



EXERCISE 1

Installing RStudio and Setting up the Environment

Introduction

RStudio is a free and open-source integrated development environment (IDE) that uses the programming language R for statistical analyses. We will be using an R script to create and run a random forest model to classify land cover types in eastern Oregon.

Objectives

- Orient yourself to the data, script, and environment

Required Data

- **rf.zip** – zip archive containing course data for Introduction to Random Forests

Prerequisites

- **You have ESRI ArcGIS 10.2 installed on your computer** and have basic understanding of how to use the software.





Table of Contents

| | |
|---|---|
| Part 1: Install Software..... | 3 |
| Part 2: Review course data in ArcMap..... | 3 |
| Part 3: Open and Explore RStudio | 6 |



Part 1: Install Software

A. Download and install R and R-Studio (IDE)

1. **RStudio** requires that you have **R version 2.11.1** (or higher) installed as well. If you don't have this, follow the link to navigate to the **RStudio CRAN mirror** (<http://cran.rstudio.com/>) and follow the instructions to download and install R.
2. Next, navigate to the **RStudio download page**:
<https://www.rstudio.com/products/rstudio/download/>
3. Click on the link recommended for your operating system to download the appropriate version of **RStudio** (e.g., RStudio 0.99.887 - Windows Vista/7/8/10). This should automatically start the **RStudio** download.
4. When the download is complete, navigate to your **Downloads** folder and run the executable (e.g., **RStudio-0.99.887.exe**) to install RStudio.
 - i. Double click to run the executable and start the install.

Note: if your computer indicates that you need administrator access to install, try right clicking on the executable and select **Run Elevated**. When prompted, in the **Powerbroker for Windows Authorization** pop up window, indicate the software installation **Justification** (e.g., installation of approved software).

- ii. Follow the subsequent prompts to complete this installation.

Part 2: Review course data in ArcMap

In these exercises, we will be creating a land-cover map of a subset of eastern Oregon using a subset of a National Agriculture Imagery Program tile. The course zip folder contains data (imagery and training data) and scripts necessary to run the model.

