MAP UNIT SUMMARY REPORT - PART 2 - INSTRUCTIONS

8/3/2016

Dylan Beaudette, Digital Soil Mapping Specialist, SSR-2, Sonora, CA

Jennifer Wood, Soil Data Quality Specialist, SSR-2, Davis, CA

Russ Almaraz, GIS Specialist, SSR-2, Davis, CA

# PART 2 - MAP UNIT SUMMARY REPORT TUTORIAL

*Tip*: Make this document easier to use by checking the Navigation Pane box under the View menu

# **Step 1.** For background information, read document titled “Map Unit Summary Report – Part 1 – Background”

# Step 2. Set up R Studio

You will be running the report in R Studio, which you already have loaded on your computers. The file that contains the script that generates the report is provided to you as an “R Markdown” file, titled “GIS-summary-by-MU.Rmd” file. Two additional files are included with the report to assist with setting up the environment and to point the report to the spatial data: “load-packages.R” and “config.R”.

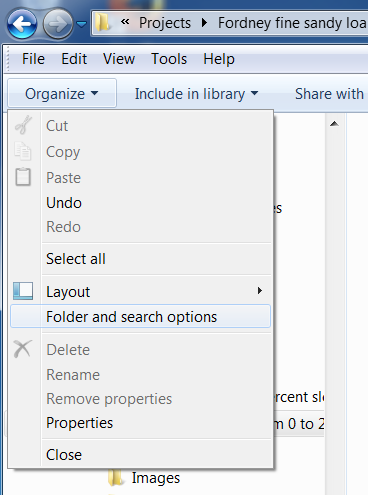
To set up your R Studio environment for the first time, read and follow the instructions in [Appendix 1](#_Appendix_1_-).

# Step 3. Prepare your files and folders

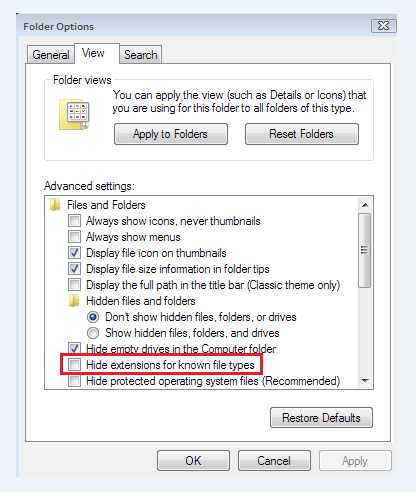
## Change your folder options in Windows Explorer to show file extensions

This will help you manage the different file types you will be using.

In Windows Explorer -> Organize menu -> Folder and Search Options



In View tab, uncheck the box to hide file extensions



## Files provided to you in folders and located in the suggested office shared drive folder location

This section will specify where the map unit summary report data layers and R files are to be located in the geodata file structure on your office shared drive. This fixed location is to ensure easy access for technical support and for remotely installing file updates. See [Appendix 2](#_Appendix_2_–_1) for a summary of the intended file structure.

### Overview

* All data folders for the map unit summary report are labeled as MUSum\_\*, such as MUSum\_PRISM or MUSum\_10M\_SSR2.
* The abbreviation “MUSum” is spoken as “M U Sum”, for “map unit summary”
* Master copies of the MUSum data folders will be stored in your office shared data drive, in the “project\_data” folder of the “geodata” folder.
* All master files in the MUSum\_\* named folders and their associated files are considered to be “read only”. No alterations are to be made without authorization of the RO.
* The MUSum DEMs and derivatives, due to their integer format, are appropriate for the MUSum Model only. Other files are suitable for other uses, such as the PRISM\_800m and MLRA hillshades. These are to be located in the proper thematic folder in the “geodata” directory. As a note, the PRISM\_800m folder is an exact copy of the MUSum\_PRISM folder.

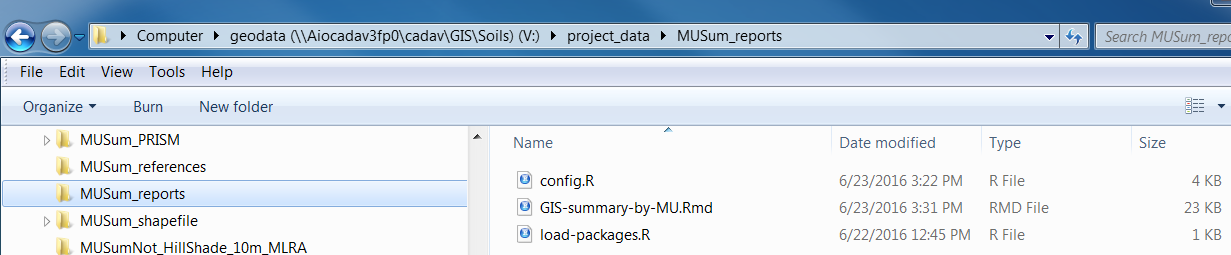
### R files

Store the master copies of these files according to the following file structure on your shared office drive. These files are not set up for simultaneous usage. These are small files, so for active use they will be copied into project folders for custom project configuration.

Report document: geodata/project\_data/MUSum\_reports/GIS-summary-by-MU.Rmd

Report configuration document geodata/project\_data/MUSum\_reports/config.R

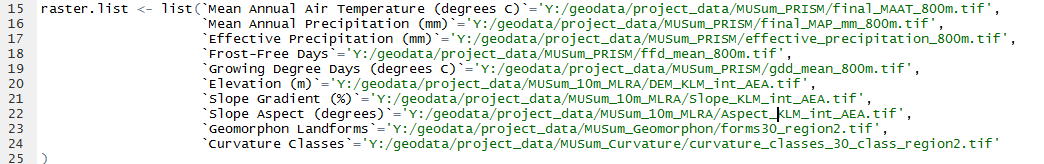
RStudio package loading document geodata/project\_data/MUSum\_reports/load-packages.R



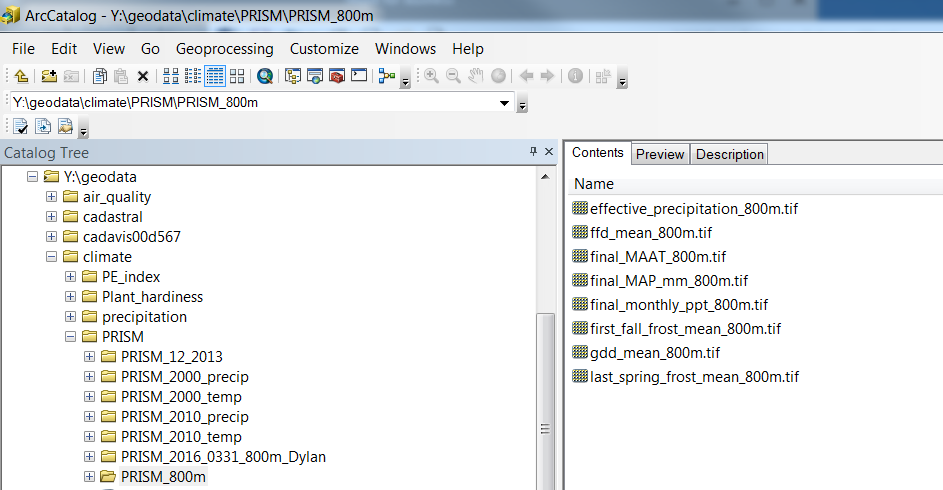
### Raster files

Store all MUSum\_\* folders on the office shared drive that contains the “geodata” folder, in the “project\_data: folder. The “config.R” file (as distributed) includes references to this folder structure and should only require a change of drive letter to use the data from the shared drive. Optionally, if the user wants to run the report using data on their local disk, the entire set of folders can be copied onto their local disk and the drive and folder locations would be edited accordingly. The report runs faster from copies on the local disk.

From the “config.R” file:



All of these files have been optimized for use by this R-based Map Unit Summary report. While it is not advisable to use the DEMs or derivatives in other applications, we do encourage you to use the PRISM products in other applications. So that others can find the files for other applications, please copy the 800m PRISM files into a new shared drive folder titled “geodata/climate/PRISM/PRISM\_800m”.



