Instructions: Setting up course software

Knowing R, H2O Flow and WEKA will give you a powerful, flexible and scalable set of tools to manipulate and analyse big data.

Installing R and RStudio

You will be using R, a language and environment widely used for statistical computing and graphics, and RStudio, an open-source integrated development environment (IDE) for R.

1) Download and install R. Follow the instructions on the R Project website.

> **Need help?**
>
> See the comprehensive FAQ and How to Documents on the R Project website.

2) Download and install RStudio Desktop ensuring you choose the installation for your platform.

> **Need help?**
>
> See the comprehensive FAQ and the support forum on the RStudio website.
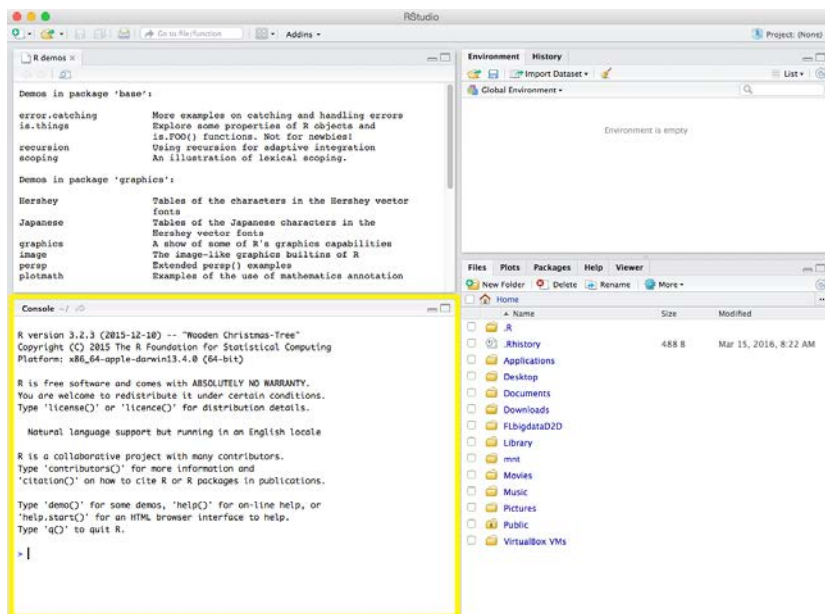
Installing R packages

You also need to install the R packages we'll use to explore and manipulate our data:
- dplyr
- ggplot2
- AER
- rpart
- rpart.plot

1) Open RStudio.

2) Enter each command below, one line at a time, into your RStudio console.

```
install.packages('dplyr')
install.packages('ggplot2')
install.packages('AER')
install.packages('rpart')
install.packages('rpart.plot')
```


RStudio console (highlighted in yellow)

Cheat sheets are your friend

Check out some of the cheat sheets available for RStudio:

Data wrangling with dplyr and tidyr

Data Visualization with ggplot2

Downloading datasets

Download the dataset you'll use for our week 1 exercises.

1) Download the bank customer dataset:

   a) Go to QUT-BDA-MOOC/FLbigdataStats in Github.

   b) Select **Download ZIP**.

2) Unzip the archive and save the bank_customer_data.csv file in the following location:

   `~/FLbigdataStats/bank_customer_data.csv`

   Note: The ~ refers to your Home directory. Not sure where your home directory is?

   In RStudio:

   a) Select **Tools**, then **Global Options**.

   b) Select **Browse** (next to 'Default working directory when not in project').

      This will show you the current home directory for RStudio on your machine.

Installing H2O

H2O can be installed via RStudio. You'll need to open RStudio, then copy and paste each of the following commands, one line at a time, into your RStudio console. The H2O website contains more complete instructions should you still have issues.

```
install.packages("h2o")
library(h2o)
localH2o = h2o.init(ip = "127.0.0.1", port = 54321)
```

These commands first install the H2O R package, then load the library and finally initialise the H2O sever. When the server is initialised for the first time the following H2O jar file will automatically be downloaded and give the following output.

```
H2O is not running yet, starting it now...
Performing one-time download of h2o.jar from
        http://s3.amazonaws.com/h2o-release/h2o/rel-
        knuth/11/Rjar/h2o.jar
(This could take a few minutes, please be patient...)
```

Note: Java Runtime is required by R and H2O. R will attempt to install it when loading H2O if you do not already have it installed.

You will need to install Java Runtime manually if the automated install fails. You will first need to ensure you have a Java-compatible version of OSX if you are a Mac user. (If you have started the H2O install process, simply close and re-open R, then continue at the next command line you were up to.)

> **Familiarise yourself with H2O**
>
> Use the R Tutorial on the H2O Documentation website to learn about using H2O with R.
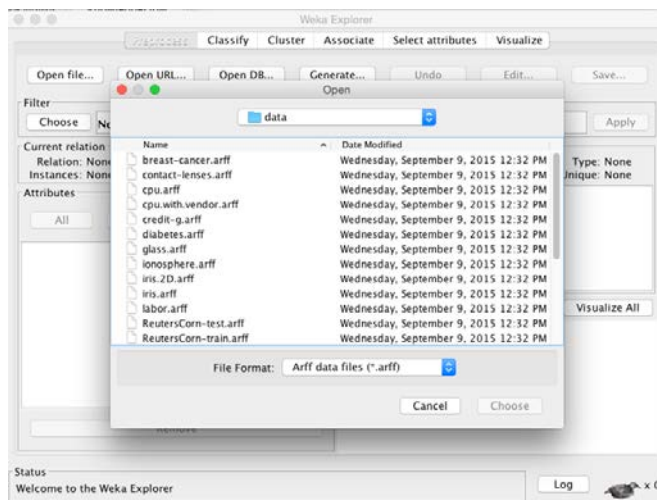
Installing Weka

Go to the Weka website to download and install the latest stable version for your platform.

Weka sample datasets

Check the sample datasets provided with Weka have installed correctly.

1) Open Weka and and select **Explorer**.

2) Select **Open File** to open your default Weka data directory.

3) Check you can see Weka's sample data files in .arff format.

   a) Check the Weka install package if the sample data files are not in your default Weka data directory. Copy the data directory from the Weka install package to your default Weka data directory.
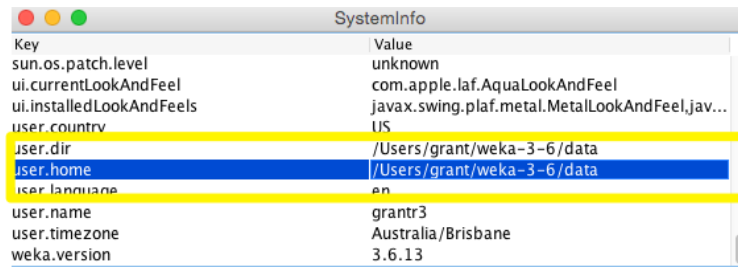
Can't find your default Weka data directory?

1) Go to the **Help** menu and select **System Info**.
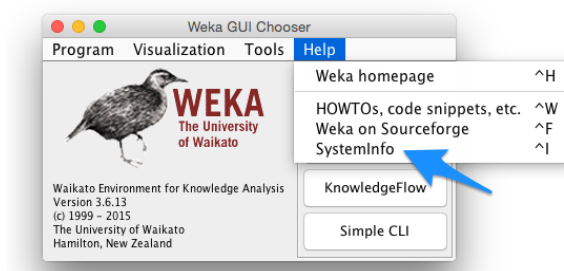
2) Scroll through the list of settings to find 'user.home'. This shows your default Weka data

directory.