Introduction to FarmShare for Statistical Packages (Mac)

This guide is intended to help first-time users understand how to remotely access statistical software programs on Stanford's FarmShare from a Mac.

FarmShare is one of Stanford's shared research computing environments. It allows users to access statistical and other software programs at no cost on their own computers through the internet. It also connects them to Stanford's powerful computers for much faster processing of computing-intensive tasks than would be possible on single desktop or laptop. This document explains how to access FarmShare through Mac's built-in UNIX command line in Terminal or through X-Windows on your Mac to log in to your Stanford AFS account.

To use FarmShare, you must have a SUNet ID (Stanford University Network Identifier). If you do not have a SUNetID, visit www.stanford.edu/services/sunetid for more information on obtaining one.

Table of Contents

Quick Start: Accessing Software in FarmShare

As an overview, the steps to using software (in this example, Stata) are:

1. Download and install XQuartz
2. Upload files you want to access in the software to your AFS space (afs.stanford.edu)
3. Open XQuartz
4. Log on to Farmshare by typing the following in XQuartz. Instead of ***user***, type your SUNet ID. You will need to do 2-step authentication.
    - ssh -X ***user***@rice.stanford.edu
5. In XQuartz, type the following commands to begin Stata:
    - module load stata-se/15
    - xstata&

The document below provides more detailed explanations and steps.

Introduction

Why would I use FarmShare for statistical packages?

- You want to access statistical software for *free*.
- You want to access statistical software *remotely*, rather than from your dorm or library computer cluster.
- You need a *more powerful version of statistical software*. Your dataset includes too many variables to open in your Stata IC (you need Stata SE/MP) on your personal computer.
- The tasks you need to perform are *computationally intensive* and take up all of your person computer's resources (leaving you unable to watch that silly cat video your mom sent you while also running complex algorithms in Stata).

In this document, words appearing in **bold type** are commands in Unix or a statistical software package and should be typed as shown. Conversely, words that are *italicized* are not supposed to be typed verbatim; instead, you are to substitute another word for them. For example, when you see *filename* in this document, you should substitute the actual filename you would like your file to have. It is also important to remember that Stata, R, and Unix are case sensitive, though SAS is not. In Stata, R, and Unix, name1, Name1, and NAME1 are all different.

Installation of Software for Using FarmShare

The primary software you need to install for using FarmShare on Mac is a so called X Window System. The X Window System (also X-Windows or 'X') is an interactive graphical computer environment that runs on OS X and UNIX-like computer operating system. Therefore, it is straightforward to get X working on a Mac. Before you begin, you may want to check which version of OS X is on your Mac by clicking the Apple menu and selecting 'About This Mac'. The OS X version determines how you will access the X-Windows utilities.

If you do not have XQuartz or another implementation of X-Windows installed on your computer, download the XQuartz software here (http://xquartz.macosforge.org/landing/).

Additional useful software for working with FarmShare, such as the Kerberos Commander and Stanford OpenAFS for Mac, is available at Essential Stanford Software (EES): https://itservices.stanford.edu/service/ess/mac.

FarmShare Environments

When accessing FarmShare, users have a choice of which environment to use with different levels of computing power and different capabilities. Cardinal, which has the least amount of computing power of the FarmShare environments, is for low-intensity tasks such as checking email and so is not recommended for analyzing data.

Rice is likely the most commonly used for data analysis. It has the advantage of offering a good amount of computing power but also allows users to access a larger range of statistical software packages (SAS, Stata, R) than is possible in Wheat or Oat. It also works for interactive (where users type and submit one line of syntax at a time and view the results as they go) as well as non-interactive jobs (where users type and submit all syntax at once). Additionally, rice offers an x-simulator through X-Windows, which features a graphical user interface among others for Stata, SAS, and R users. The graphical user interface versions of statistical software programs have multiple windows and so have a similar feel to how the programs operate on a personal computer or a Mac.