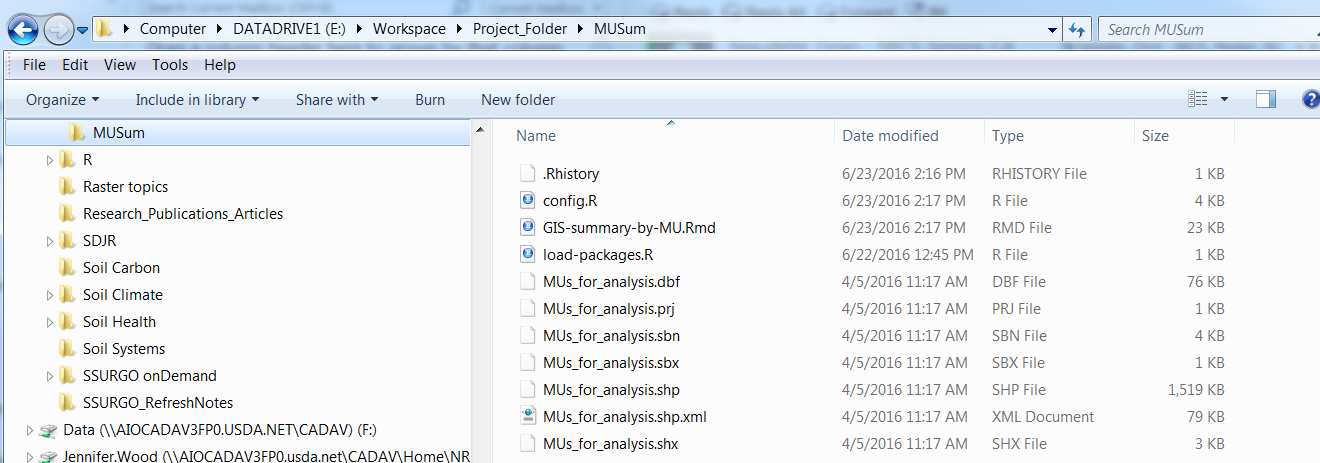
## Files provided by you

When running the report, you will be specifying a working directory that will be used to store output from the report. For the report to run correctly, the report file itself and the data configuration file should also be in the working directory.

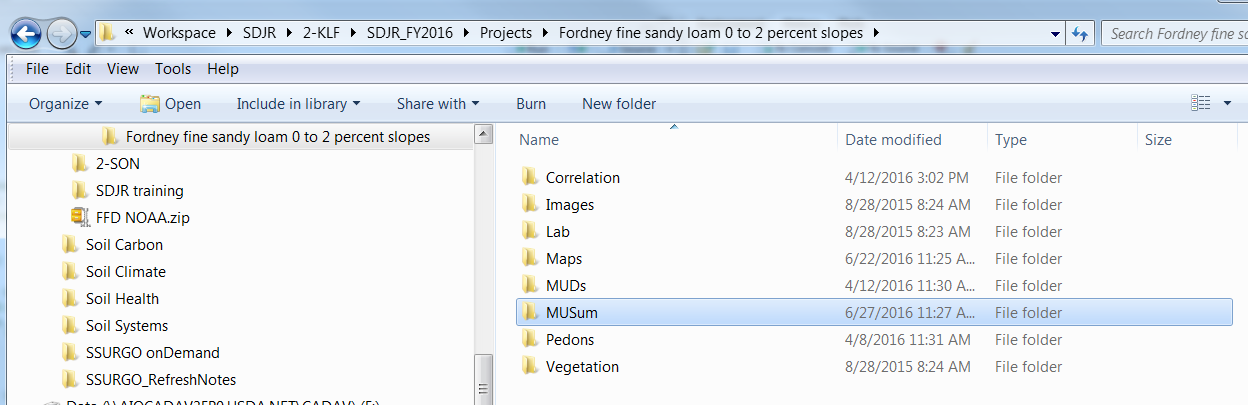
Copy all files in the MUSum\_reports folder into a folder titled MUSum (spoken as “M U Sum”, for “map unit summary”) that you create in your Project folder.

Copy your input map unit polygon data (.shp or .gdb) into the MUSum folder in your Project folder

In the following example, and as written into the “config.R” file, the “MUSum” folder is created within “Workspace/Project\_Folder” and stored on a local drive. The report will create an “output” folder in the working directory in which it will place output files. You can put your map unit shape file here or you can direct the script to where your shape file is located. In this example, and as written into the “config.R” file, the shape file is named “MUs\_for\_analysis”. You will edit the script in the config.R file to direct the report to the data location.



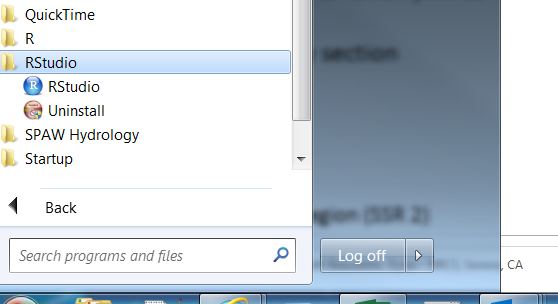
The path to your Project folder might look more like this:



See [Appendix 2](#_Appendix_2_–_1) for a summary of the intended file structure.

# Step 4. Run the report

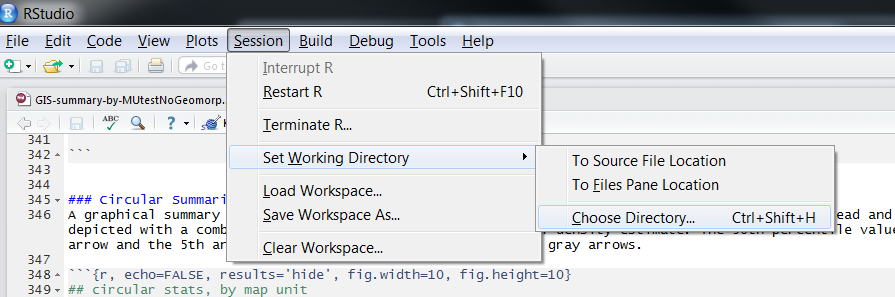
Open R Studio: Start – Programs – RStudio folder – Rstudio



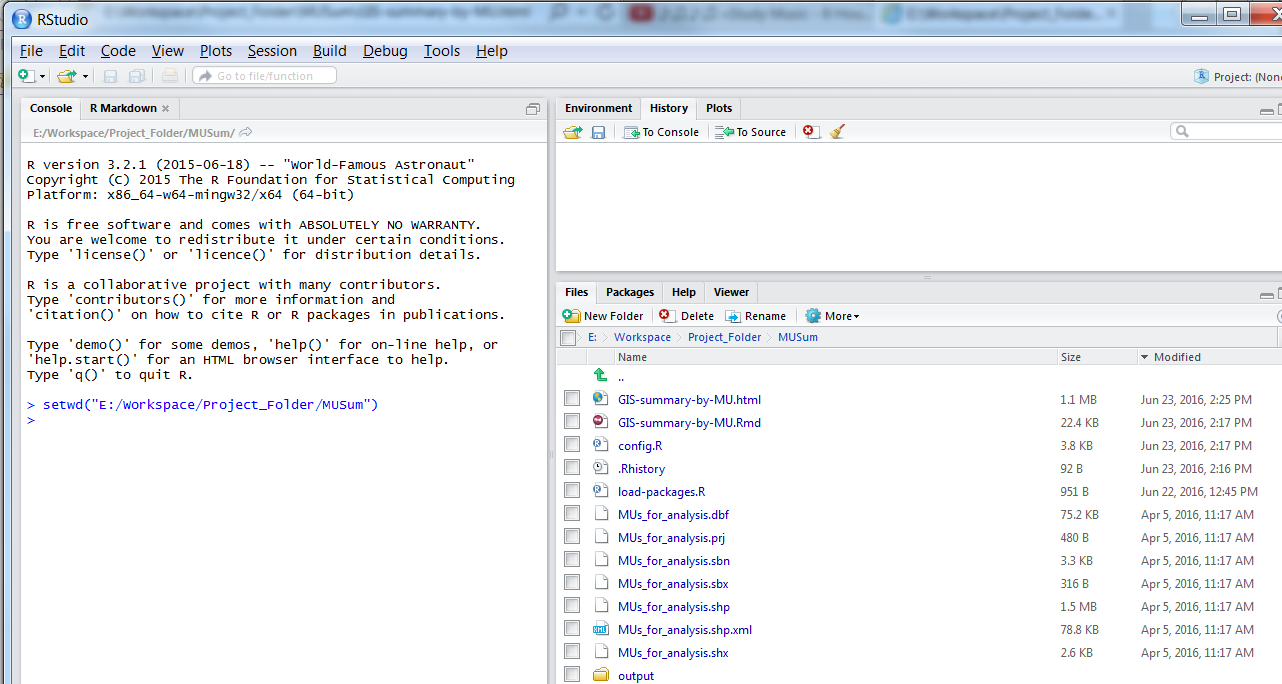
## Set the working directory

On the R Studio toolbar: Session – Set Working Directory- Choose Directory

Set the working directory to a folder titled MUSum that you place in your project folder, and in which you previously placed the GIS-summary-by-MU.Rmd, the config.R, and the load-packages.R files.



