MAP UNIT SUMMARY REPORT- PART 2 - INSTRUCTIONS

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. R Studio is a program that you already have loaded on your computers.

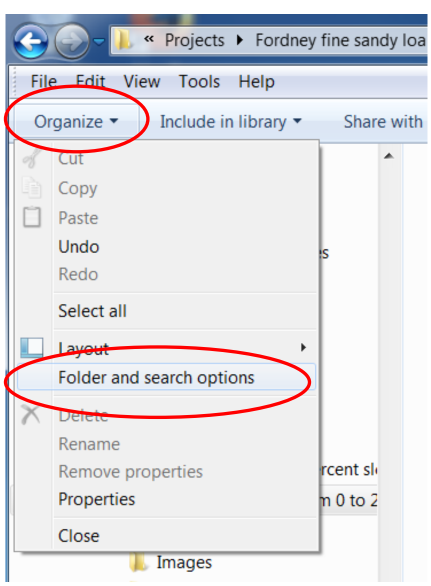
To set up your R Studio environment for the first time, read and follow the instructions in Appendix 1.

Step 3. Prepare your files and folders

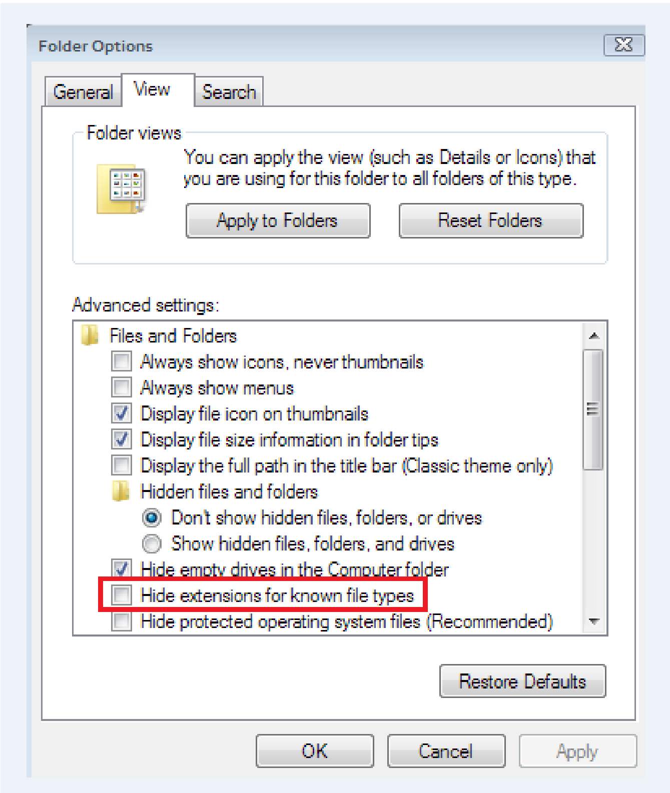
Change your folder options in Windows Explorer to show file extensions

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



Store the geodata files provided to you in the office shared drive folder location specified in Appendix 2, and make copies of these files onto your local disk in a convenient folder structure as suggested in Appendix 2.

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 for a summary of the intended file structure.

Overview

Data folders that contain files that are specifically designed for this for the MUSum report are labeled as MUSum_*, such as MUSum_10M_SSR2.

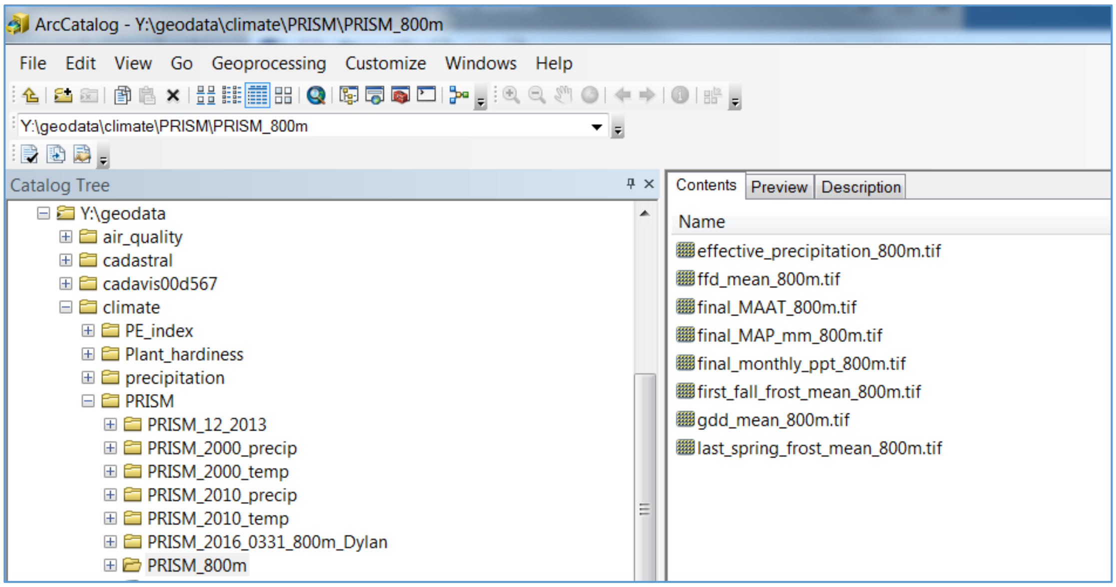
The abbreviation “MUSum” is spoken as “M U Sum”, for “map unit summary”

Master copies of the MUSum data folders will be stored in your office shared data drive, in the appropriate folder in the “geodata” folder.

All master files on your office shared drive 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. These are stored in the "project_data" folder. Other files are suitable for other uses, such as the PRISM_800m and geomorphons. These are to be located in the proper thematic folder in the "geodata" directory.

For example, for the PRISM files:



<u>Example</u>

The report works fastest when input files are stored on a local drive. Appendix 2 has a suggested file folder structure for copying the master files on your office shared drive onto your local drive. The "config.R" file (as distributed) includes references to this folder structure.

From the "config.R" file:

```
raster.list <- list(
  continuous=list(
    `Mean Annual Air Temperature (degrees C)`='E:/Workspace/geodata/PRISM_800m/final_MAAT_800m.tif',
    `Mean Annual Precipitation (mm)`='E:/Workspace/geodata/PRISM_800m/final_MAP_mm_800m.tif',
    `Effective Precipitation (mm)`='E:/Workspace/geodata/PRISM_800m/effective_precipitation_800m.tif',
    `Frost-Free Days`='E:/Workspace/geodata/PRISM_800m/ffd_mean_800m.tif',
    `Growing Degree Days (degrees C)`='E:/Workspace/geodata/PRISM_800m/gdd_mean_800m.tif',
    `Elevation (m)`='E:/Workspace/geodata/DEM_KLM_int_AEA.tif',
    `Slope Gradient (%)`='E:/Workspace/geodata/Slope_KLM_int_AEA.tif'
    #`Annual Beam Radiance (MJ/sq.m)`='E:/gis_data/ca630/beam_rad_sum_mj_30m.tif,
    #`(Estimated) MAST (degrees C)`='E:/gis_data/ca630/mast-model.tif'
    # `Compound Topographic Index`='E:/gis_data/ca630/tci30.tif',
    # `MRVBF`='E:/gis_data/ca630/mrvbf_10.tif',
    # `SAGA TWI`='E:/gis_data/ca630/saga_twi_10.tif'
  ),
  categorical=list(
    `Geomorphon Landforms`='E:/Workspace/geodata/Geomorphons/forms10_region2.tif',
    `Curvature Classes`='E:/Workspace/geodata/MU_Curvature/curvature_classes_10_class_region2.tif',
    `NLCD 2011`='E:/Workspace/geodata/land_use_land_cover/nlcd_2011_cropped.tif'
  ),
  circular=list(
    `Slope Aspect (degrees)`='E:/Workspace/geodata/Aspect_KLM_int_AEACopy.tif'
    )
)
```

