

Data Science and Machine Learning Essentials

Lab Setup

Overview

This course includes labs that require you to conduct data science experiments in Microsoft Azure Machine Learning (Azure ML) Studio and to develop Python and R code. This setup guide describes how to prepare for the labs.

Setup Tasks

To prepare the lab environment, you must perform the following tasks:

1. Create an Azure ML account
2. Download and extract the lab files
3. Install R and RStudio
or
4. Install Python Anaconda

What You'll Need

To perform the setup tasks, you will need the following:

- A Windows, Linux, or Apple Macintosh computer.
- A web browser and Internet connection.

Create an Azure ML Account

Azure ML offers a free-tier account, which you can use to complete the labs in this course.

Sign Up for a Microsoft Account

1. If you do not already have a Microsoft account, sign up for one at <https://signup.live.com/>.

Sign Up for a Free Azure ML Account

1. Browse to http://bit.ly/azureml_login and click **Get started now**.
2. When prompted, choose the option to sign in, and sign in with your Microsoft account credentials.
3. On the **Welcome** page, watch the overview video if you want to see an introduction to Azure ML Studio. Then close the **Welcome** page by clicking the checkmark icon.

Note: Your free-tier Azure ML account allows you unlimited access, with some reduced capabilities compared to a full Microsoft Azure subscription. Your experiments will only run at

low priority on a single processor core. As a result, you will experience some longer wait times. However, you have full access to all features of Azure ML.

Install R and RStudio

R is a programming language for conducting statistical analysis processes and visualizing data. You will create R code in some labs in this course.

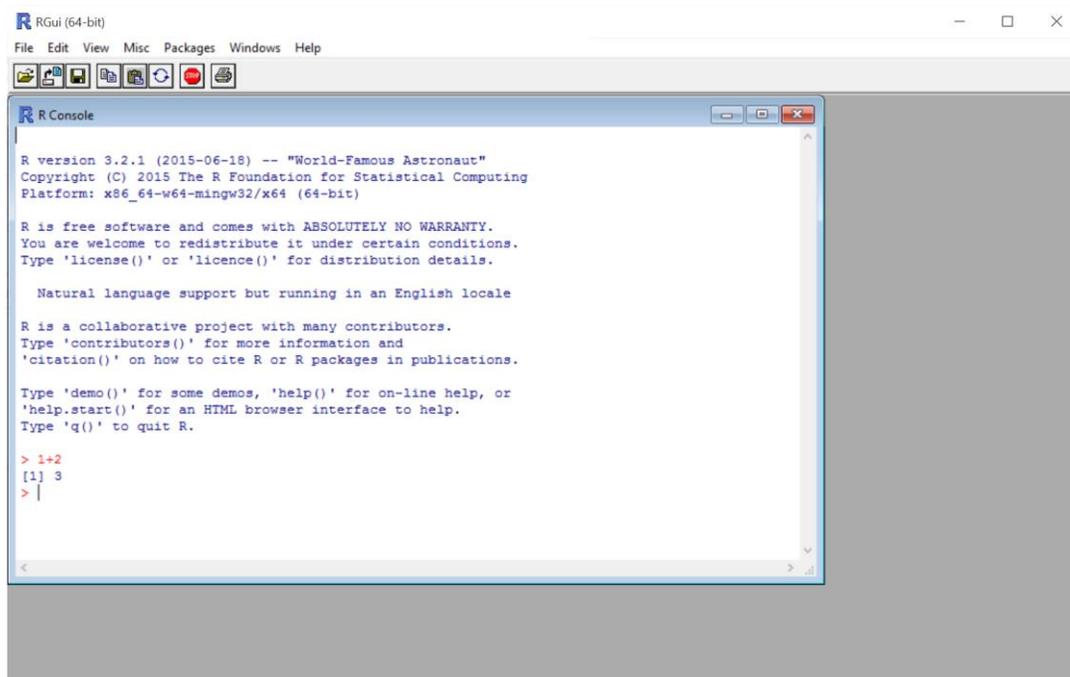
Note: In this course, you can choose to complete programming exercises in Python or R (or both). If you plan to use R, complete this procedure to install the R runtime and development tools. If you do not plan to use R, you can skip this procedure.

Install R

1. In a web browser, navigate to <https://cran.r-project.org/bin/>.
2. Open the folder for your operating system (Linux, MacOS, MacOSX, or Windows) and follow the instructions to install R.

Note: Linux users must choose the subdirectory for the Linux type and version. Windows users with 64-bit computers can install both 32 and 64 bit versions of R.