Once you set your Working Directory, the files will show up in the lower right window, under the Files tab.

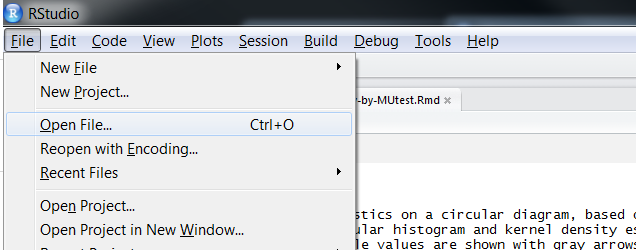


*Tip*: R studio will set the working directory to the folder where your GIS-summary-by-MU.Rmd file is located, if you open it from Windows Explorer.

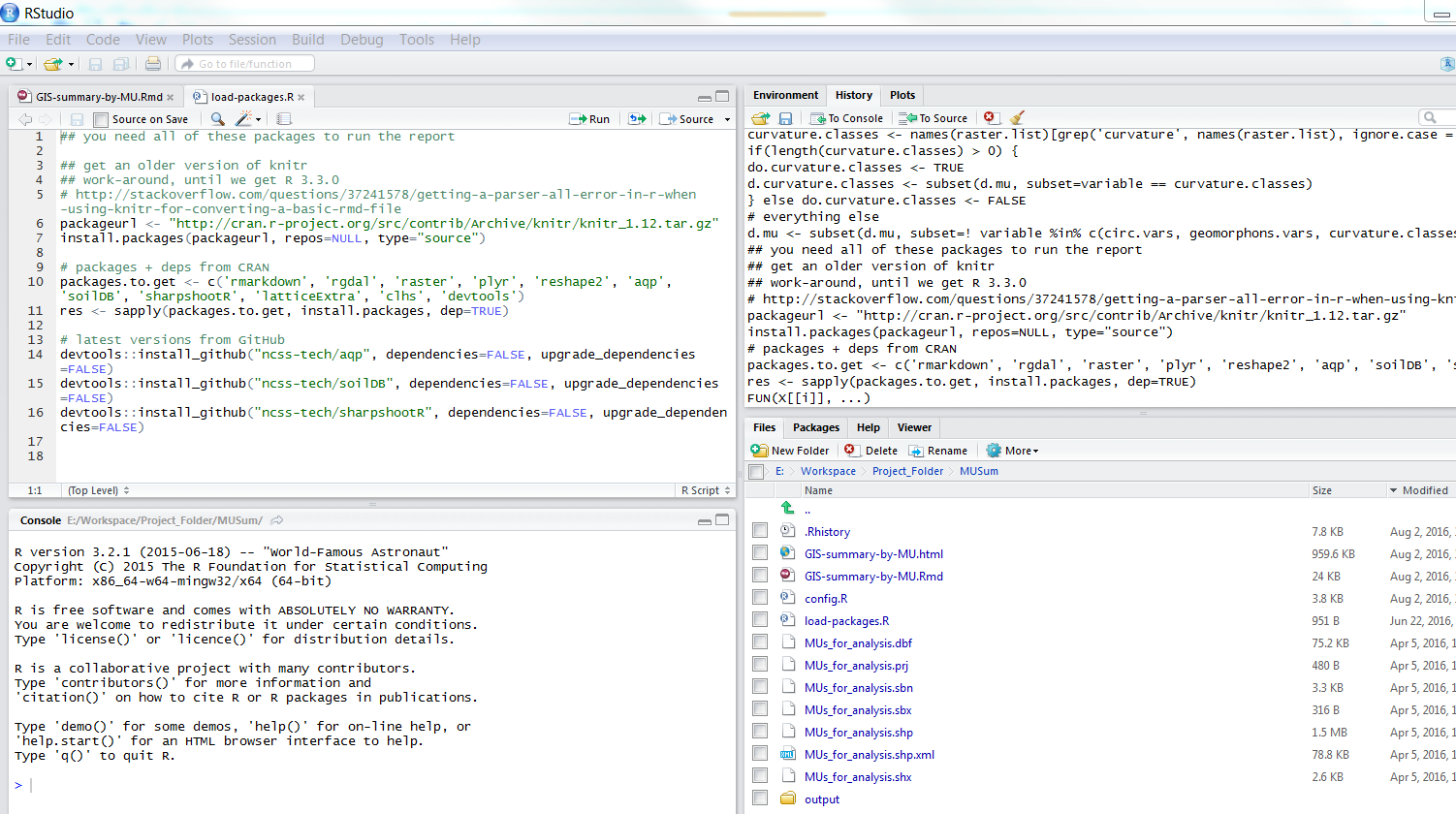
## Load the R packages that you will need to run the report

Open the file ‘load-packages.R’ from either File-Open or from clicking on it in the Files window. *You only have to do this one time when you first set up RStudio on your computer unless a new version of R and RStudio are installed on your computer. Also, do not update any packages from the Packages tab in R Studio – the report is designed to use packages that only work with the current version of R that is installed on our computers.*