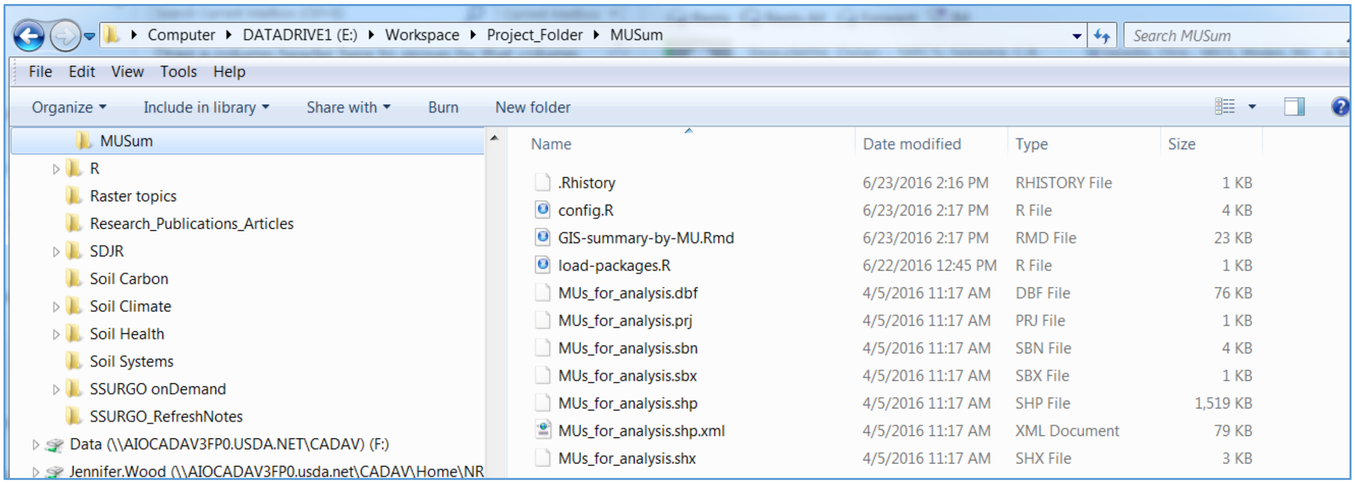
All of these files have been optimized for use by this R-based Map Unit Summary report. While it is not advisable to use the DEMs or their derivatives in other applications, we do encourage you to use other products not provided in a folder named “MUSum_*” in other applications.

Locate the soil polygon files that the report will be analyzing

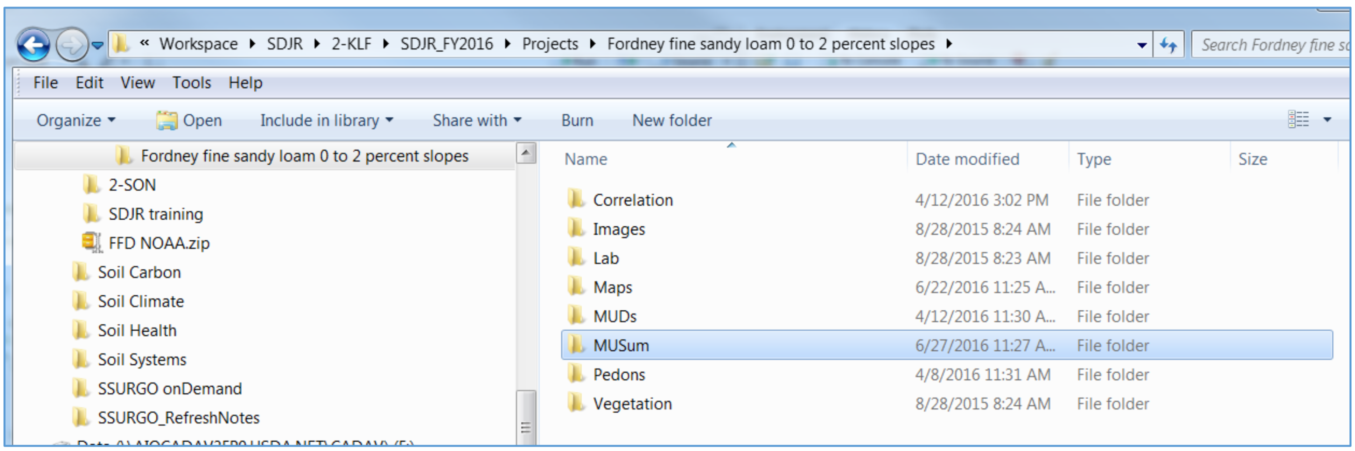
When running the report, you will be specifying a working directory that will be used to store output from the report.

You can copy your input map unit polygon data (.shp or .gdb) into the MUSum folder in your Project folder or edit the config.R file to direct it somewhere else.

In the following example, and as written into the downloaded “config.R” file, an “MUSum” folder was placed within the “Workspace/Project_Folder” on the local drive. You can put your map unit polygon file here or you can direct the script to where it is located. The report will create a folder called “MU-comparison” folder in the working directory in which it will place various files. 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.



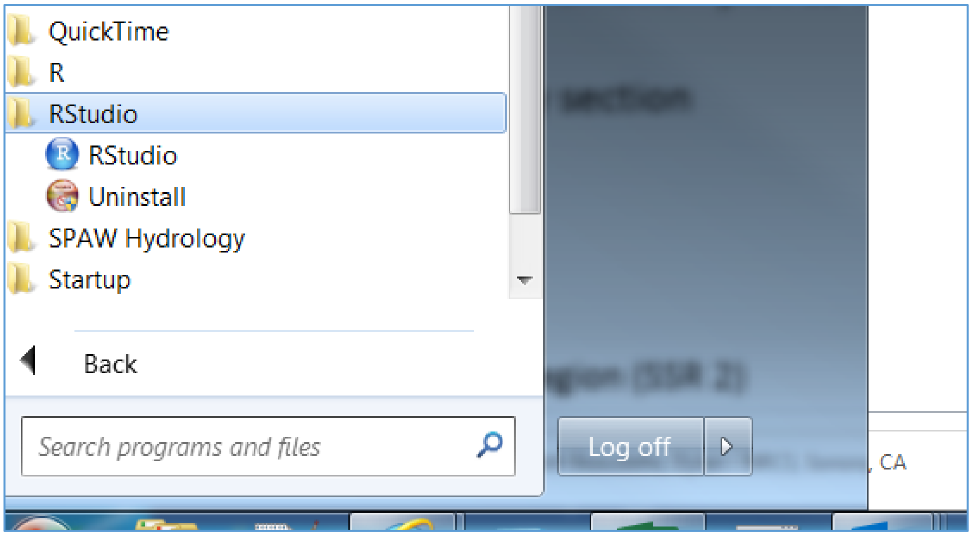
The path to your Project folder might look more like this:



See Appendix 2 for a summary of the intended file structure.