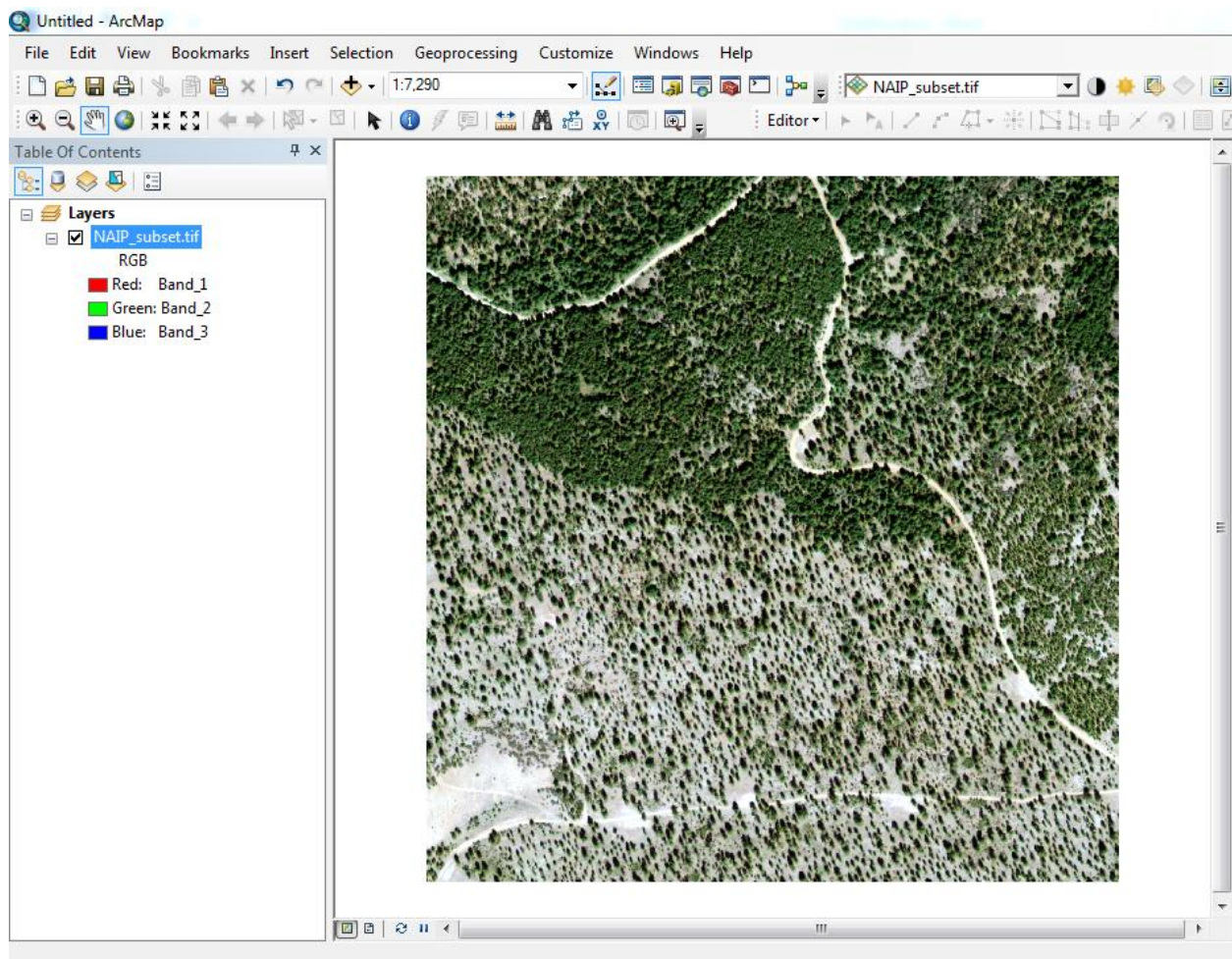
A. Open the data in Windows Explorer

1. Extract the course data, **rf.zip**, to your local drive
 - i. Create a course folder on your local drive (we suggest a folder called RF in the root C:\ drive – for example, **C:\RF**)
 - ii. Right-click on the zip folder, and choose to unzip the contents to your new course folder.
2. The contents of the zip file will contain:
 - i. **Data:** all of the imagery and shapefiles necessary for running random forest (RF) model.
 - ii. **References:** pertinent literature
 - iii. **Scripts:** folder containing the R script for running a thematic classification using a random forest algorithm.
3. Open the **Data** folder.
 - i. You should see a few more subfolders: **Imagery**, **Output**, and **Shapefile**.
 - ii. Open these subfolders and explore the contents.
 - (a) You should notice that the **Output** folder is currently empty (this will be where you write or save your files)

- (b) The **Imagery** folder contains two key layer stacks. One layer stack is a subset of a NAIP scene including only its optical bands (which include red, green, blue, and near infrared). The second layer stack contains a Normalized Difference Vegetation Index (NDVI) layer in addition to the aforementioned spectral bands.
- (c) The **Shapefile** folder contains a point shapefile used to train the random forest model.