On the R Studio toolbar – File - Open File

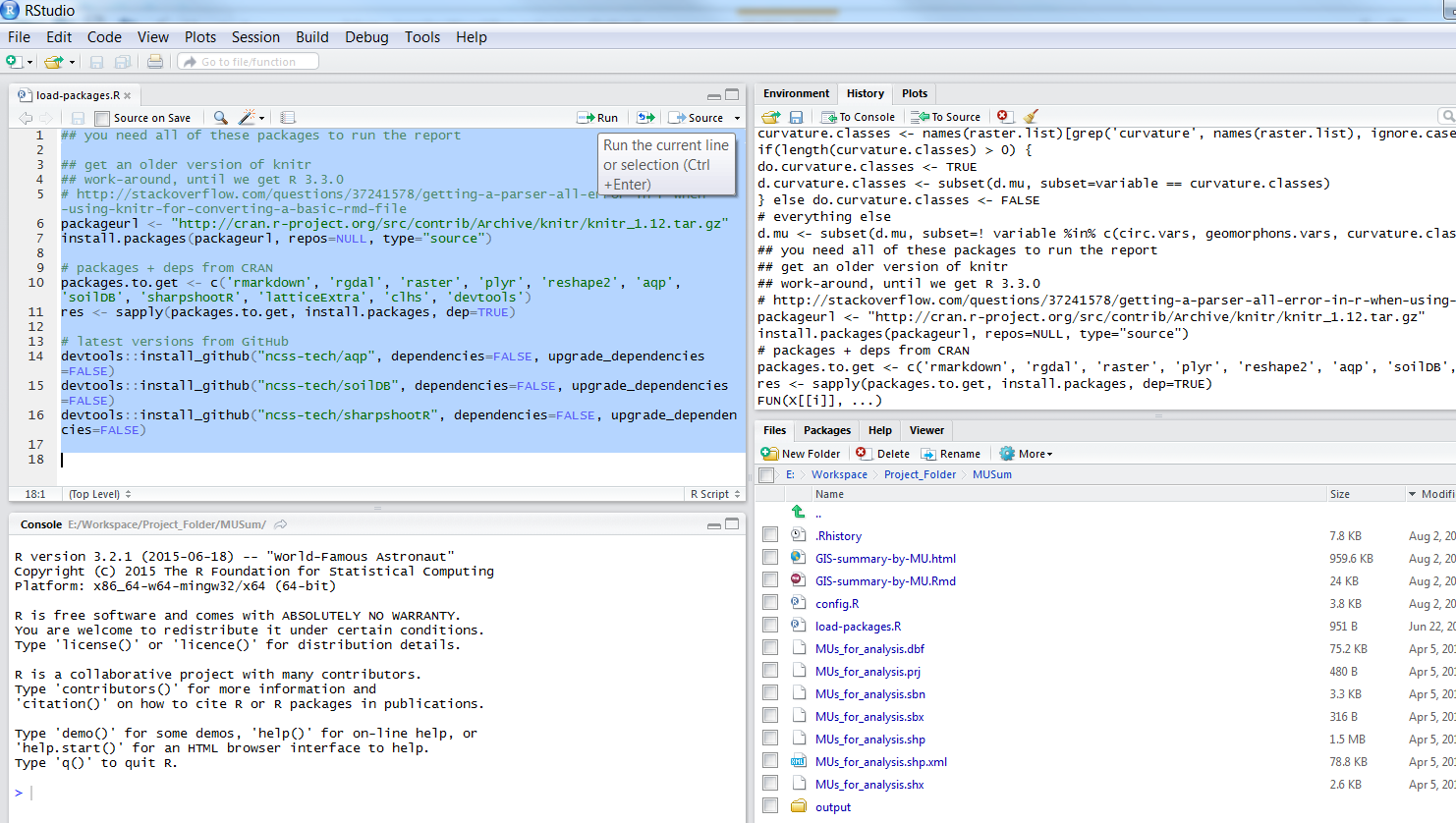


Or, you can open it by clicking on it in the Files window

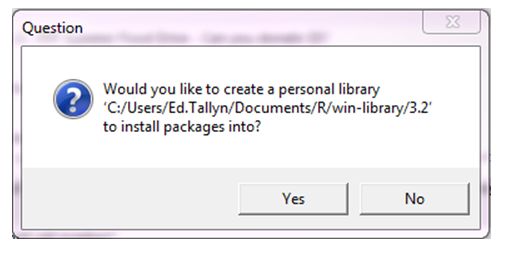


It will open in the ‘Script Window’.

Highlight the whole block of script and press Ctrl-Enter or the Run button on the file toolbar. This will begin loading packages from the internet and could take up to 5 minutes. When it is done, you can close the file.



You might get a window that gives you this message – choose Yes. This window can get buried under other open windows, so if nothing seems to be happening, minimize windows to see if it is there.

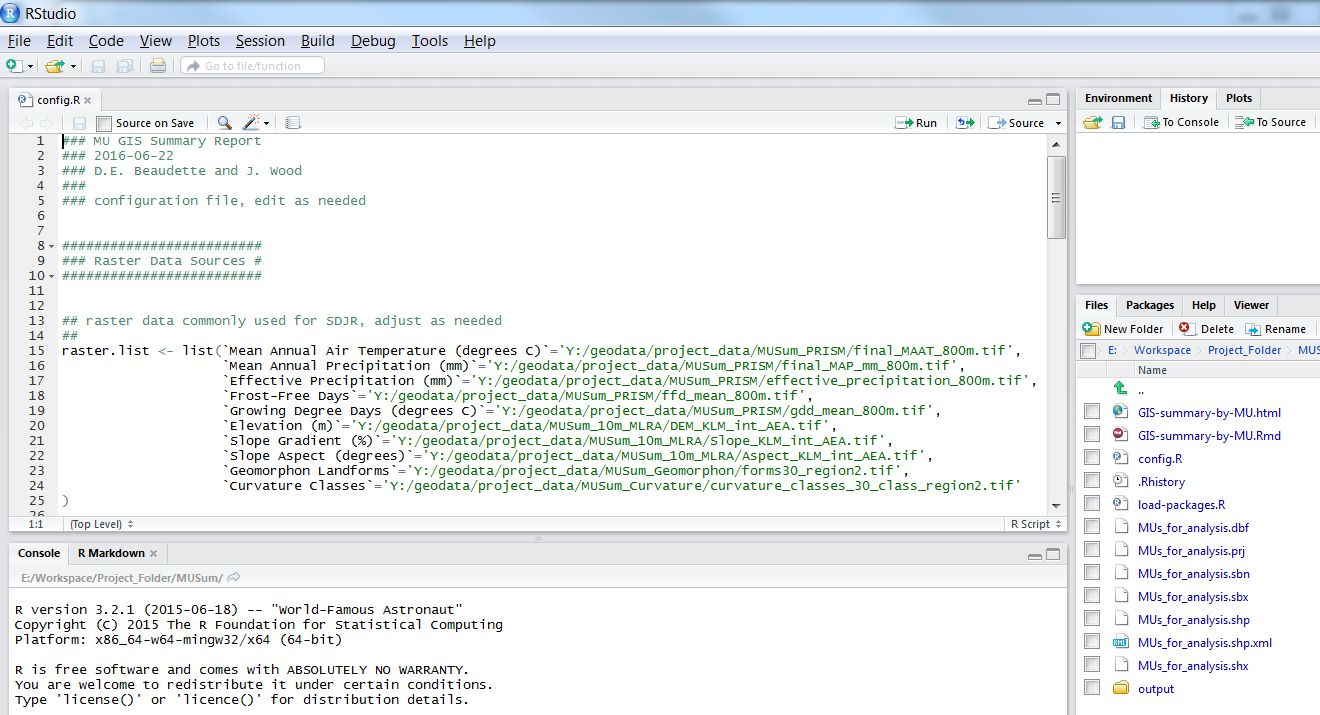


## Set the input data (rasters) in the report using the config.R file

The raster files are the environmental data you are interested in summarizing for your map units.

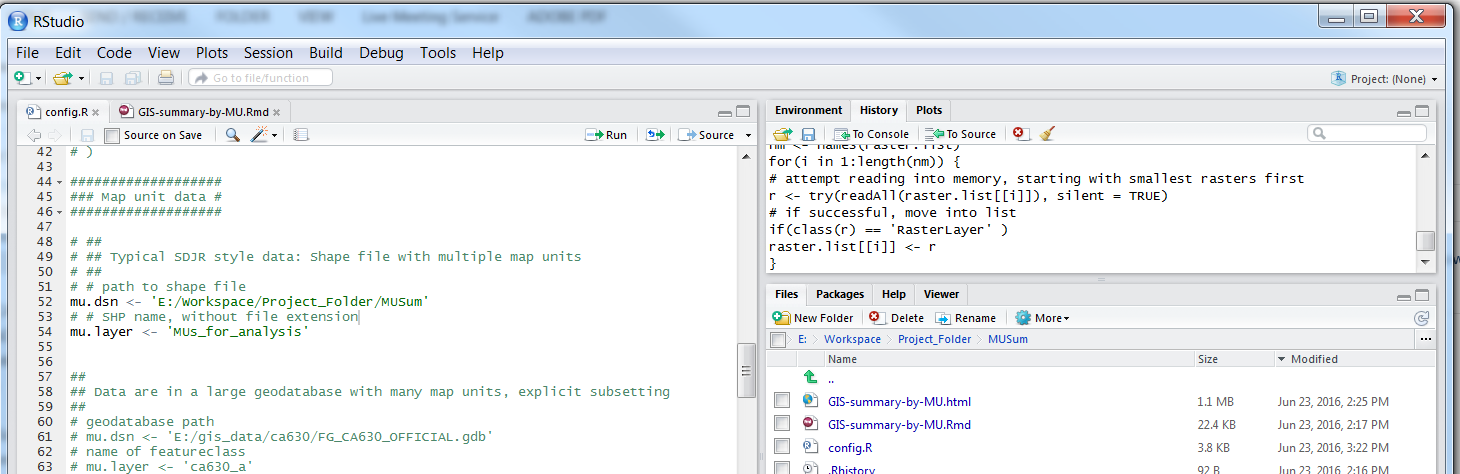
Open the file ‘config.R’ by either clicking File-Open or by clicking on the file in the File Window. It will open in the Script Window.

