Stats 133: Lab 10

April 5, 2016

In this lab, we will learn how to run .R scripts remotely on a server! This will involve uploading the .R file from your computer to the server, running the .R script in batch mode, and then retrieving the resulting .rda output back onto your local machine.

1 Task 1: Editing the .R File

1. Download the .R file called Lab10.R and save it to a directory on your computer. Notice that this file includes a function, `simpleFun()`, as well as code to call this function and code to save the return value from the call.
2. In this file, update the function call by replacing "Smith" with your last name and "10/31/2011" with your birthday.

We are now going to learn how to upload this file with R code to the Berkeley statistics servers and run it there. This can be useful in situations where your code requires more computational power than what your computer offers.

2 Task 2: Getting a Statistics Class Account

If you have yet to register for a Statistical Class Account, ask your instructor for the required form. You can't do this lab without an account. Your username and password are at the top of this form.
Follow the instructions that appear in the terminal to set up your new account. Note that your new password needs one uppercase letter and one number.

3 Task 3: SSH into the Berkeley Statistics Computing Servers.

SSH is a method that allows us to execute commands on a remote computer using your very own home computer!

To use ssh, you need to start your computer's terminal. Please use the terminal in your DataScience Toolbox virtual machine.

Once you're in the terminal, you can SSH into the Berkeley servers by using the command:

```
ssh your_account_login_name@gandalf.berkeley.edu
```

The "gandalf.berkeley.edu" is the name of the server that we are accessing. In addition to gandalf, there are several other servers, including arwen, beren, roo, and treebeard. You may use any of these servers. To learn more you can visit the page `http://statistics.berkeley.edu/computing/servers/compute`.

4 Task 4: Looking Around

Lets take a brief look at the some of the command line tools available to you.

1. Type

   ```
   ls
   ```

   into the terminal and hit enter. What do you see?

2. Use the

   ```
   cd
   ```

   command to access the Documents directory.

3. In the documents directory, use the

   ```
   mkdir
   ```

   command to make a new directory named `lab10Dir`.

4. Use the `cd` command to access this new directory.

5. In this new directory, type

   ```
   touch myFile.txt
   ```

   This command is useful for testing purposes. It creates a new file named `myFile.txt`.

6. Type `ls` into the terminal to see the new document you created.

7. Delete this document using the

   ```
   rm -i myFile.txt
   ```

   command. Note we use the option `-i` so that we are prompted to confirm we really want to remove the file.

8. Finally, type `pwd` into the terminal. This will give you your present working directory. We're going to need this in just a second so be sure you keep it handy.

5 Task 5: Uploading Your .R File to the Server

We can't upload .R files to the server directly. Instead, we need to use something called `scp` Here's how it works.

1. If you are on the server type exit to return to your DataScience toolbox terminal.

2. We need to put the file Lab10.R in your DataScience toolbox home directory. We discussed in class how to transfer a file from your local home directory to the Data Science toolbox home directory (see lecture 27).

3. Once you're in the folder that contains your Lab10.R file, use `pwd` to get the path to your Lab10.R file.

4. In the terminal, use the following command to upload your Lab10.R file to the Berkeley statistics server. You may have to change the server name from gandalf if you are using a different server. Below, we break the line into two lines for readability purposes. However, it should be entered into the terminal as a single line (don't hit enter halfway through!)

```
scp PathToLab10/Lab10.R your_account_name@gandalf.berkeley.edu:
/accounts/class/s133/your_account_name/Documents/.
```

Note the general form for scp is

```
scp pathToFileOnYourComputer/file.extension username@server:/PathToCopyFileInto
```

6 Task 6: Running Your .R File from the Server

Lets go back to the terminal window that's still running on the server. If you type `ls`, you should now see a copy of your Lab10.R. Luckily, running this file on the server is easy. Simply execute the command

```
nice R CMD BATCH Lab10.R outfile.Rout &
```

This will run the Lab10.R file in batch mode. The "nice" argument at the beginning tells the server the job is low priority (don't interrupt other people). The ampersand at the end of the expression is saying "execute this task in the background." That way, you can still use the terminal while your script is running!

Once the script finishes (if it takes more than one second something is wrong), you can type `ls` to see the output of your Lab10.R file. In particular, you should now have `lab10Results.rda` in your working directory. In addition, you should have the file `outfile.Rout` in your working directory. Examine the contents of this file using the `more` command,

```
more outfile.Rout
```

Notice that it contains the printout of an R session. The last lines of this file tell you the time it took to run the code. It's good to check this file to confirm that the code ran as expected.

7 Task 7: Pulling Your Results Back to Your Local Machine

Lets finish by retrieving a copy of the `lab10.rda` file onto your local machine from the server. This can also be accomplished via scp. The general command takes the form

```
scp username@server:/PathToFileOnServer/myFile.ext /PathToCopyIntoOnMyMachine
```

In your case, this will look something like

```
scp your_user_name@gandalf.berkeley.edu:
/accounts/class/s133/your_user_name/Documents/lab10.rda ~/.
```

Once you have successfully retrieved this file from the server, be sure to upload it to b-courses. This upload will count as your submission for the lab.

If you wish to view your lab10.rda type `load("~/lab10.rda")` to load the file in your global environment. You can then type `myData` to view the list.

Since you're done running commands directly on the server, you can now close your connection to the server by typing `exit` and hitting `enter` on the terminal window that is running the server. You can now close your terminal windows as well.