For more information on FarmShare environments:
https://itservices.stanford.edu/service/sharedcomputing/environments

Accessing Files

Because FarmShare is not on your local computer, you need a method to access your files while in the FarmShare environment and then retrieve finished files from the FarmShare session to bring to your local computer. The simplest approach is to use Stanford's AFS (Andrew File System) space. AFS is a file system that enables efficient file sharing between clients and servers. AFS files are accessible via the Web (afs.stanford.edu) or through file transfer programs such as OpenAFS (see SSDS guide ***Introduction to AFS*** for more details).

The first step is to using FarmShare in any capacity is to copy your relevant files (datasets, command scripts, etc.) into your AFS space by going to afs.stanford.edu and using the appropriate menu items to copy or upload your files.

Two Methods of Using FarmShare

There are two primary ways of interacting with FarmShare:

(1) Interactively through an X-simulator or
(2) Submitting code as "batch jobs".

Using software interactively though X enables the use of statistical program's familiar Graphical User Interface (GUI) to run new analyses and view output in real time. For non-computationally-intensive tasks, this is the most user-friendly.

Submitting code as "batch jobs" for statistical software involves users submitting the full script (e.g., Stata's .do files) of the entire analysis in advance, and after submitting the "job" to FarmShare's servers, users wait for the script to conclude running and view output files. This method of using FarmShare is useful for computationally-intensive tasks for the ability to take advantage of parallel processing and not being required to stay connected to FarmShare after submitting the job code.

(1) Accessing FarmShare in Interactive Mode

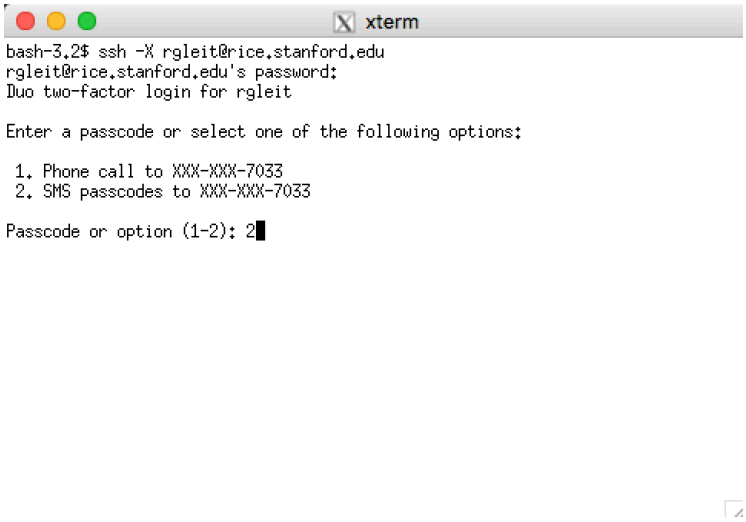
Logging in to FarmShare

You will access FarmShare using **XQuartz**. If you are running Mac OS 10.5 "Leopard" or later versions, you can open **XQuartz** from the **Applications/Utilities** menu under the **Go** menu (or simply by typing **XQuartz** in Spotlight). Once XQuartz is open, you will log into a Stanford Mac by using a command and your SUNet ID. Type:

ssh -X *your_SUNet_ID*@**rice.stanford.edu** (note the space after '**ssh**' and '**-X**')

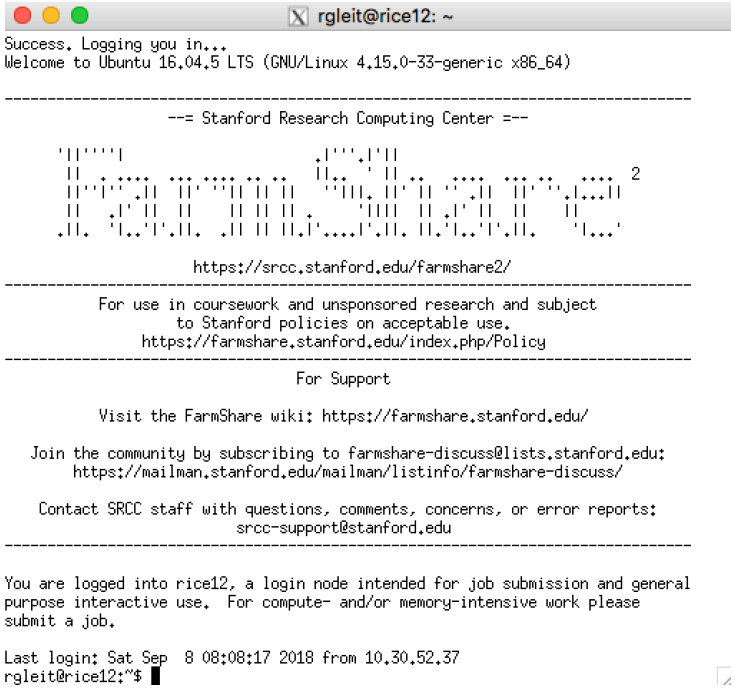
This procedure will connect you to Stanford's rice servers. (See section on ***FarmShare Environments*** above. If you need more computing power, also see the separate SSDS document for ***Using Stata and R in Wheat or Oat***)