Step 4. Run the report

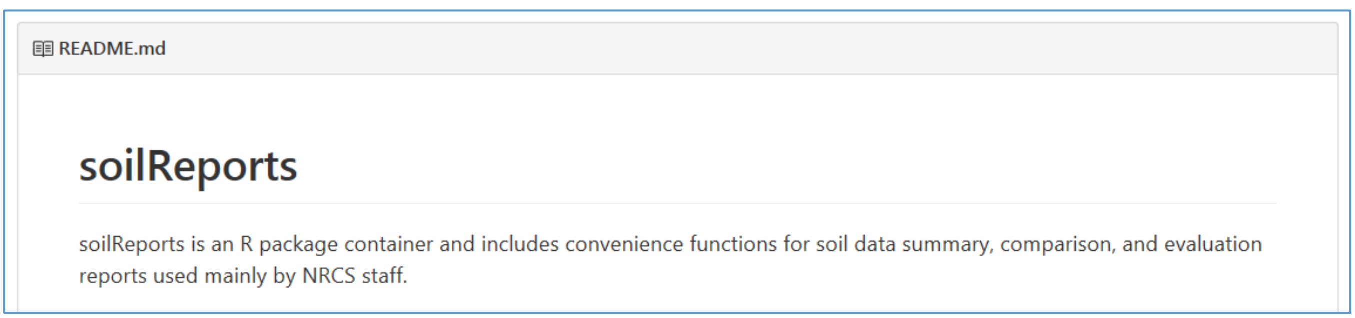
Open R Studio: Start – Programs – RStudio folder – Rstudio



All R files are obtained by running lines of script found on GitHub (a document editing and management website), at this location:

https://github.com/ncss-tech/soilReports

Scroll down to the section titled soilReports



Open R Studio if it is not open already.

For first-time report users
Scroll down to the “Pre-Installation (NRCS only)” section. Read the description and copy and paste the first line of script into your R studio console and press enter. Repeat with the second line of script and inspect the results, see example below.

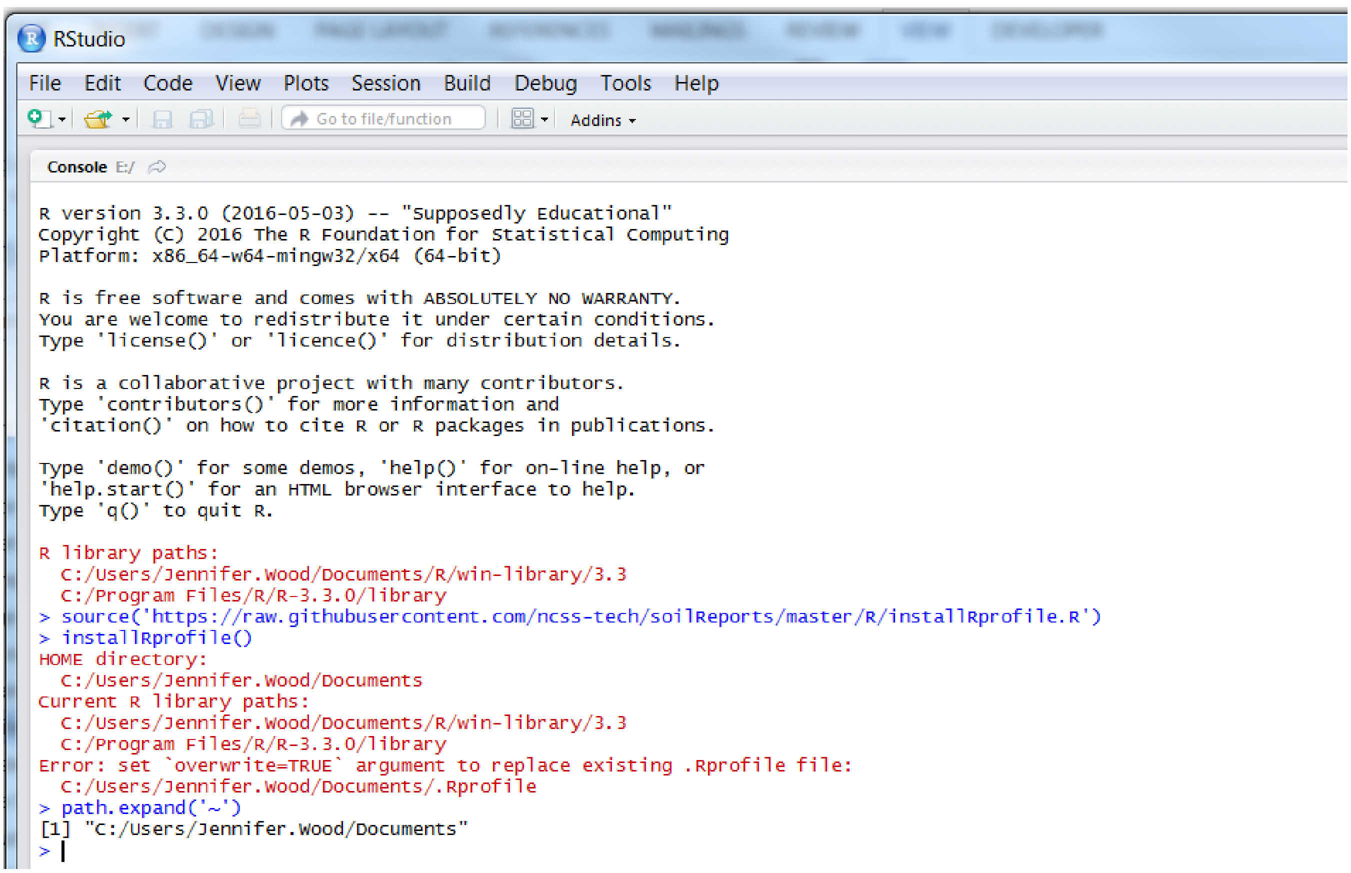
Pre-Installation (NRCS only). This is only required once.

On many of our machines, the `$HOME` directory points to a network share. This can cause all kinds of problems when installing R packages, especially if you connect to the network by VPN. The following code is a one-time solution and will cause R packages to be installed on a local disk by adding an `.Rprofile` file to your `$HOME` directory. This file will instruct R to use `C:/Users/First.Last/Documents/R/` for installing R packages. Again, you only have to do this **once**.

```
# run this in the R console
source('https://raw.githubusercontent.com/ncss-tech/soilReports/master/R/installRprofile.R')
installRprofile()
```

The following code can be used to "see" where the `$HOME` directory is. The result should look like "C:/Users/First.Last/Documents"

