Supplementary: Class Virtual Machine

You can now access the [class virtual machine](https://spark-public.s3.amazonaws.com/datasci/Coursera-Data-Science-Ubuntu.ova), an ubuntu 12.04 32-bit linux image preconfigured with all the software you will need for the course.

You can also get a torrent for the class virtual machine by going to the following URL: <https://spark-public.s3.amazonaws.com/datasci/Coursera-Data-Science-Ubuntu.ova?torrent>

You will need a virtual machine player to use this resource: VirtualBox, the free VMPlayer, or the native Windows player should all be able to convert the image or read it directly.

If you are using VirtualBox, you may want to install ["Guest Additions"](https://www.virtualbox.org/manual/ch04.html#idp46785384343968) and set up a [shared folder](https://www.virtualbox.org/manual/ch04.html#sharedfolders) to move files back and forth.

You are not required to use the virtual machine; you can install the (modest) necessary software yourself. At a minimum, you will need Python 2.7, SQLite, R, and git. (Note that R will not be required for this first course, but will be used in later courses.)

The user account password for the virtual machine is 'password'

If you have trouble accessing or using the VM, please post a message to the forum so the staff or your classmates can respond.

Supplementary: Github Instructions

Throughout the course we may or may not provide you with starter files (resources and code skeletons) for the assignments.

We will provide these files to you online through a Github repository rather than through Coursera.

Source control software addresses the problem of having multiple collaborators reading and writing different parts of a shared set of files.

Git, in tandem with Github, is the source control software we’ll be using. Github is an online service that provides free git repositories.

**Cloning the repository:**

**Virtual Machine Users**

The repository is already cloned and located in the home directory.

**Windows Users**

If you don't have Git installed, installation instructions can be found here: <http://git-scm.com/download/win>*\*\**

You can install a GUI on top of GIt from here: <http://windows.github.com/>

Once the GUI is installed, you can navigate to the repository page, <https://github.com/uwescience/datasci_course_materials>, and click on the 'Clone in Windows' button.

**Linux Users**

If you don't have Git installed, installation instructions can be found here: <http://git-scm.com/download/linux>

If you have Git installed, you need to clone the repository. From a terminal, navigate to where you want the repository to be located and perform a clone.

git clone https://github.com/uwescience/datasci\_course\_materials.git

**Mac Users**

If you don't have Git installed, installation instructions can be found here: <http://git-scm.com/download/mac>

If you have Git installed, you need to clone the repository. From a terminal, navigate to where you want the repository to be located and perform a clone.

git clone https://github.com/uwescience/datasci\_course\_materials.git

**Updating the repository:**

**Virtual Machine, Linux, and Mac Users**

Navigate to the repository and perform a git pull

cd datasci\_course\_materials

git pull

**Windows Users**

Perform a pull on the datasci\_course\_materials through the GUI interface. If you set the repository to stay in sync, the gui will give you notifications where there are updates to be pulled.