Scroll down to about line 15 and you will see a list of rasters. Either set up the data to match the provided folder path and file names or adjust the paths and file names to suit your data. *Pay attention to capitalization and backslash direction*. If you change any of the information, save the file by clicking File-Save on the R Studio toolbar or pressing Ctrl-S.

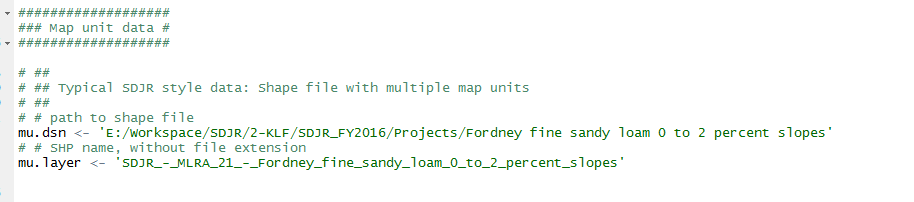


## Specify the map unit/s you are evaluating also using the config.R file

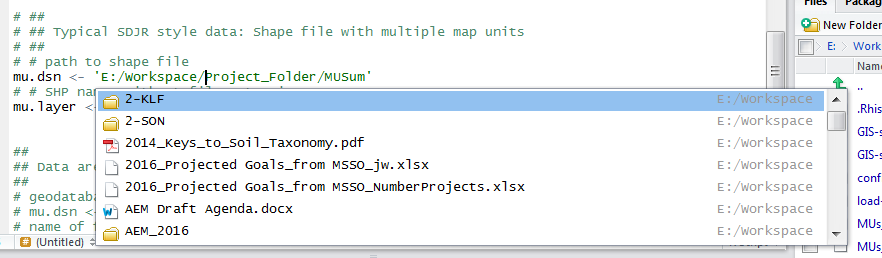
Scroll down to line 52. Specify the path and folder name where the map unit shape file is located. Scroll down to line 54 and specify the shape file name, without the file extension. *Pay attention to capitalization*. See the examples in “config.R” for the specification of data sources.



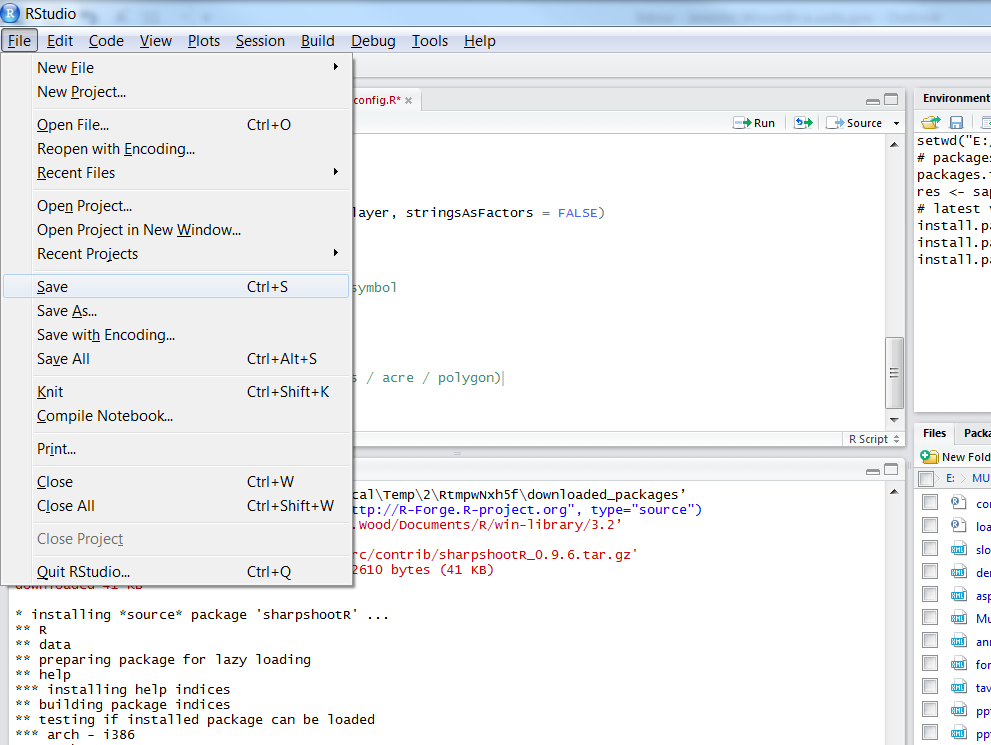
Your path name might look something like this:



*Tip:* To edit path names in the config.R file, place your cursor to the left of any “/” in the path name. Press the Tab button and a list of folders will appear where you can arrow down to choose desired folder or file.



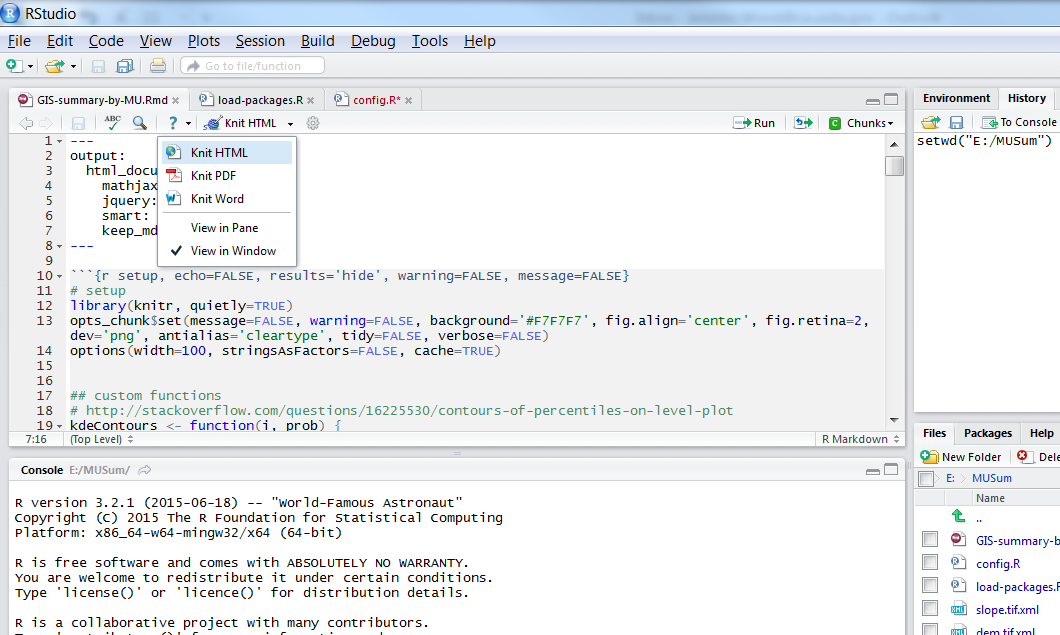
If you change any of the information, save the file by clicking File-Save on the R Studio toolbar or pressing Ctrl-S.



## Run the report by “Knitting” the file into an HTML file

Open the file titled ‘GIS-summary-by-MU.Rmd’, by either clicking File-Open or by clicking on the file in the File Window. It will open in the Script Window.