3. Verify installation by starting R and entering a simple R expression such as $1 + 2$ (which should produce the result 3) as shown in the following image.



```
RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
R version 3.2.1 (2015-06-18) -- "World-Famous Astronaut"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

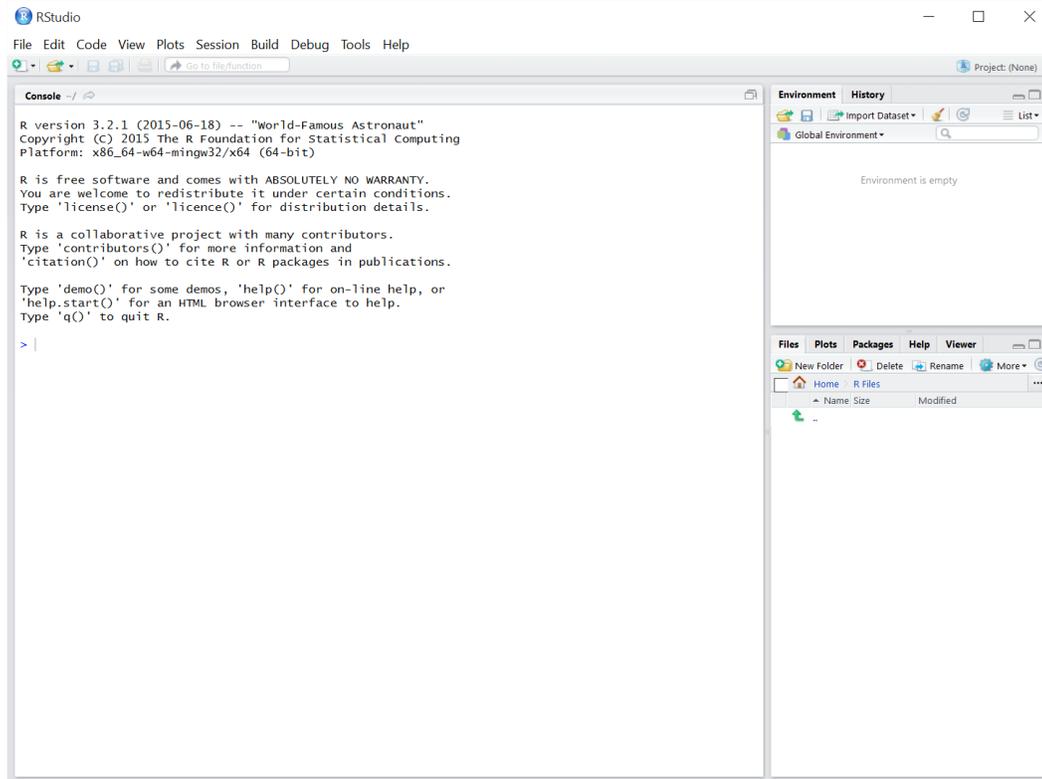
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 1+2
[1] 3
> |
```

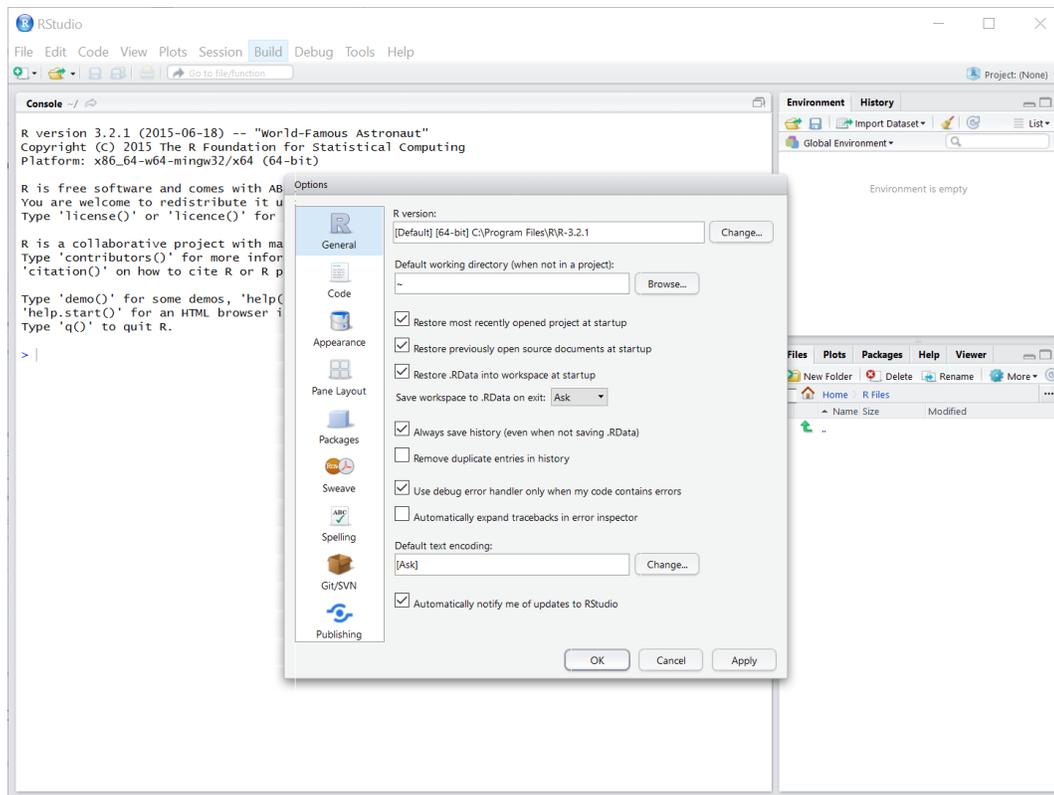
4. Close R.