Once you have typed in the command, you will be asked to enter your SUNet password. When typing your password, text will not appear in the terminal window. The connection is encrypted and secure. After that, type the number of your preferred two-step authentication method (e.g., 1 = Phone call, 2 = SMS passcode) and follow the corresponding instructions.



Side note: To avoid the two-step authentication method every time you connect to FarmShare, follow the instructions under *Duo Two-Factor* here. To change your ssh_config file, open Finder, press Shift + Command + G (*Go To*). Enter the path **/etc/ssh/ssh_config**, open the file ssh_config, and save changes to the same document.

The terminal and command should look like this:



Loading Software into a FarmShare Session

Once in FarmShare, you will want to access the statistical software program of your choice. For Stata and SAS, which are not open source programs like R is, you will need to start by loading the program. The first step to loading a program is to type

module avail

at the FarmShare prompt and hit enter on the keyboard. You should see a list of available programs appear on your screen. Locate the program you would like to use and then type

module load *program_name*

where *program_name* is the name of the desired program exactly as it appears. Then hit enter. Note that the FarmShare environment is case sensitive, so you will need to use the same upper-case and lower-case letters in the name of the program as it appears on your screen.

To pull up a list of programs that you have loaded into your FarmShare session, type

module list

and hit enter. To have a set of modules loaded automatically every time you start a session, load the modules you want to use and type:

module save

```
rgleit@rice03: ~

rgleit@rice03:~$ module load matlab/r2018a stata-se/15 sas/9.4
rgleit@rice03:~$ module list

Currently Loaded Modules:
  1) matlab/r2018a   2) stata-se/15   3) sas/9.4


rgleit@rice03:~$ module save
Saved current collection of modules to: default
rgleit@rice03:~$
```

The modules you save will be automatically loaded into your future sessions. If some modules are still unavailable in your next session, type **module update**.

To look for a particular version of a program use '**module spider** *program*' and load the software version you want to use. For example, to see all the available versions of Stata type:

module spider *stata*

```
rgleit@rice03: ~
rgleit@rice03:~$ module spider stata

-----------------------------------------------------------------------------
  stata-mp:
-----------------------------------------------------------------------------
    Description:
      Integrated statistical software package for data analysis, data
      management, and graphics.

     Versions:
        stata-mp/14.2
        stata-mp/15

-----------------------------------------------------------------------------
  To find detailed information about stata-mp please enter the full name.
  For example:

     $ module spider stata-mp/15
-----------------------------------------------------------------------------

-----------------------------------------------------------------------------
  stata-se:
-----------------------------------------------------------------------------
    Description:
      Integrated statistical software package for data analysis, data
      management, and graphics.

     Versions:
        stata-se/14.2
        stata-se/15

-----------------------------------------------------------------------------
  To find detailed information about stata-se please enter the full name.
  For example:

     $ module spider stata-se/15
-----------------------------------------------------------------------------


rgleit@rice03:~$
```

To unload all the modules loaded into your session type:

module purge

Once your software is loaded (see **module load** above), there are two ways to start using the program.