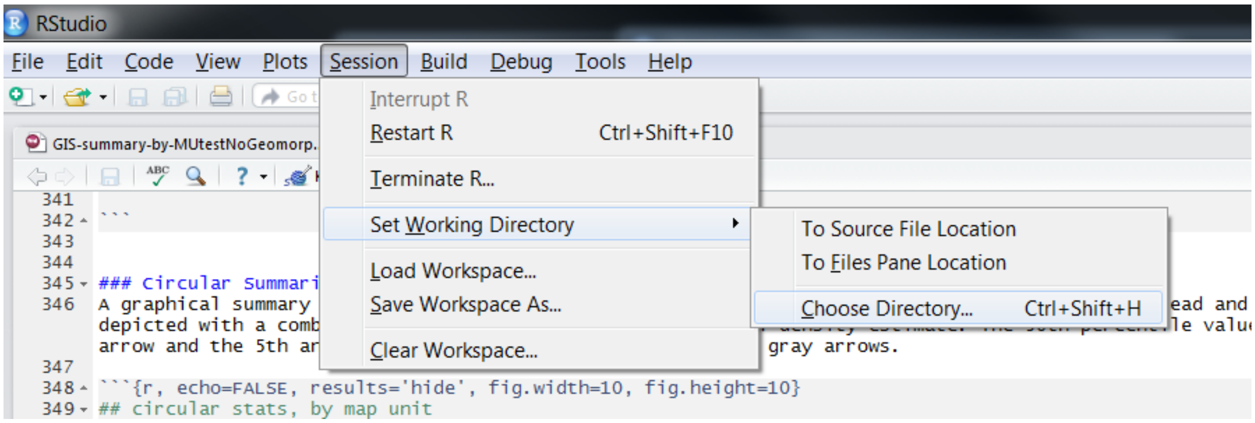
```
# run this in the R console
path.expand('~')
```

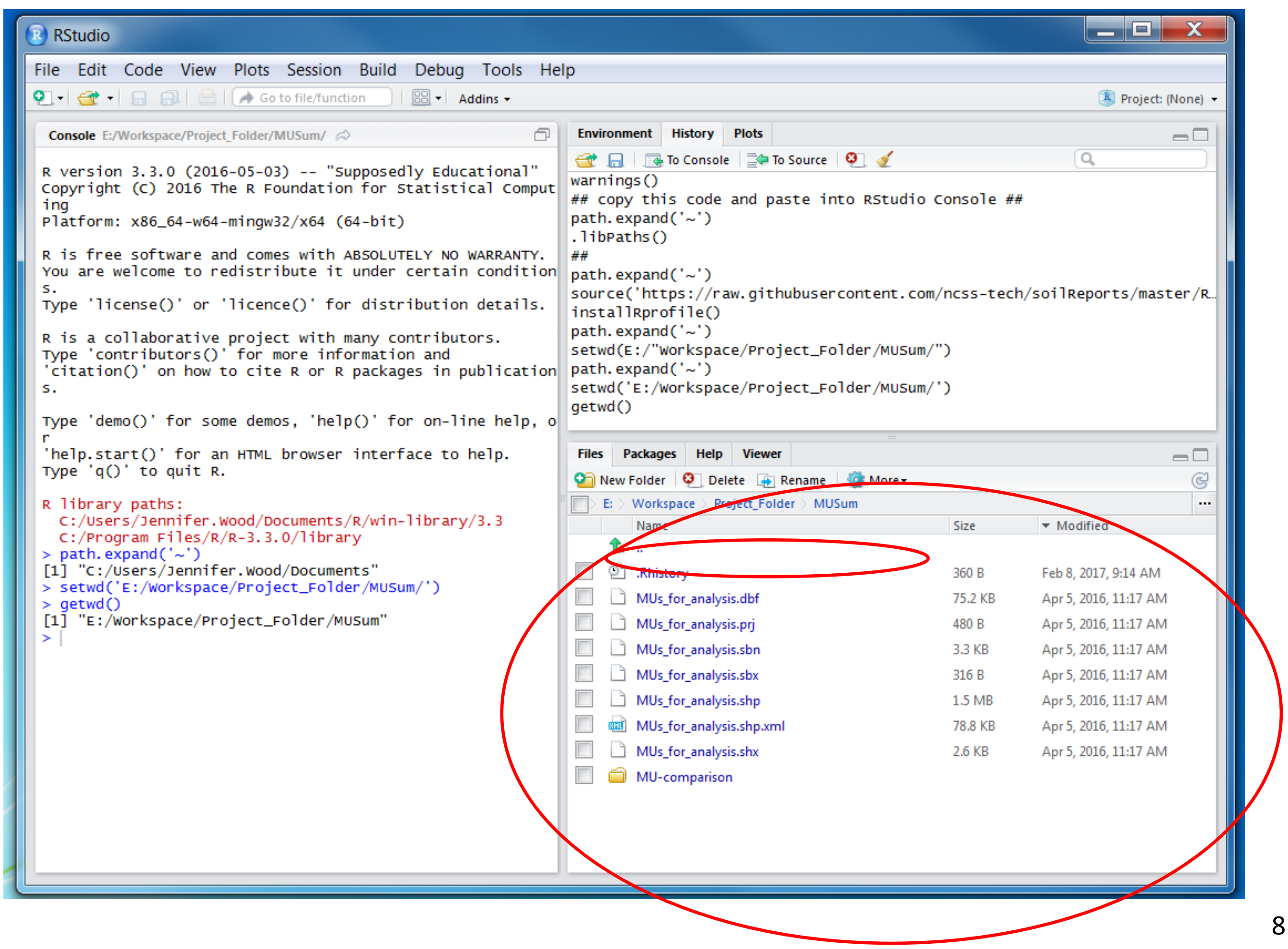


Set the working directory

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. See [Appendix 2](#) for an example.





Tips: You can also set the working directory by using the More button in the File window in R. Alternatively, you can use the command setwd () directly in the console as shown in the example above, and getwd() to see the working directory. Also, If you open an R file from Windows Explorer, R Studio will automatically set the working directory to the folder where that file is located.

For first-time use and when a new version of soiReports is released, install and run the soilReport from GitHub

Go to the soilReports page on GitHub - https://github.com/ncss-tech/soilReports

Scroll down to the section titled “Installation of the soilReports package….”. Copy the chunk of script in the grey block and paste it into your R console window, and press enter.

Installation of the soilReports package. Only required for first-time use of soilReports and when a new version of soilReports is released.

The current version of `soilReports` is installed with the following code:

```
# need devtools to install packages from GitHub
install.packages('devtools', dep=TRUE)

# get the latest version of the 'soilReports' package
devtools::install_github("ncss-tech/soilReports", dependencies=FALSE, upgrade_dependencies=FALSE)
```

```
> install.packages('devtools', dep=TRUE)
Installing package into 'C:/Users/Jennifer.Wood/Documents/R/win-library/3.3'
(as 'lib' is unspecified)
Warning in install.packages :
  dependency 'BiocInstaller' is not available
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/devtools_1.12.0.zip'
Content type 'application/zip' length 432180 bytes (422 KB)
downloaded 422 KB

package 'devtools' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Jennifer.Wood\AppData\Local\Temp\1\Rtmpw9WcDE\downloaded_packages
>
> # get the latest version of the 'soilReports' package
> devtools::install_github("ncss-tech/soilReports", dependencies=FALSE, upgrade_dependencies=FALSE)
Downloading GitHub repo ncss-tech/soilReports@master
from URL https://api.github.com/repos/ncss-tech/soilReports/zipball/master
Installing soilReports
"C:/PROGRA~1/R/R-33~1.0/bin/x64/R" --no-site-file --no-environ --no-save --no-restore --quiet CMD INSTALL  \
  "C:/Users/Jennifer.Wood/AppData/Local/Temp/1/Rtmpw9WcDE/devtoolsfdc3fe167f7/ncss-tech-soilReports-00c1bea"  \
  --library="C:/Users/Jennifer.Wood/Documents/R/win-library/3.3" --install-tests

* installing *source* package 'soilReports' ...
** R
** inst
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
*** arch - i386
*** arch - x64
* DONE (soilReports)
>
```

For first-time use, when a new version of soilReports is released, or when you need original report versions, load the soilReports library and download the required files.

