# Workshop preparation: software installation tasks

This guide will describe how to install three programs that will be used within the data visualisation workshop. These three programs are R, RStudio, and GitHub. It will discuss what each of these programs are, how they are useful, and how to install them. GitHub is optional, but learning more about it, and how it integrates with RStudio, can pay dividends, especially if you plan on using R extensively and/or in collaboration with others.

# R

R is a statistical programming language. It is free, open source, and popular with professional statisticians. It also has a notoriously steep learning curve, which I will try to help everyone overcome by the end of the workshop.

To install R, go to the Comprehensive R Archive Network (CRAN), and select the correct installation option for the computer you will be using in the workshop:

<http://cran.r-project.org/>

**Please install R before either of the other programs.**

# RStudio

RStudio is a free integrated design environment (IDE) for R, which makes R easier to work with. It contains a large number of features which make it very popular, including tools for managing and maintaining complex research by designating particular folders on your computer as ‘R Projects’, which can then be archived, stored and managed through Git/GitHub. Used well, this means you never need to suffer the disaster of losing all your research due to a hard drive breaking down or a file getting deleted by accident. Perhaps most importantly (!), RStudio makes R *look* a lot nicer!

To install RStudio, please go to the following location and select the free, desktop option. Further details about RStudio and its features are available on the website.

<http://www.rstudio.com/products/RStudio/>

**Please install RStudio only *after* installing R.**

# Git/GitHub

Git is a set of tools for managing collections of files, especially text files containing programming code. This includes R script files, and RStudio contains some features for working closely with Git to make sure projects and code is effectively developed and maintained.

GitHub is both a website and data storage facility that ‘hosts’ collections of files managed with Git, and a graphical user interface that makes Git, which is primarily operated with the command line, a bit easier to use.

Within this workshop, Git/GitHub will be used only in quite a simple and minimal way, as a place you can go to to get the latest version of the workshop material. This material will include slides and handbooks like this one, as well as data and code. These material are available from the following location:

<https://github.com/JonMinton/AQMEN_Data_Vis_Workshop>

There are two buttons available near the bottom right hand side of the website: one button saying ‘clone in desktop’, and another saying ‘download zip’. Using either option, it should be possible to download the material, arranged in the same directory structure as I have developed it. This will mean you have all the data and material you need for some of the practicals during the workshop.

For now, I would recommend just using the url above as a file download location. To do this, please wait until the date of the workshop and go to the url above, click on the ‘Download Zip’ button, and unzip the files. An alternative approach is described in the ‘Extra credit’ section below.

# Extra Credit (Optional)

1. Please send me a dataset you would be interested in working with, and happy for others to work with, during the workshop. Such datasets should ideally be contained in a single file, in a delimited text file format such as .txt or .csv, and under 50Mb in size. Please also provide a clear description of the dataset and your research interests in the data. Please only send the dataset if you have appropriate permissions for the data to be publically available, as I will add the dataset to the above GitHub repository.
2. Using the github url above, click on the ‘Clone in Desktop’ option rather than the ‘Download Zip’ button. This will be a first step towards learning more about using Git/GitHub’s features and benefits. If you would like to learn a little about GitHub through doing, I would recommend clicking on ‘Clone in Desktop’ now, on using the machine you will be bringing to the workshop, and following the instructions. This should install a simple GitHub GUI which will make it easier to manage Git file collections, known as repositories or ‘repos’. The contents of the files hosted at the url address will change as we get closer to the session. After ‘cloning’ the ‘repo’ on your machine, I would recommend trying to update it to the latest version closer to the date of the workshop. This is known as a ‘pull’ or a ‘fetch’ operation in Git/GitHub lingo.
3. Set up a directory as a project within R Studio, sign up for GitHub, and upload the project to a GitHub repository. Change the contents of the R project, and update the repository with the changes.

If you have any questions please feel free to contact me: [jonathan.minton@glasgow.ac.uk](mailto:jonathan.minton@glasgow.ac.uk)

Best of luck. I look forward to meeting you in Glasgow!

Dr Jon Minton

11 September, 2014