Getting Started

There are many ways to run code and, depending on the kind of environment in which you end up working or studying after graduation, your experience might differ quite a bit from that of other students; however, in general, a good deal of contemporary data science research is written in either R or Python (or a mix of both!) running in a Jupyter Notebook.

Installing the Programming Environment

Over the years, we have experimented with a range of approaches to setting you up with a programming environment: VirtualBox; Vagrant; Docker; and Anaconda Python directly. Each of these has pros and cons, but after careful consideration we have come to the conclusion that **Docker** is the most robust way to ensure a consistent experience in which all students end up with the same versions of each library, difficult-to-diagnose hardware/OS issues are minimised, and running/recovery is the most straightforward.

Some students are unable to run Docker on their Windows machines, in which case **Anaconda Python** can be used with the configuration file that we provide. However, if your machine runs Docker then you **must** use Docker: this isolates the programming environment form your computer, ensuring that nothing is clobbered by accident, and guaranteeing that you are working with the same version of every library that we are. Anaconda is only for emergencies.

To install the Programming Environment follow the instructions in Practical 1 here (Task 5, but 4 and 3 are recommended)