(A) To use without the graphical interface:

Simply type the name of the program on the command prompt (*R, stata, matlab,* etc) to start using the program. For example:

```
rgleit@rice03: ~
rgleit@rice03:~$ module load stata-se/15
rgleit@rice03:~$ stata

  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

25-user Stata network perpetual license:
       Serial number:  401506258593
         Licensed to:  Software Licensing
                       Stanford University

Notes:
     1.  Unicode is supported; see help unicode_advice.

.
```

Once a program is started, you can type software-specific commands directly on the command prompt. Note that, for most of the programs, the default working directory will be your AFS space at Stanford (see ***Accessing Files*** section above). That is, the software you are using will be searching for files and exporting output from and to your AFS space. Make sure that all the files and scripts for your session are uploaded into your AFS folder, which you can access at afs.stanford.edu. You can also change the working directory using the corresponding commands for each software.

(B) Or, to use the interactive version with graphical user interface:

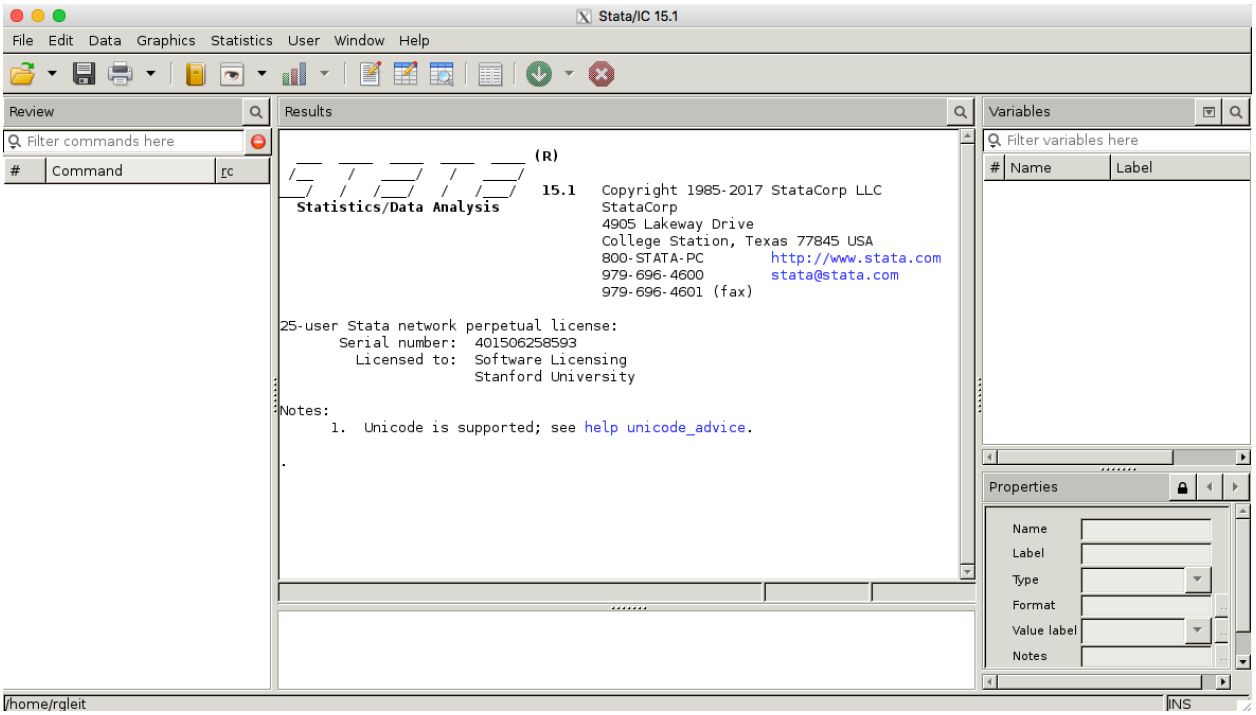
Type the name of the program with an x preceding it. Only a few programs have an X-Windows version available. Here are some of them:

Software	Command
Calculator	*xcalc*
Stata	*xstata*
SAS	*sas*

Matlab	*matlab*

Side note: We suggest adding '&' to the end of the program name (e.g. *xcalc&*). The ampersand (&) tells a program to run in the background. This is an optional function, but without it, the Console window is "tied up" (unusable) until the xterm program had been terminated. If you want to run the X-Windows version of Stata, type '*xstata&*'.

