Setting Up

> ⚠️ **Changes before 22/23**
>
> ☐ Add Learning Outcomes for each week
> ☐ More time on Git/GitHub and pull/push/browsing version history; point to setting quantitative research questions (CASA0007 Week 1) but shift focus on to policy. Explain why there are two assessments and why one has group work with a peer component.
> ☐ Remove Docker installation practical from FSDS; keep in QM.
> ☐ Jon to create short video on *purpose* of Docker that can be used in QM and discussed in FSDS (as part of a data science 'workflow')
> ☐ Add discussion of .gitignore file to GitHub practical (and make sure they add .gz, .csv, and .zip!)

Overview

> ❗ **Important**
>
> This week's learning objectives are:
>
> 1. A basic understanding of the data science 'pipeline'.
> 2. An understanding of how data scientists use a wide range of 'tools' to *do* data science.
> 3. A completed installation/configuration of these tools.

This week is focussed on getting you set up for the rest of the course in terms of having the requisite software installed and accounts configured so that you can keep track of your work, write code, and track changes. However, you should *also* see this as preparing the foundation not only for your remaining CASA modules (especially those in Term 2) but also for your post-MSc career: the ability to manage and version code (GitHub); collaborate around a shared codebase (GitHub+Markdown); and produce reproducible code (GitHub+Docker) is integral to modern software development and data science.

Setting Up

You should also see this session as connecting to *Quantitative Methods* Week 1 content on 'setting quantitative research questions' since the main assessment will require you to develop a data-led policy briefing. In other words, you'll need to map current policy on to one or more research questions that can be quantitatively examined using the tools and techniques acquired over the course of the term! While you don't need to start work on this yet, you should keep it in the back of your mind for when you come across readings/results that you'd like to explore in more detail.

Preparation

Although none of these activities are *compulsory* in advance of the first session, getting your computer set up to code *does* take time and most of these preparatory activites are fairly straightforward... with a few exceptions noted below. If you are able to get these tools installed in advance then you can focus on the taught content in the first two practicals rather than also wrestling with an installation process. This will also give us more time to help you if you discover that you're one of the unlucky few for whom getting set up is a lot more work!

> 💡 Tip
>
> Complete as many of these activities as you can:
>
> 1. Go through the computer health check.
> 2. Have a go at installing the Command Line Tools for your operating system.
> 3. Have a go at installing the programming environment.

The last of these is the stage where you're most likely to encounter problems that will need our assistance, so knowing that you need our help in Week 1 means that you can ask for it much sooner in the practical!

Class

In this week's workshop we will review the module aims, learning outcomes, and expectations with a general introduction to the course.

> ⚠️ Changes before 22/23
>
> ☐ Standardise delivery by TAs where practicals are TA-led (e.g. FSDS): clearer guidance on how to take students through each week.
> ☐ Clearer articulation of difficulty levels/targets for components *within* each week's practicals.

☐ Make links between lecture and practical content explicit; ideally trace a question through the whole process (*e.g.* "If I wanted to know if and where blue plaques are clustered in London how would I find out?")

Practical

This week's practical is focussed on getting you set up with the tools and accounts that you'll need to across many of the CASA modules in Terms 1 and 2, and familiarising you with 'how people *do* data science'. Outside of academia, it's rare to find a data scientist who works entirely on their own: most code is collaborative, as is most analysis! But collaborating effectively requires tools that: get out of the way of doing 'stuff'; support teams in negotiating conflicts in code; make it easy to share results; and make it easy to ensure that everyone is 'on the same page'.

> ❗ Important
>
> The practical's learning objectives are:
>
> - Get you up and running with coding and collaboration tools.
> - Provide you with hands-on experience of using these tools.
> - Configure your programming environment for the rest of the programme.

> 💡 Tip
>
> To save a copy of notebook to your own GitHub Repo: follow the GitHub link, click on `Raw` and then `Save File As...` to save it to your own computer. From there, move it to your `fsds` folder and follow the steps to `add` and `push` the file to GitHub.

If you have already installed the programming environment then the practical can be downloaded from GitHub. Since there is no code to run this week, you can also view it online in your web browser.