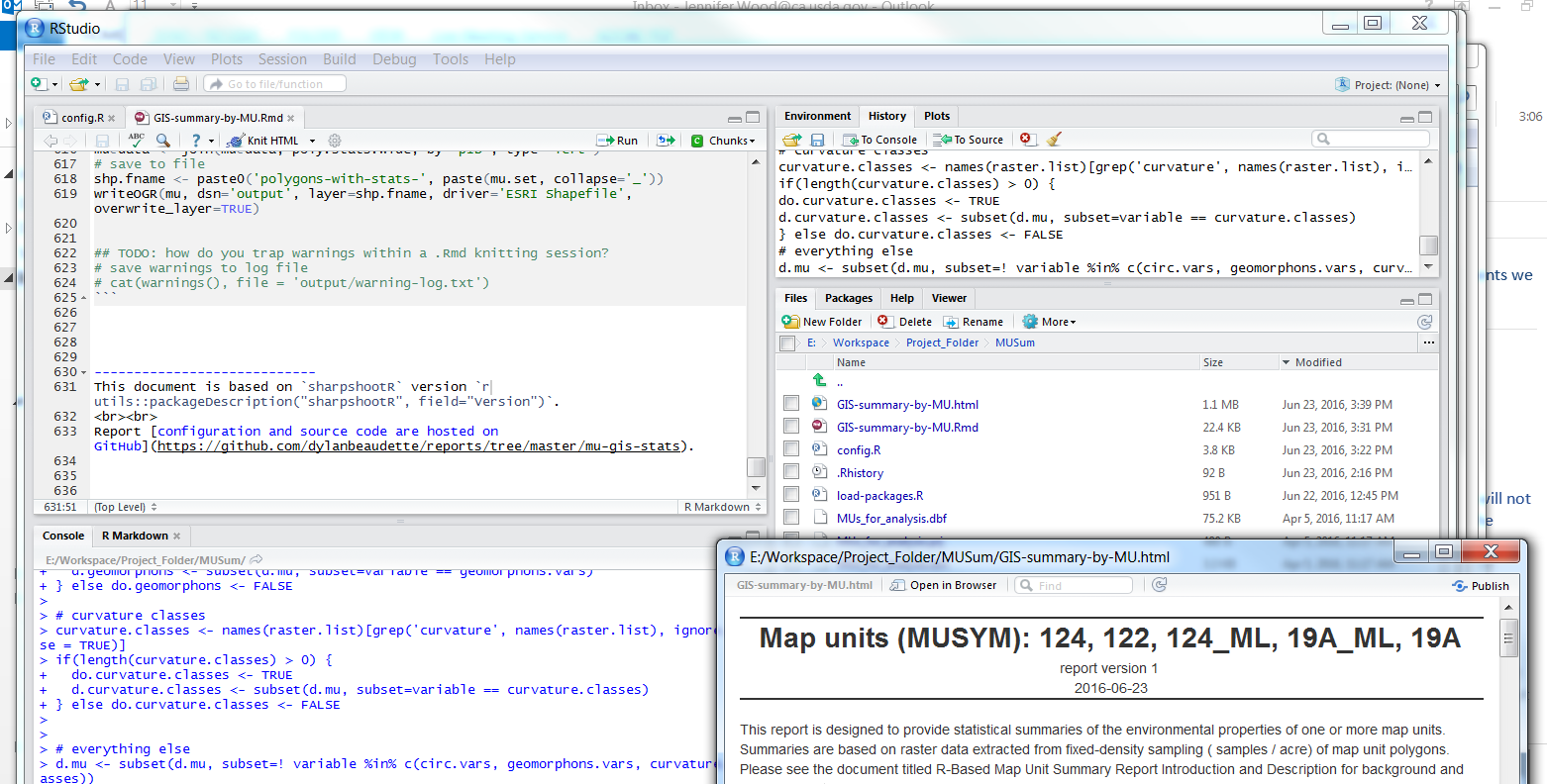
Run the report by clicking on the Knit HTML button on the Script Window toolbar. If Knit PDF or Knit Word is showing, use the drop down menu to choose Knit HTML.



## Inspect the report – see [Appendix 3](#_Appendix_3_–) for example html file.

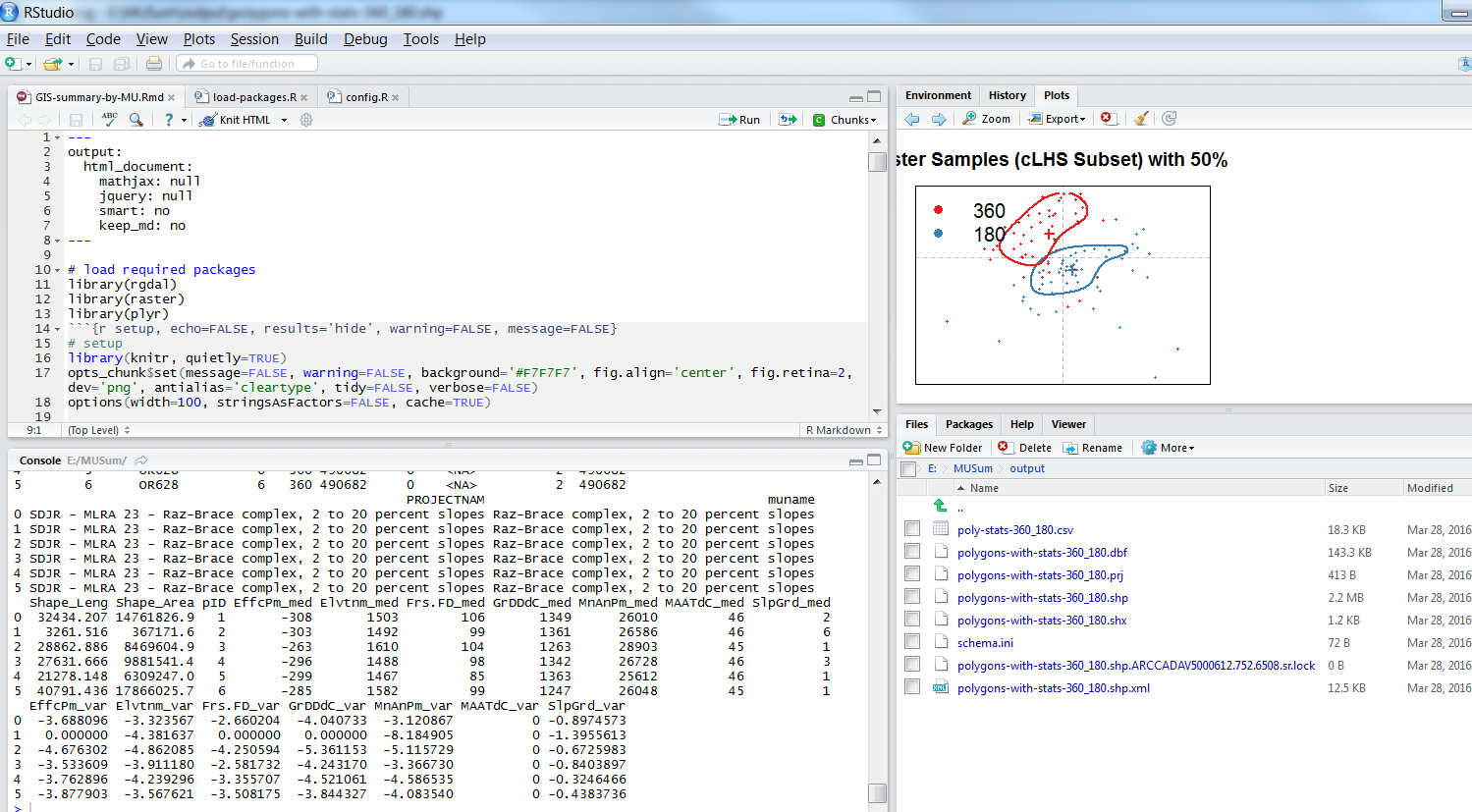
The report might take several minutes to run, depending on the size of the area you are analyzing and the size of the DEM and derivatives. You will know when the file is done when the script finishes running in the Console window, and a ‘>’ symbol appears again, ready to receive more code to execute. An html output document opens automatically. It will automatically be saved in the working directory with the same name as the report file name, but with an .html file extension.

You can close the automatically opened html file. To open it again and to see it displayed correctly there are a couple of options. If your default browser is Mozilla Firefox, it will open correctly either directly from RStudio or from Windows Explorer. Firefox also has the advantage of being able to bring an individual report graph to view in a separate window, in a larger size by right clicking and selecting the appropriate view option. If your default browser is Internet Explorer and you don’t want to change it to Firefox, you should open it from Windows Explorer and click on the “allow blocked content” warning.