For example, the command '*xstata&*' launches a graphical user interface that gives you the same windows as if you had STATA installed on your personal computer.



Remember to leave your X terminal window open while using X-Windows. If you close the terminal, you will terminate your software session.

Once you are done, exit the statistical software via the drop-down menu (if using GUI), and close the terminal by typing '**exit**'.

(2) Accessing FarmShare in Batch Mode

Working with emacs

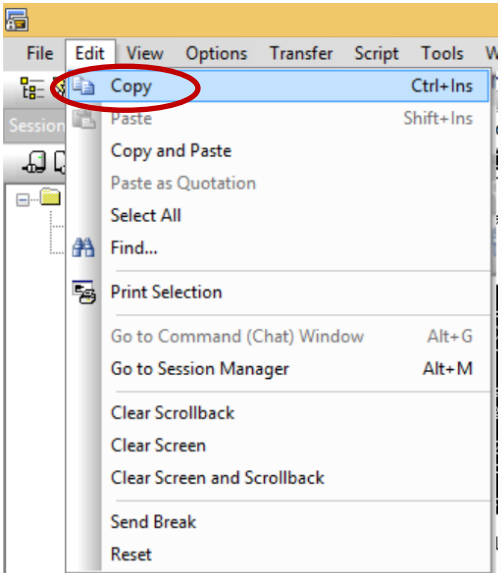
To use statistical software programs in batch mode, you will need to type all of the commands you wish to run in the program into a text editor and then save the commands as a file that you will then submit to the program. This section gives tips for working in the emacs text editor within FarmShare.

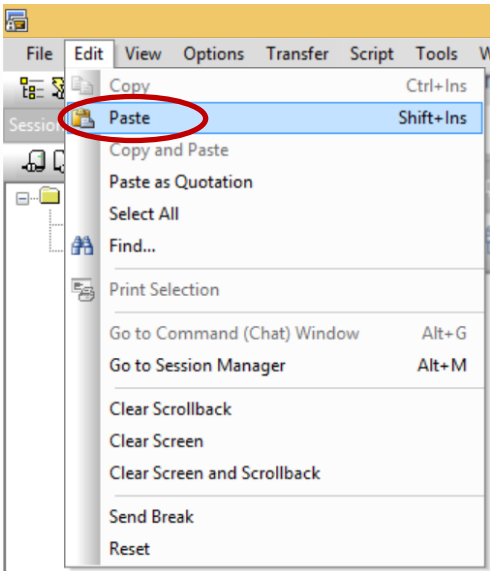
To create a new file or to open an existing file in emacs, at the FarmShare prompt type

emacs "*filename*"

where "*filename*" is the name of the file you are creating or would like to open. Then hit enter. Note: The quotation marks are only necessary if the name of the file has spaces in it; if the file name has no spaces, you can simply type it without the quotation marks.

Once the file is open, you can type commands into it, but there are some quirks to working within emacs that make it different from working in a program like Microsoft Word. First, if you need to delete something you just typed (i.e., delete characters and spaces in a backward motion from where the cursor is), hit delete on your keyboard. The backspace key will not have the intended results. Additionally, the shortcut keys for copying and pasting do not work within emacs. Instead, after selecting the section you want to copy, go to the **Edit** drop-down menu at the top left of the FarmShare screen and select **Copy**, and then move your cursor to where you want the copied text to appear and select **Paste** from the **Edit** menu, as shown here:





When you are done working with your file in emacs, save your changes and exit emacs (but not FarmShare).

When you do so, you will see a prompt to save changes to your file if it contains unsaved work. Follow the directions in the prompt to either save or discard changes.