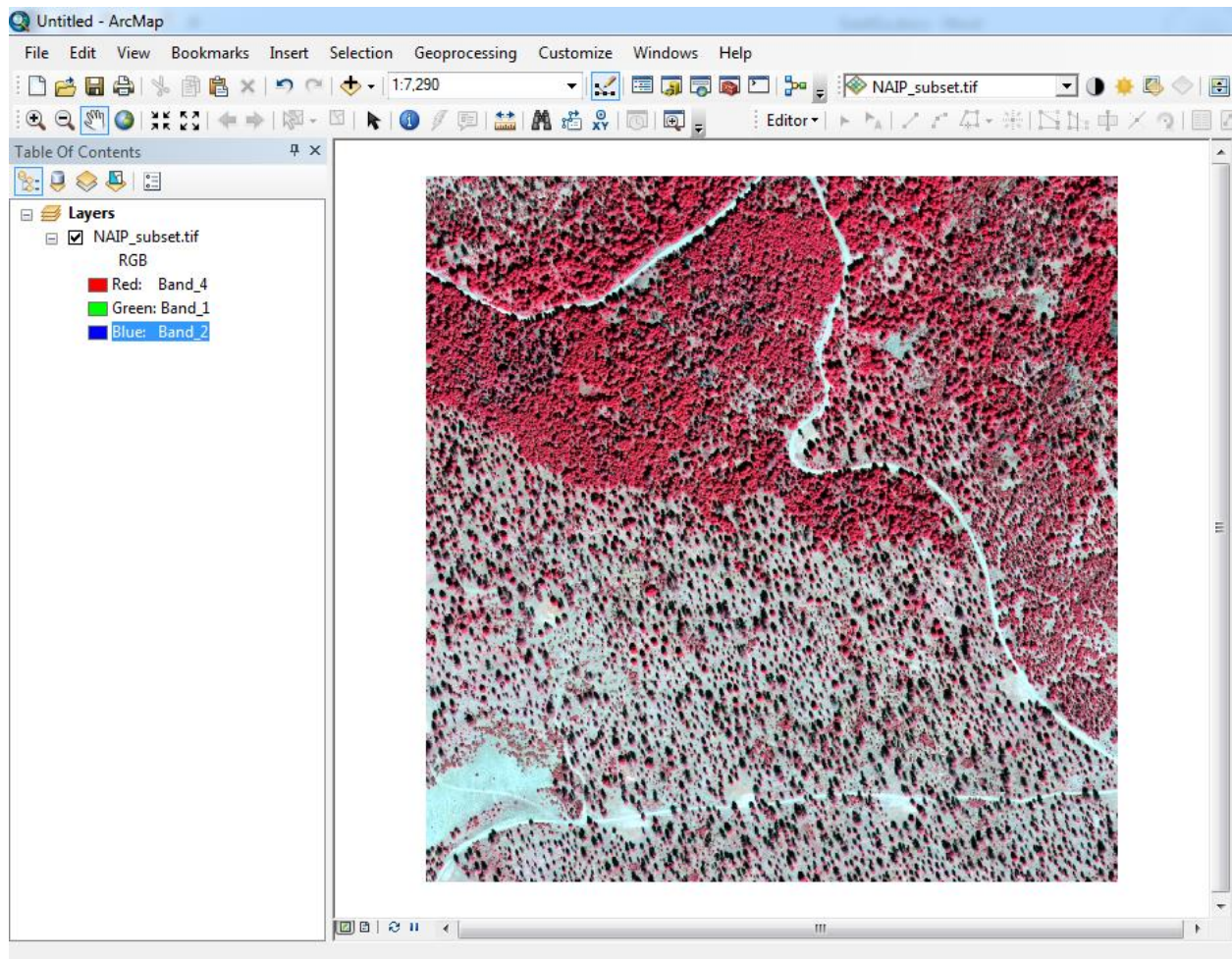
B. Examine the imagery in a GIS

1. Launch **ArcGIS** by choosing the **Start Menu, All Programs, ArcGIS, and ArcMap 10.2**
2. Open a **Windows Explorer** window by pressing the **Windows key** and **E** on your keyboard.
3. Navigate to your **course data folder**, and access the imagery by double-clicking the **Data** subfolder and then double-clicking the **Imagery** subfolder.
4. Select the file **NAIP_subset.tif** and drag it into your **ArcMap viewer**. The image will display as a natural color composite by default.



5. If you prefer looking at imagery with as a false color composite, feel free to adjust the imagery to a color infrared composite now.

- i. Click the **red rectangle** next to **Red: Band_1** and choose **Band_4** from the resulting drop-down menu. This will assign the red color gun to the near infrared band.
- ii. Click the **green rectangle** next to **Green: Band_2** and choose **Band_1** from the resulting drop-down menu. This will assign the green color gun to the red band.
- iii. Click the **blue rectangle** next to **Blue: Band_3** and choose **Band_2** from the resulting drop-down menu. This will assign the blue color gun to the green band.



Note: This image shows a subset of a NAIP tile taken over the southern portion of the Malheur National Forest in eastern Oregon. The image shows a dry, bright soil background with forest cover (conifer; dark red) ranging from sparse to moderately dense. A road cuts through the image. In addition to forest cover and ground, the image also shows shrub cover and shadows.

C. Load the training data

1. From the **course folder**, navigate up a level back to the **Data** subfolder and double-click on the **Shapefile** subfolder. Select **train_subset.shp** from the folder and drag it into the **ArcMap** viewer.
2. In the **Table of Contents**, right-click on **train_subset** and choose **Open Attribute Table**.

Note: Each point is attributed with a unique id (**FID**). Each point is also attributed by one of the four class names (**Class**) and an associated numerical class identifier (**Class_Num**). These classes include **Ground**, **Shadow**, **Shrub**, and **Tree**, which will be classified with a random forest model.

