Analysis request an R mentor. The purpose of the scheme is to provide a better on-the-job R learning experience and raise awareness of the preferred Data and Analysis ways of working that will for instance enable people to get up to speed more quickly with others' code. The scheme is also open to non-coders who need to use the Analytical Platform to advise them through the learning process, and for those commencing a more complex project involving the use of R.

To request an R mentor please complete this mentee form. If you could become an R mentor (we have a shortage of mentors) please complete this mentor form. For more information please contact Jessica Dawson.

1.3 Internal Training Group materials

This internal training is recommended for those working in Data and Analysis as it enables you to get up and running with R, SQL, Python and GitHub using the MoJ Analytical Platform. The main introductory R, SQL and GitHub sessions are run live in February/March, June/July and October/November each year while you can also work through R, SQL and Python sessions yourself using the training material and/or recordings. To learn more about the sessions currently available and how to access the material and recordings, go to the Internal Training Group materials chapter.

1.4 Coffee and Coding and Bitesize sessions

The internal training (see above) is complemented by coffee and coding and bitesize sessions. The contacts are Katharine Breeze and George Papadopoulos respectively.

1.5 Analytical Function training opportunities

There are now many technical Analytical Function training opportunities for analysts including about topics not presently covered internally e.g. an introduction to python. Examples useful for RAP practitioners include Best practice in programming – clean code – GSS (civilservice.gov.uk) and an Introduction to unit testing – GSS (civilservice.gov.uk). You can view such opportunities via:

- A user-friendly list (but not updated since Jan 2021)
- GSS Training Courses (many which are open to all analysts)

1.6 Datacamp

Paid Datacamp licenses are beneficial to cover gaps in current training provision that are not picked up by either internal or GSS/Analytical function training currently e.g. more advanced R, SQL, and Python skills. You can read more about Datacamp [here](#).

If you would like to have a Datacamp license please contact Aidan Mews or Rosina Costello providing a brief explanation about why it will be important for your work along with approval from your deputy director.

1.7 Other assistance

Technical help can be requested via the following Data and Analysis slack channels:

- `intro_R` which provides support to those starting out in the world of R and RAP.
- `R` which is for beginners and experts alike.
- `sql`
- `python`

You may also find useful a trello board providing Links to further free online analytical training including in R and R cheatsheets.

Chapter 2

Analytical Platform

2.1 Introduction

To gain an overview of the Analytical Platform watch this 2-3 min introductory video. For more technical details, you can also read the user guidance section About the Analytical Platform.

2.2 Summary of key terms

It will help you to be familiar with the following key terms:

- **Analytical Platform (AP):** A data analysis environment providing modern tools and key datasets for MoJ analysts. AP contains training documents, resources, and access to various analytical software such as Rstudio and Jupyter.
- **Control Panel:** A place to navigate to Rstudio, Jupyter, S3 Buckets etc
- **RStudio:** Development environment for writing R code and R Shiny apps
- **JupyterLab:** Development environment for writing Python code including Python notebooks
- **Git:** Version control software that enables multiple people to make separate changes at the same time.
- **GitHub:** A web-based interface that uses Git and on which you publish and share your version-controlled code. You use Git locally (e.g. using RStudio) to track versions of your code, and then submit those changes to Github.
- **GitHub Repositories (Repo):** Similar to setting up a project folder on DOM1 shared drive to save work and share with others. Files on Github

Repos represent the definitive version of the project. Everyone who works on the project makes contributions from their own personal versions.

- **Amazon S3:** A web-based cloud storage platform. Where your home directory stores working copies of code, data and final analytical outputs should be stored in Amazon S3. Access to amazon S3 buckets can be managed.
- **Slack:** Collaboration tool where you can get technical support for Analytical Platform tools such as R, Python, Git. You can share knowledge, submit admin requests and communicate quickly with other AP users.

2.3 Getting set up

First, so that you can request any help you need when getting set up on the Analytical Platform, it is recommended that you join the Data and Analysis Slack. You can read a very brief overview here with further information including how to sign-up here.

Second, follow the steps in the Getting Started section of the Analytical Platform User Guide. You need to:

- Set up a GitHub account with two-factor authentication.
- Set up a Analytical Platform account.

Once you have completed the above two steps, you should be able to access the Analytical Platform Control Panel and from there open the relevant tool (e.g. RStudio for R).

Third, as code written on the Analytical Platform should be stored in a Git repository on GitHub, complete the steps to configure Git and GitHub for the Analytical Platform. You can learn more about GitHub by viewing this introduction and by attending the internal introduction to GitHub.

For those that need to get set up to use Athena databases for SQL (in R or Athena) on the Analytical Platform, please follow the additional instructions in the “Training Requirements” section of the Introduction to SQL training repository.

2.4 Managing data

Once you have got set up on the Analytical Platform, do read about the following data management/handling topics:

- How data are held on the Analytical Platform and finding the data you need. You can read about the three different data storage options (Amazon S3, Curated databases and home directories).
- Working with Amazon S3, including how to upload data files and download data files using Amazon S3.
- Information governance procedures to be followed, in particular when moving any data onto the Analytical Platform, when a data movement form must be completed.
- Data retention policies including when deleting data means they are permanently deleted.

Chapter 3

Internal Training Group materials

This internal training is recommended for those working in Data and Analysis as it is run using the MoJ Analytical Platform. There are live R, SQL and GitHub training sessions, which are generally run in February/March, June/July and October/November each year. You can also work through the R and SQL sessions yourself using the training material and recordings. There is also Python training in development; we are considering how best to run these.

3.1 R Training

The following sessions are available; click on the links to view the material in the GitHub repositories. The first two are particularly recommended for new starters.

- Introduction to R
- R Charting
- R Markdown
- Interfacing Excel with R
- Writing Functions in R
- Developing R packages & RAP ways of working

The recordings of these R training sessions are available on the R training MS Stream Channel.

3.2 GitHub Training

Whether seeking to use R, SQL or Python, it is recommended that all newcomers attend this Introduction to GitHub (click on link to view material in GitHub repository) session.

3.3 SQL Training

All newcomers who are to use SQL should attend the Introduction to SQL (click on link to view material in GitHub repository) session. A recording of this session is available on the SQL training MS Stream Channel.

3.4 Python Training

These have not yet been run as live sessions, but you can work through the material at your own pace.

- Introduction to Python
- AWS Tooling demos

3.5 Get involved!

One great way of learning is by teaching. If you would be interested in being part of the R, SQL and/or Python training groups, whether booking courses/allocating places, designing training, or presenting, please contact Aidan Mews, Georgina Eaton or George Kelly.

If you have any questions please contact Aidan Mews, Georgina Eaton or George Kelly.