In the soilReports GitHub page, scroll down to the section titled "Loading the soilReports library….". Copy the chunk of script in the grey block and paste it into your R console window, and press enter.

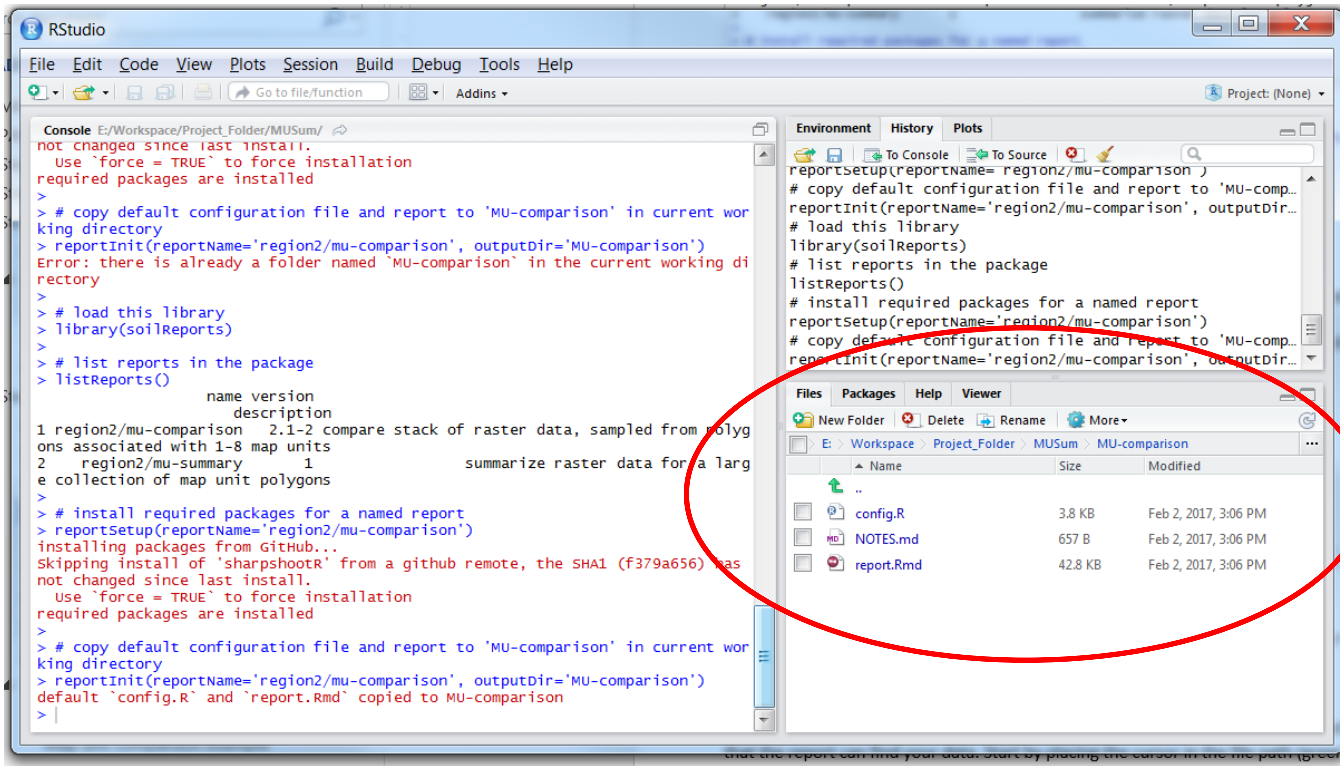
Loading the soilReports library and downloading the required files. Only required for first-time use, when a new version of soilReports is released, or when you need an original version of the config.R and report.Rmd files.

The `soilReports` package contains reports and associated configuration files. The following steps perform all required setup for the **region2/mu-comparison** report, then copies the configuration (config.R) and report (report.Rmd) files to a folder that it creates named 'MU-comparison' in the working directory. Edit the `config.R` file (or replace it with an existing config.R in the working directory) so that it points to the correct raster layers and map unit polygons. "Knit" the report file by opening `report.Rmd` and clicking on the "Knit HTML" button. The package will put a 'report.html' file in the MU-comparison folder and will create a folder named 'output' for report-generated shapefiles.

```
# load this library
library(soilReports)

# list reports in the package
listReports()

# install required packages for a named report
reportSetup(reportName='region2/mu-comparison')

# copy default configuration file and report to 'MU-comparison' in current working directory
reportInit(reportName='region2/mu-comparison', outputDir='MU-comparison')
```

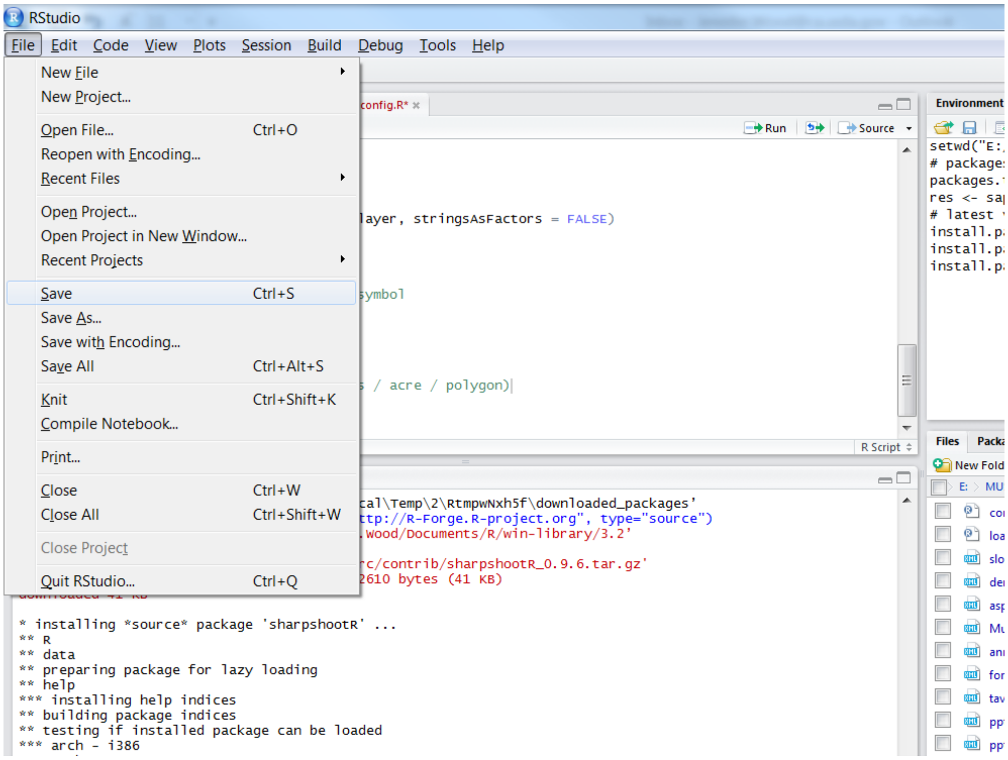
```
>
> # load this library
> library(soilReports)
>
> # list reports in the package
> listReports()
                   name version                                                            description
1 region2/mu-comparison   2.1-2 compare stack of raster data, sampled from polygons associated with 1-8 map units
2    region2/mu-summary       1                  summarize raster data for a large collection of map unit polygons
>
> # install required packages for a named report
> reportSetup(reportName='region2/mu-comparison')
installing packages from GitHub...
Skipping install of 'sharpshootR' from a github remote, the SHA1 (f379a656) has not changed since last install.
  Use `force = TRUE` to force installation
required packages are installed
>
> # copy default configuration file and report to 'MU-comparison' in current working directory
> reportInit(reportName='region2/mu-comparison', outputDir='MU-comparison')
default `config.R` and `report.Rmd` copied to MU-comparison
>
```

You will now see a folder named mu-comparison in your file window, and if you open that, you will see the ‘config.R’ and the ‘report.Rmd’ files placed there by the soilReport script.