3. Close the attribute table.

Part 3: Open and Explore RStudio

A. Orient yourself to the Graphical User Interface (GUI)

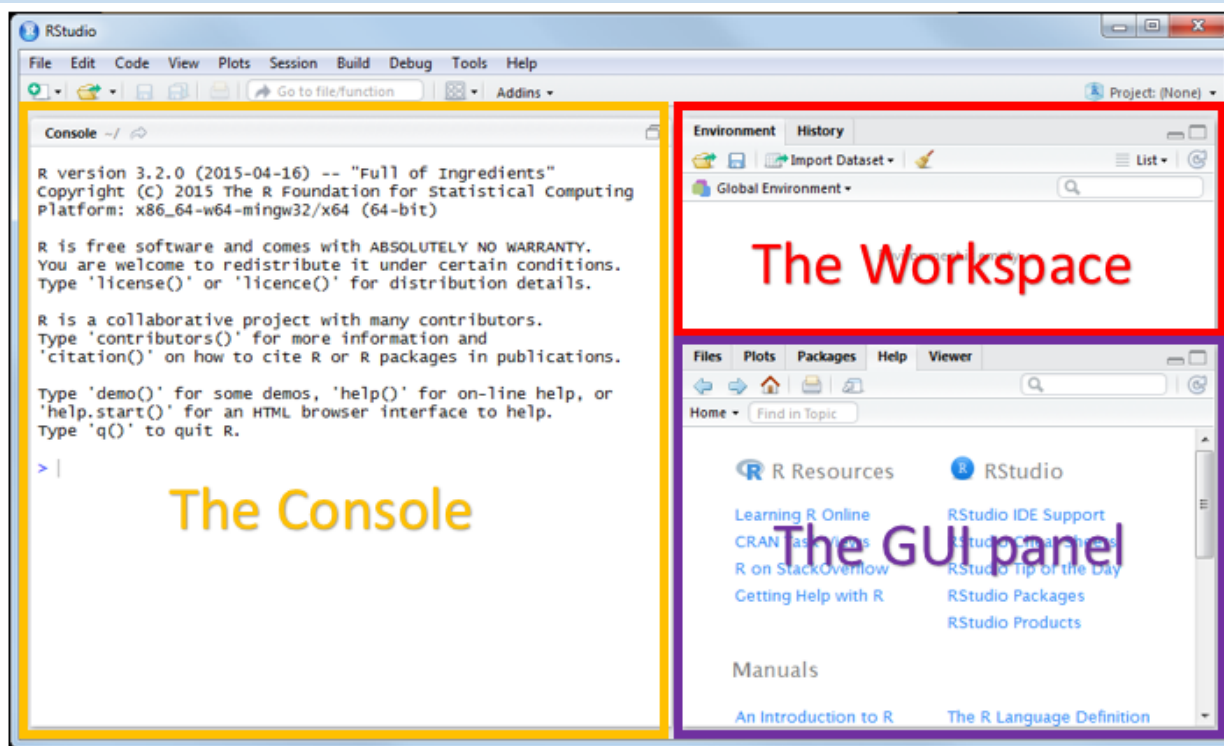
1. Launch **RStudio** by clicking the **Start Menu**, **All Programs**, **RStudio**, and then **RStudio**
2. There are three windows/pods within the RStudio GUI: The Console, The Workspace, and The GUI Panel.

RStudio is a free, open-source integrated development environment (IDE) for R. R comes with its own text editor and RStudio is not required to work in R. However, RStudio offers several convenient features for managing the R environment, making it easier to use R interactively.

The Console: the command prompt within RStudio. This is where you execute your script.

The Workspace: displays and catalogs the elements in your workspace. This helps us remember the 'what' and 'how' of our code in our scripts.