For more tips on working with emacs, see the emacs Cheat Sheet here:
https://ccrma.stanford.edu/guides/package/emacs/emacs.html

Submitting files with code to statistical software programs

The following table gives the Slurm syntax to submit a file named *filename* to R, SAS, and Stata. Note that *filename* must not contain spaces (so name the file *file_name* rather than *file name*). Type the specified syntax at the FarmShare prompt when you are in the folder/directory in FarmShare that contains the file you wish to submit. (See the section below titled ***Common Slurm Commands*** for directions to change folders/directories in FarmShare.) For the syntax below to work for SAS and Stata, you will need to have loaded the needed program into your FarmShare session. (See the section above titled ***Loading Software into a FarmShare Session***.) You will also need to load R into your FarmShare session if you wish to use the newest version available (v3.3.0) but not if you would like to use the default version of R (v3.0.2).

Program	Syntax	Notes
SAS	**sas –memsize 0** *filename* **&**	The **memsize** option allows SAS to use the amount of memory needed for the job, even if this amount is greater than the default memory size allowed for SAS jobs in the system.
Stata	**stata –b do** *filename* **&**	This syntax assumes that *filename* has the extension **.do**; when creating new files with a text editor, make sure to give them the **.do** extension in the filename (e.g., *filename.do*).
R	**R --vanilla --no-save <** *filename*	The **--vanilla** part of the syntax specifies that R will be used in batch mode instead of interactive mode.

To loop through a program's command line coding file multiple times, you will also need to create a ".sbatch" file, which tell servers to open "module" (program) and run script (e.g., .do file) on dataset multiple times. You can tell servers to execute submit file with code:

rice02> **sbatch** *example.sbatch*

To loop through submitting this X times, enter:

rice02> **for i in {1..***X***}; do**

sbatch *example.submit*

done

Common Slurm Commands

This section focuses on basic commands in Slurm to help you navigate once in FarmShare, and the subsequent section will go over how to access statistical software packages in batch and interactive mode in FarmShare.

Once you are in FarmShare, you will need to use some Slurm commands to work with files and folders where you store your work. Files, such as data files, that you need to access in your FarmShare session should be saved in your AFS space. When you first connect to FarmShare, you will be in your home directory, or folder, in your AFS space and will have access to the files stored there. The commands below allow you to switch folders and manage your files. To use the commands, type them as they appear, substituting relevant folder and file names, and then hit enter on your keyboard. Note that Slurm is case sensitive, so you must type the name of folders and files exactly as they are. Both the commands and their results are listed in the table below.

Slurm commands for working with folders and files in FarmShare	
Command	**Result**
cd "*folder name*"	Navigates to "*folder name*" and grants access to the files it contains
cd ~	Navigates to home directory (the home folder in your AFS space)
ls	Pulls up a list of files in the current folder
ls -l	Pulls up a list of files in the current folder with details such as the files' dates
cp "*original filename*" "*copied filename*"	Creates a copy of "*original filename*" and gives it the name "*copied filename*"
rm "*filename*"	Deletes "*filename*"

Note: The commands above to navigate to folders are for the rice session within FarmShare. If applicable, you will also need to specify within the statistical software package that you use (SAS, R, or Stata) which directory or folder you would like to work in.

If you are new to working with your AFS, you can read more by clicking on the link for Introduction to AFS on this page: https://ssds.stanford.edu/software-resources/getting-started-guides-documents

For More Information and Assistance

The IT services at Stanford have a collection of webpages that provide information on FarmShare, including what it is and how to access it. You can find them here: https://itservices.stanford.edu/service/sharedcomputing

The FarmShare wiki page is available at (including tutorial):

https://srcc.stanford.edu/farmshare2

You can get e-mail support at research-computing-support@stanford.edu

You can also get help from the Stat, Math, Algorithmic and Computational Consulting (SMACC) group at Stanford here: http://stanford.edu/~rezab/smacc/

Accessing FarmShare on a PC

To access software on a PC, see the document ***Introduction to FarmShare for Statistical Packages (PC)***.

SSDS Software Services at Stanford

If you have questions about using statistical software packages in a FarmShare environment, please contact the software consultants at Social Science Data and Software. Our website is http://ssds.stanford.edu. The software consultants are available during the academic year on a walk-in basis. Please see our website for current walk-in hours.

Social Science Data and Software

Document revised: 9/8/18