## Inspect the output files

The report will also create a .csv file raster statistics for each of the environmental variables within each polygon. It will also produce a shape file with the same values in the attribute table. These files will be saved in a folder called ‘output’ in your working directory. These files can be used for other analyses and visualization.



## Using the report output data

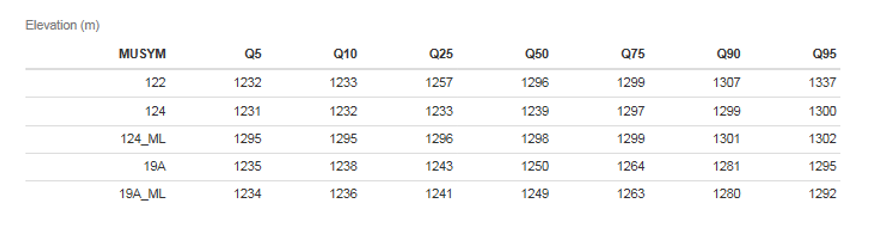
### Simplest example

The simplest use of these output data is to generate low-rv-high values of various environmental data across the spatial domain for the population of NASIS components in the data mapunits.

Remember, just as in the Lucas model output in Crystal reports, the data generated is for the map unit as a whole. You have to decide if the components need to be differentiated within the map unit based on your map unit model.

To populate NASIS for GIS data, we recommend that you use the 10th, 50th, and 90th percentile values, for low-rv-high respectively. You can obtain these from the tabular summaries in the .html file or in the .csv files.

For instance, for elevation:



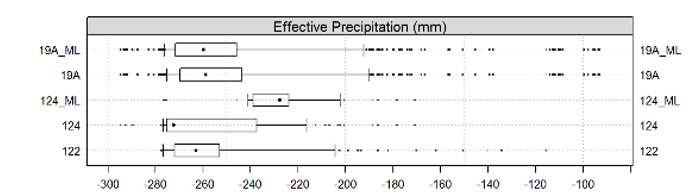
### Map unit comparison example

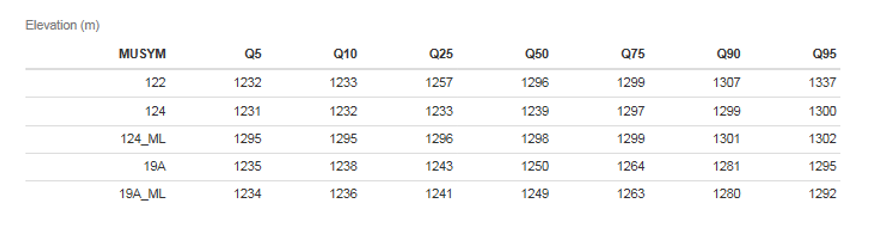
A more complex situation occurs during SDJR or initial activities when you are evaluating two different map units/sets of delineations to decide if they are the same or different based on various environmental parameters.

In this case, in addition to inspecting the tabular summaries, you can inspect both the box and whisker plots and the density plots to help decide if the distribution of the data across delineations warrants one, two, or more map units. The box and whisker and density plots show distributions for individual variables. The multivariate analysis evaluates the all variables at once in one big mash-up. This is where you have to think about the source data and understand why any differences are showing up and if they are reflecting important differences in the underlying data.

If you decide that there are two sets of delineations that you decide to combine, they you would recreate a new polygon file with those two sets merged, and populate the database based on the combined data output.

In the following situation, it was already decided that there was a small set of delineations that should be included with a larger set (map unit 19A) across a SSA join. A set of polygons that included join units had already been created, named 19A\_ML. Below, the box and whisker plot for the map unit 19A vs 19A\_ML for is shown for precipitation. The tabular data for 19\_ML would be used to populate NASIS.





After inspecting the output, you may decide that delineations should be combined in different ways. In that case, you could create a new map unit polygon file with different sets of delineations belonging to newly configured map units. You would then rerun the report on the new map unit polygon file containing the newly configured map units.

# Appendix 1 - RStudio Tutorial - Getting started

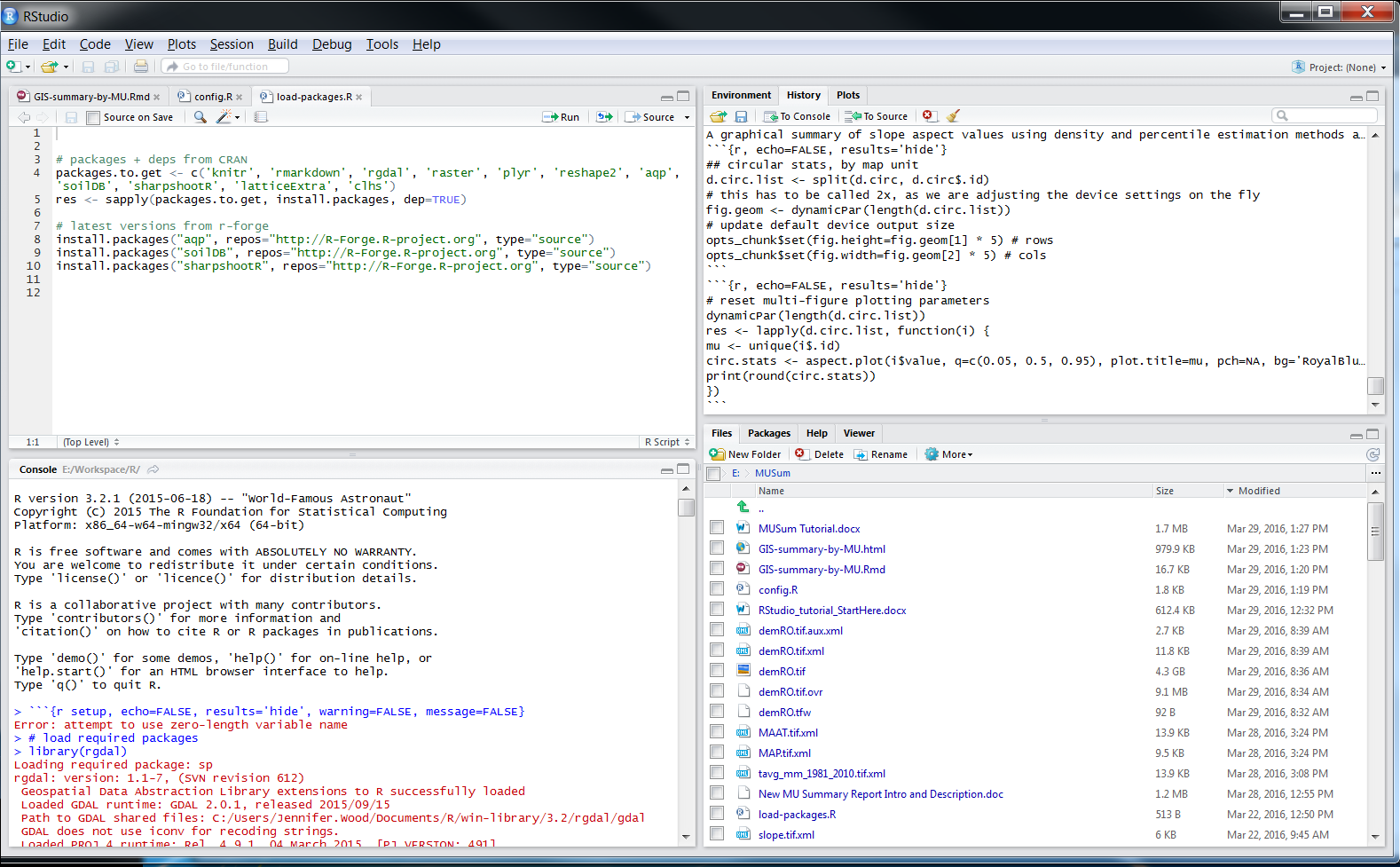
Check that R Studio is installed on your computer. Request the installation from IT, if not.

Open R Studio: Start – All Programs - Rstudio. The default display for R studio has four windows. We recommend a particular layout that can be specified from the Tools-Options menu. Check that the Options settings are correct for the General and Pane Layout tabs, see screen shots below.