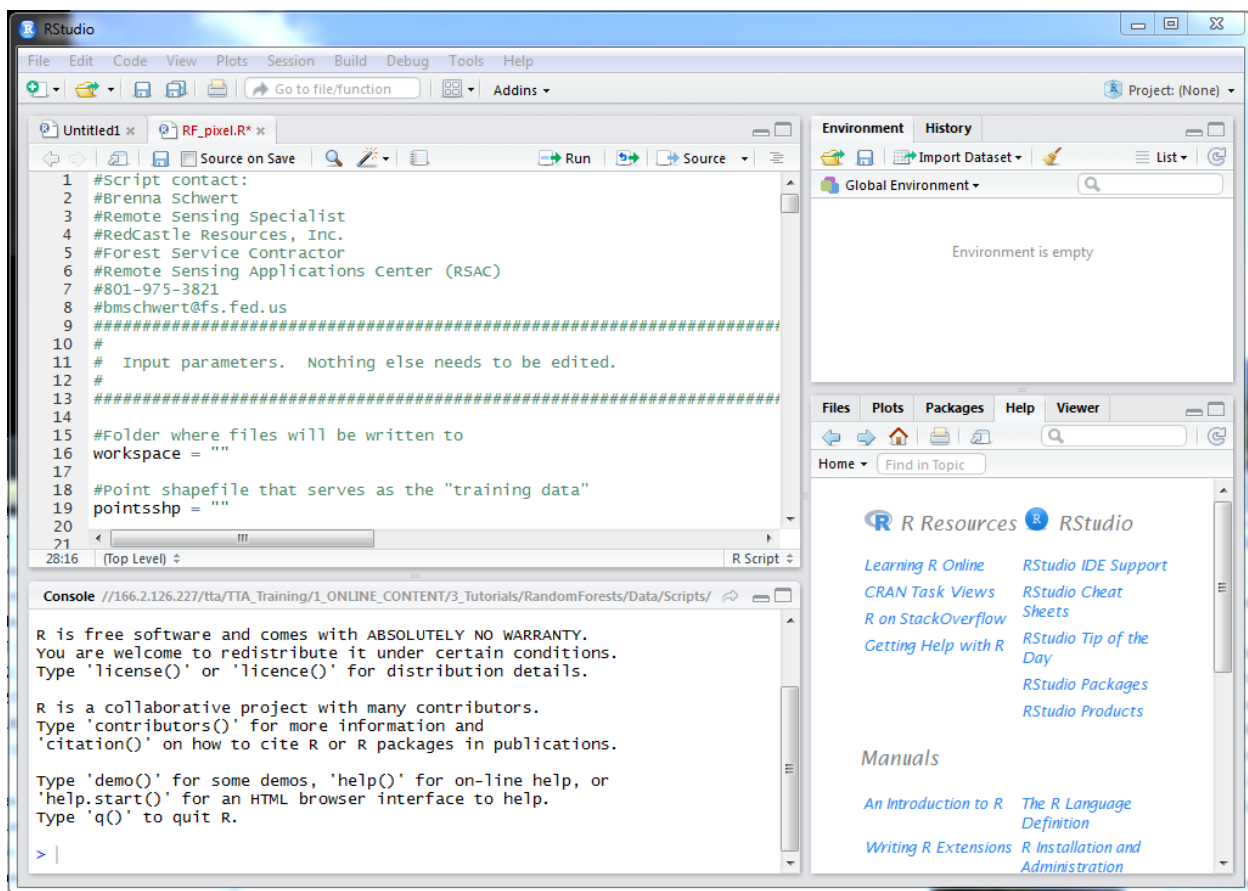
The GUI Panel: access files, plot graphics, R package, and R help. In addition, it allows you to export your graphics to the clipboard.



B. Investigate the random forest script

1. In **RStudio**, select **File** and then **Open File...**
2. Navigate to your **course folder**, double-click on the **Scripts** subfolder, select **RF_pixel.R** and click **Open**.

Note: this will load your script into a fourth window above The Console.



Note: In R, pieces of code can be commented out using the pound sign (`#`). These lines will not be read by the software, but they often provide useful information for the end user. In the above screen capture, a block of text is commented out – this is used to indicate the contact information of the person who created the script and note where the editable input variables are located.

3. Scroll down to the section after the **“Input parameters”**. This section includes the seven input variables that must be edited before running the script. The commented section above each variable describes the variable or action executed by each variable. For example, the variable **workspace** is describing the **path** and **folder** where the results will be saved.

```

13 #####
14
15 #Folder where files will be written to
16 workspace = ""
17
18 #Point shapefile that serves as the "training data"
19 pointsshp = ""
20
21 #The attribute name of the field in the point shapefile you want to predict
22 classfield = "Class"
23
24 #A value of "Thematic" or "Continuous". Whether or not you want to do classification c
25 predicttype = "Thematic"
26
27 #A list of all the continuous predictor images (e.g. Landsat)
28 imagelist = c("")
29
30 #A list of all the thematic predictor images (e.g. soils). This is commonly empty (i.e
31 thematicimagelist = c("")
32
33 #The output image location and name.
34 OutputModel = ""
35 #####

```

Note: When specifying filepath or the full filename, it is important to use the proper slash. While the backslash (\) is used for file locations, in many scripting languages the backslash is used as an escape character. If you wish to use backslashes in your filepaths or filenames, you must use double backslashes (i.e., \\). Alternatively, you can use the forward slash (/).

Congratulations! You have successfully completed this exercise. You now know the basics of navigating RStudio and you've had a chance to explore the course data. In the next exercise, you will run the script in RStudio.