Every time you run the report, set the input data (rasters and polygon files) using the config.R file

Tip: Before changing information in the config.R file in your Project folder, you can make a copy of the original in a convienient folder for use in subsequent analyses. However, you can always go through the previous step to get a fresh copy. To save edits in the Project config.R file, you can click File-Save on the R Studio toolbar or press Ctrl-S. Ctrl-Z undoes operations completed since the last Save.

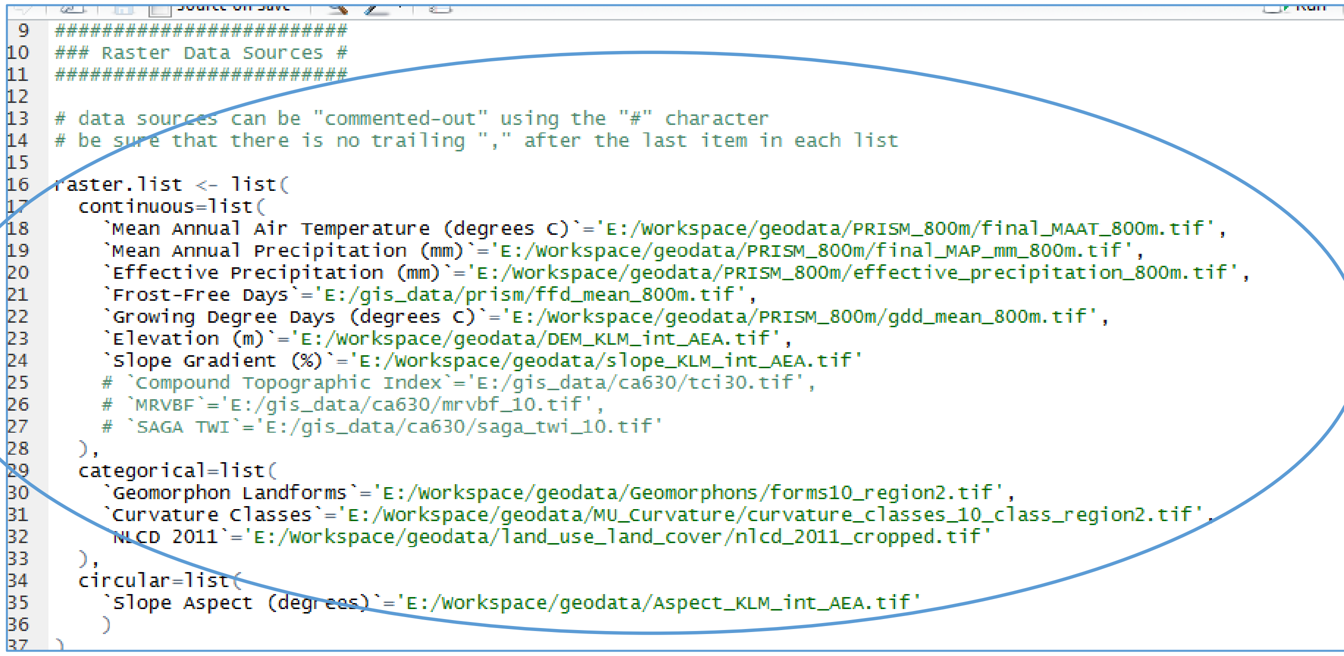


Raster files

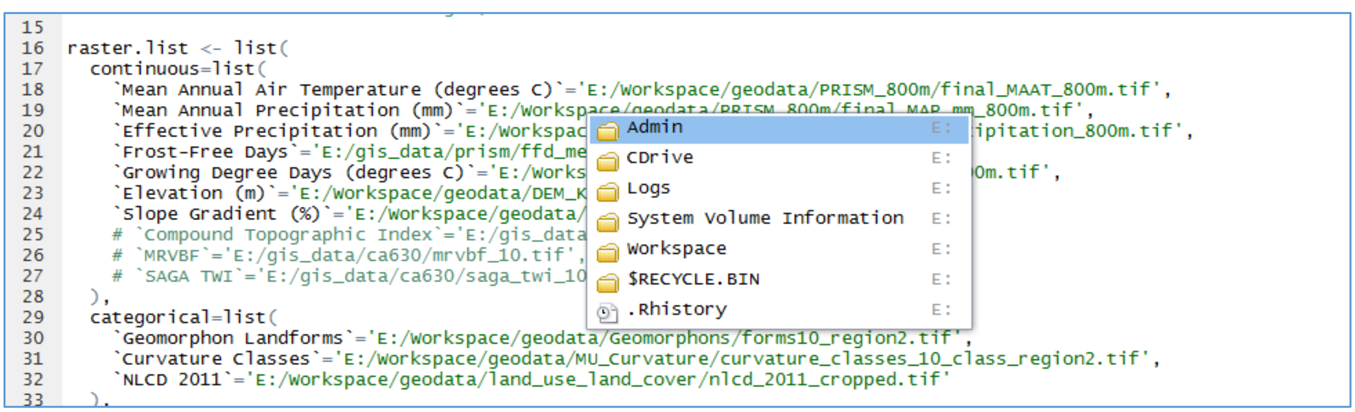
The raster files are the environmental data you are interested in summarizing for your map units.

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 Source Window.