Install RStudio

1. In a web browser, navigate to <https://www.rstudio.com/products/rstudio/download/>.
2. Run the installer for your operating system (Windows, MacOSX, Ubuntu, or Fedora) to install RStudio.
3. To verify installation, start RStudio; which should resemble the following image:



4. If you are using Windows, ensure that RStudio is configured to use the correct version of R (32-bit or 64-bit) by selecting **Global Options** on the **Tools** menu, as shown in the following image.



5. When you have verified installation and configuration, close RStudio.

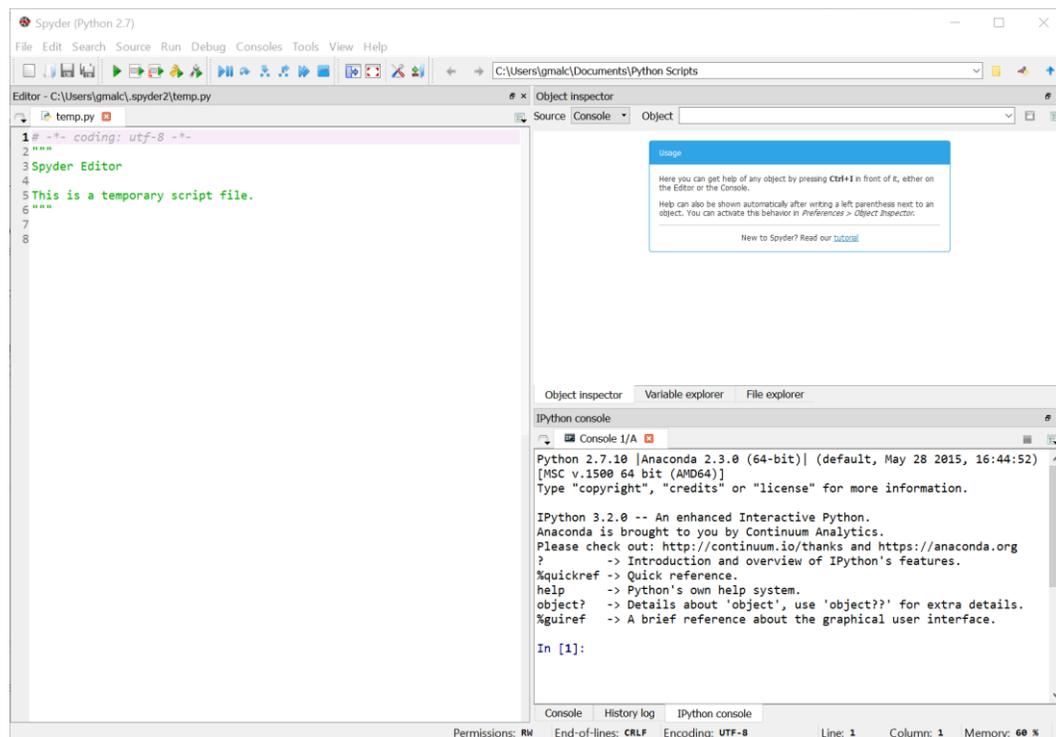
Install Python Anaconda

Python Anaconda is a distribution of Python that includes the Spyder Integrated Development Environment (IDE), which you will use to create Python code in the labs for this course.

Note: In this course, you can choose to complete programming exercises in Python or R (or both). If you plan to use Python, complete this procedure to install the Python runtime and development tools. If you do not plan to use Python, you can skip this procedure.

Install the Python Anaconda Distribution

1. In a web browser, navigate to <http://continuum.io/downloads>.
2. Choose the installer for your operating system (Windows, Apple Macintosh, or Linux).
3. Complete the installation process for Python 2.7.
4. After installation is complete, verify the installation by starting Spyder, which should look similar to the following image:



5. Close Spyder.

Download and Extract the Lab Files

Each lab in this course requires sample data files and code script files that you will use to build machine learning experiments. These files are available as zip archives in the course handouts.

Before you start each lab, download the lab files and extract them to a folder on your local computer.

Summary

By completing the tasks in this setup guide, you have prepared your environment for the labs in this course. Now you're ready to start learning how to build data science and machine learning solutions.