Click on all the tabs that are available in each window and look at all the drop-down menus.

**Screen shots for setting up your R Studio environment**

**This is how your R Studio screen will be arranged after you set up your R Studio environment**



**Source Window**

**Shows the content of files**

**Console Window**

**Where the code is run**

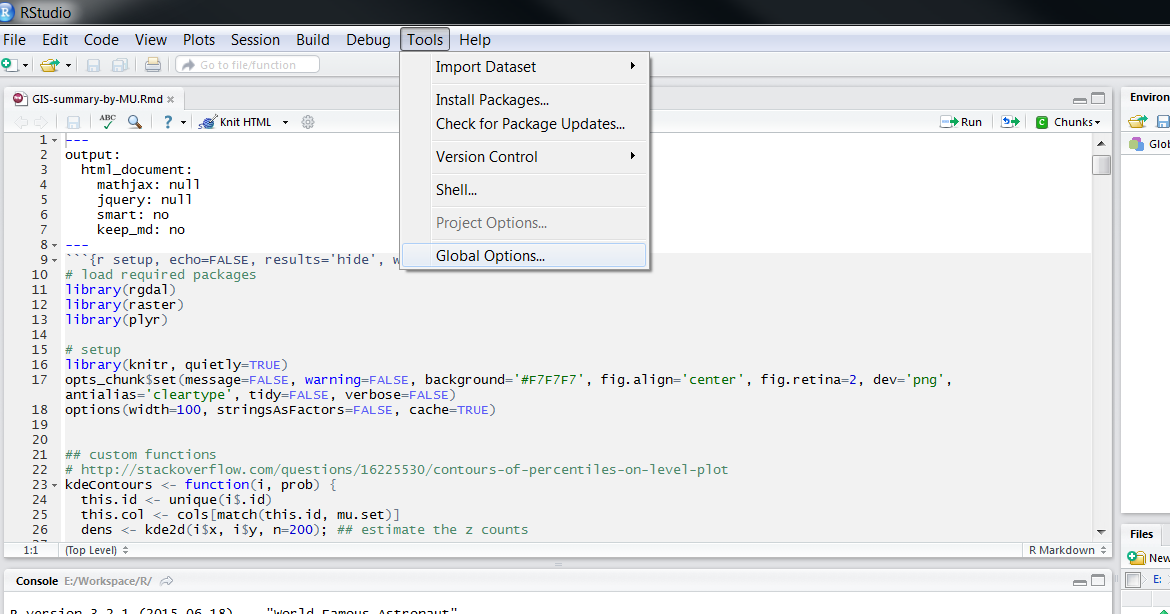
**Files, Packages, Help**

**Click on tabs to switch**

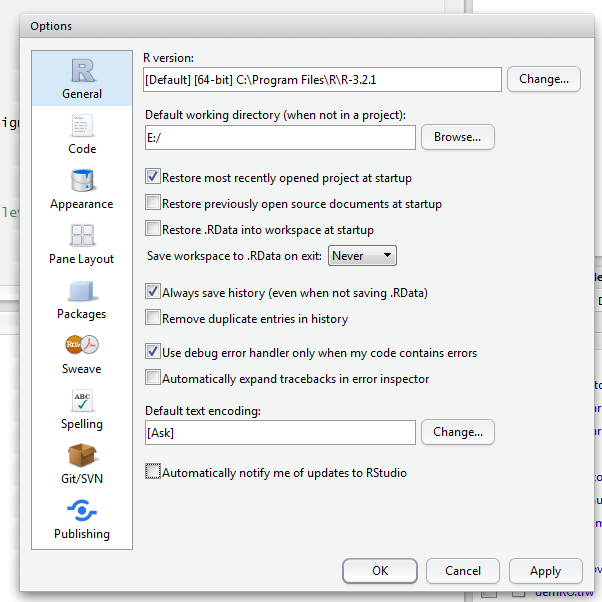
**Workspace,History,Plots**

**Click on tabs to switch**

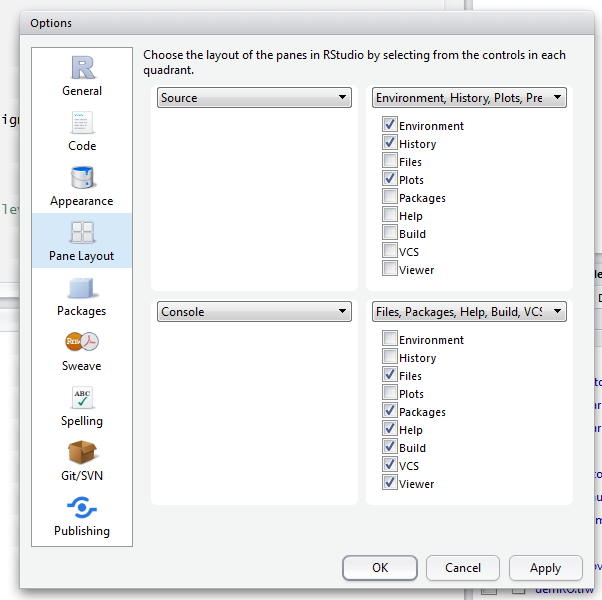
**To set up your screen like the above screenshot, click Tools- Global Options**



**In the General tab on the left, fill out the options like this:**



**In the Pane Layout tab (on the left), fill out the options like this:**



**If you have a file open, your screen will be set up as in the first screenshot above.**

**That’s It!**

# Appendix 2 – File Structure

1 Directory Structure

2 ¦--Davis Server files and folders provided to you

3 ¦ °--geodata/project data/

4 ¦ ¦--MuSum\_10m\_MSSO/ 10m DEM and derivatives

5 ¦ ¦--MUSum\_30m\_SSR2/ 30m DEM and derivatives

6 ¦ ¦--MUSum\_Curvature/ 30m curvature classes

7 ¦ ¦--MUSum\_Geomorphons/ 30m landform elements

8 ¦ ¦--MUSum\_PRISM/ 800m climate data

9 ¦ °--MUSum\_reports/ master copies of R-files

10 ¦ --geodata/elevation/derivatives/

11 ¦ --HillShades\_10m\_MLRA

12 ¦--MSSO shared data drive report runs faster from copies on local disk

you may choose to make copies of these folders on your local disk

13 ¦ ¦--geodata/project data/

14 ¦ ¦ ¦--MuSum\_10m\_MSSO/ 10m DEM and derivatives

15 ¦ ¦ ¦--MUSum\_30m\_SSR2/ 30m DEM and derivatives

16 ¦ ¦ ¦--MUSum\_Curvature/ 30 m curvature classes

17 ¦ ¦ ¦--MUSum\_Geomorphons/ 30m landform elements

18 ¦ ¦ ¦--MUSum\_PRISM/ 800m climate data

19 ¦ ¦ °--MUSum\_reports/ local master copies of R-files, do not edit

20 ¦ ¦ ¦--GIS-summary-by-MU.Rmd

21 ¦ ¦ ¦--config.R

22 ¦ ¦ °--load-packages.R run the first time only

23 ¦ °--geodata/climate/PRISM/

24 ¦ °--PRISM\_800m/ copy files from MUSum\_PRISM for use with other projects

25 ¦ --geodata/elevation/derivatives/

26 ¦ --HillShades\_10m\_MLRA

27 ¦ ¦--Project folders

28 °--[SDJR|MLRA Project Name]

29 ¦--Correlation/

30 ¦--Images/

31 ¦--Lab/

32 ¦--Maps/

33 ¦ °-- map unit polygon data (.shp, .gdb)

34 ¦--MUDs/

35 ¦--Pedons/

36 ¦--Vegetation/

37 °--MUSum/ project copy of R-files, set this as working directory

you may choose to run this on your local disk

38 ¦--output/ .csv, .shp file output automatically placed here

35 ¦--GIS-summary-by-MU.Rmd "run" this report file

36 ¦--config.R edit as needed for each project

37 ¦--load-packages.R run the first time only

38 °--map unit polygon data copy input map unit polygon data here (.shp, .gdb)

39 --\*.html html report output automatically generated here

# Appendix 3 – Example html report output file