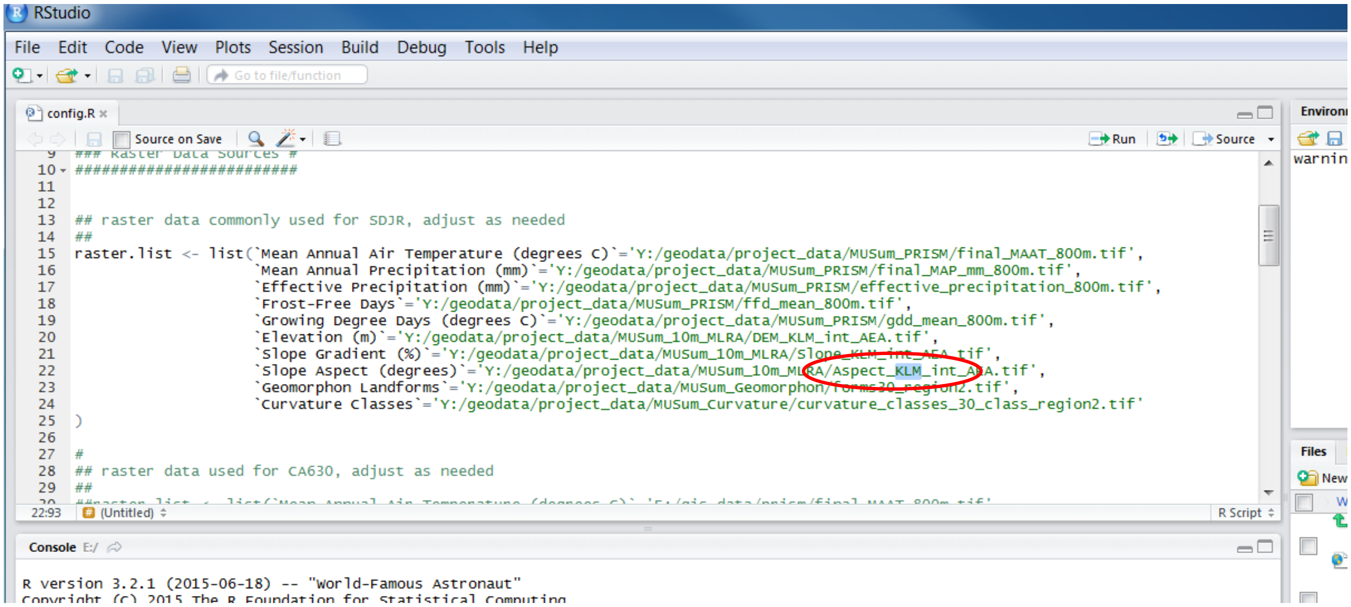
Scroll down to about line 16 and you will see a list of rasters. You will need to directly edit these pathways so that the report can find your data. Start by placing the cursor in the file path (green text) where you need to start changing the directory and/or the folder.



Tip: To edit path names in the config.R file, here is a handy tip: First make sure that the drive letter correctly specifies where your folder/file of interest resides. Place your cursor to the right 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. Repeat that until you find the folder and/or file you are looking for.



Also edit the office abbreviation in the DEM and derivative files to match yours (KEA for Kealakekua, CHI for Chico, etc).



Pay attention to capitalization and backslash direction.

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

Polygon files

Scroll down to line 52 or 65, depending on the kind of polygon file you are using.

Scroll to line 52 if you are specifying multiple map units from a geodatabase or shapefile. Edit the path and folder name that contains the file and specify the mapunits your are analyzing. The file is located by directly editing the pathway (green text).

Note: If the geodatabase containes more than 2 or 3 survey areas, it is advisable to extract the polygons as a shape file and use the procedure starting at line 63.

If you are skipping to the procedure that uses a shapefile of just the mapunits you are evaluating, ensure that these lines of code have a hashtag (#) in front of them to indicate to the report not to run those commands.

```
###################
### Map unit data #
###################

##
## Data are in a geodatabase with many map units, explicit subsetting
## Note: consider sub-setting to SHP if the geodatabase contains more than 2-3 soil survey areas
##

# geodatabase path
#mu.dsn <- 'E:/gis_data/ca630/FG_CA630_OFFICIAL.gdb'
# name of featureclass
#mu.layer <- 'ca630_a'
# map unit symbols / keys to extract
#mu.set <- c('7011', '5012', '7089')
```

Scroll down to line 63 if you are evaluating polygon data in a shapfile, and edit the pathway for the shape file, either by using the Tip above to find the folder or by typing in the pathway direcly. Ensure there is no forward slash after the folder name that contains your shapefile.

Scroll down to line 65 and either use the Tip above to find the file or type in the name of the shape file, *without the file extension*. Pay attention to capitalization.

```
56
57
58  # ##
59  # ## Typical SDJR style data: SHP with multiple map units
60  # ##
61
62  # # path to SHP
63  mu.dsn <- 'E:/Workspace/Project_Folder/MUSum'
64  # # SHP name, without file extension
65  mu.layer <- 'MUs_for_analysis'
66
67
```

Your path and file names might look something like this:

```
57
58  # ##
59  # ## Typical SDJR style data: SHP with multiple map units
60  # ##
61
62  # # path to SHP
63  mu.dsn <- 'E:/Workspace/SDJR/2-KLF/SDJR_FY2016/Projects/Fordney fine sandy loam 0 to 2 percent slopes
    /MUSum'
64  # # SHP name, without file extension
65  mu.layer <- 'Fordney_fine_sandy_loam_0_to_2_percent_slopes'
66
67
68
```

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

Other config.R settings

<u>Map unit symbol column</u>

Depending on the source of your polygon data, you may have to change the column name used for your map unit symbols.

```
############################################
### column with map unit ID / key / symbol #
############################################

# could be 'MUKEY', 'MUSYM', or any valid column name
mu.col <- 'MUSYM'

```

<u>Raster sampling density</u>

The number of points sampled by the report, in the rasters within the map units, can be adjusted.

A sampling density **between 1 and 2 points per acre** should be *sufficient* for most of the commonly used 10m to 30m data sources, and *excessive* for PRISM data (800m). The sampling density "sweet spot" for most combinations of map units and raster data sources is on the order of **1 point per acre**.

There are two cases that require further consideration of an "optimal" sampling density: 1) map unit polygons of limited extent, and 2) map units composed of very small delineations (less than 5 acres). In these cases it would be wise to increase the sampling density to a value between 2 and 5 points per acre.

[Test out the "stability" of different sampling density values using your own map units and raster data sources.]()

```

#########################################################
### polygon sampling density (samples / acre / polygon) #
#########################################################

# consider using a sampling density between 1-2 points / ac.
# increase if there are un-sampled polygons
# delineations smaller than 5 ac. may require up to 5 points / ac.
# values > 6-7 points / ac. will only slow things down
pts.per.acre <- 1


```

<u>Quantiles to display</u>

You can choose to reduce or change the quantiles in the report. We are recommending that .1, .5, and .9 are chosen for consistent population of NASIS.

```

###########################
### quantiles of interest #
###########################

# the most important quantiles (percentiles / 100) are: 0.1, 0.5 (median), and 0.9
# optionally reduce the number of quantiles for narrower tables
p.quantiles <- c(0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95)
```

<u>Adding confidence interval feature to box and whisker plots</u>

If you have a small polygon file, this feature can add information to the box and whisker plots, by changing FALSE to TRUE in this part of the config.R file.

```
########################################################
### Add estimate of confidence to box and whisker plots ###
########################################################

#enabling this feature will double the run time
#enabling this feature will add "notches" to box and whisker plots
# that are close approximations to a confidence interval around the median
# adjusted for spatial autocorrelation
correct.sample.size <- FALSE

```

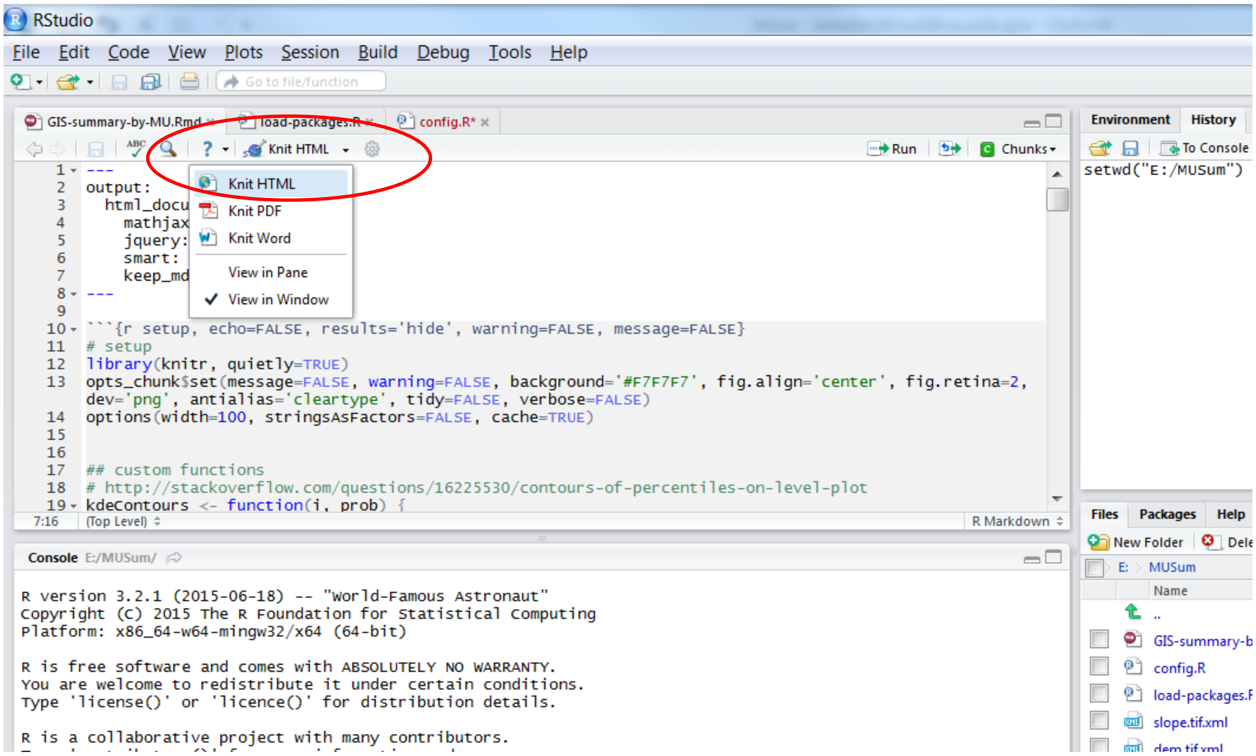
If you change any of the information, save the file by clicking File-Save on the R Studio toolbar or pressing Ctrl-S.

Tip: Now that you have set your config.R file for your typical working environment, you can copy this config.R file to other Project folders and you might only have to change the name of the shape file folder and file name. You can also copy the report.Rmd to other Project folders.

Run the report by "Knitting" the file into an HTML file

Open the file titled 'Report.Rmd', by either clicking File-Open or by clicking on the file in the File Window. It will open in the Source Window.

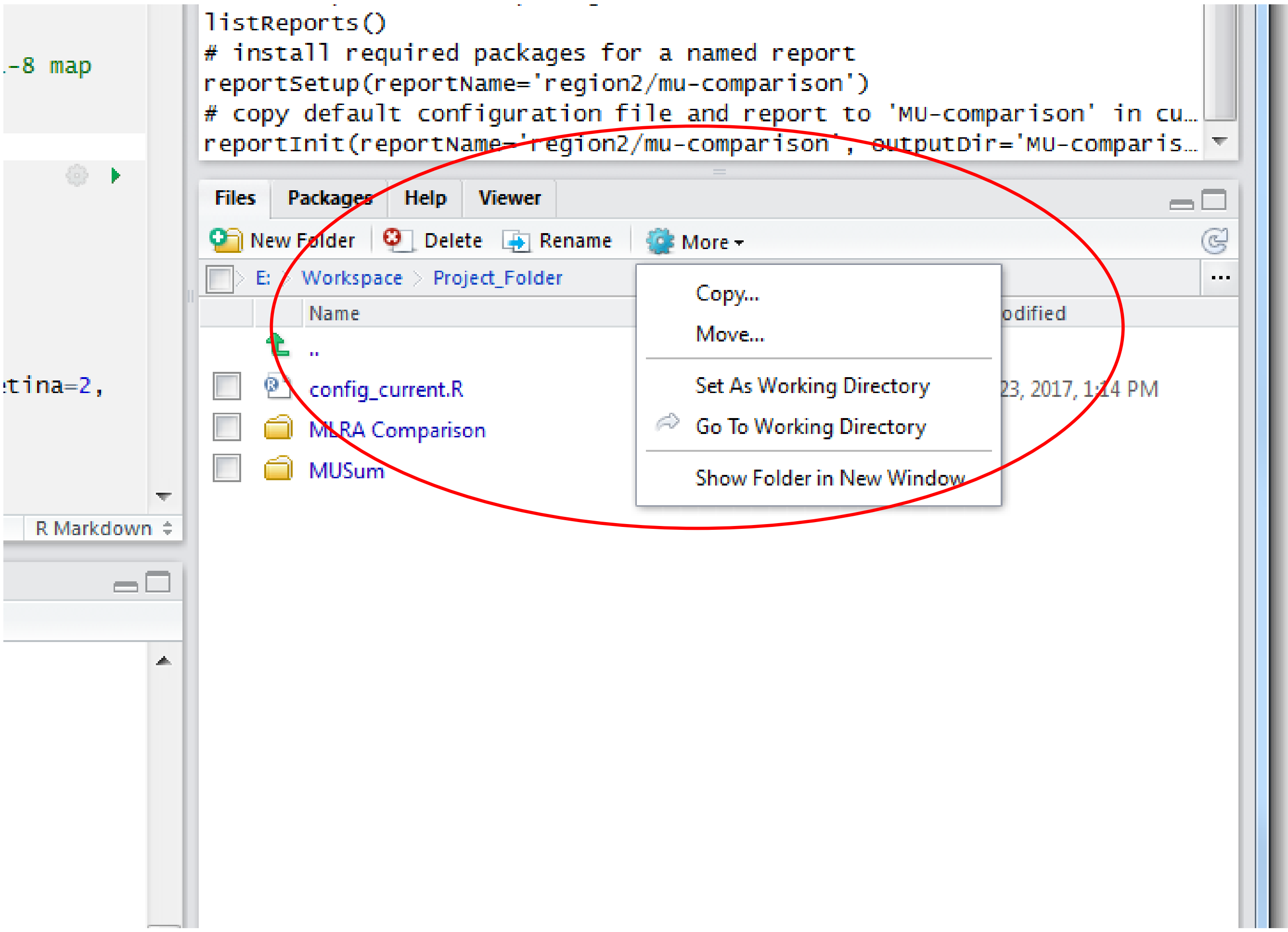
Run the report by clicking on the Knit HTML button on the Script Window toolbar. If Knit PDF or Knit Word is showing, use the drop down menu to choose Knit HTML.



Running the report again and output file management

Now you have successfully run the report, and the config.R file directed the report to the correct locations. This can become your new master config.R file (until you are notified that a new version of the report with a new config.R file is available). You could copy this config.R file directly into another Project folder for the next Project you will be working. Or perhaps you could move it to a folder higher up and name it something connotative like 'config_current.R'. You will come up with your own workflow to keep track of which file has been reconfigured to match your locally stored data.

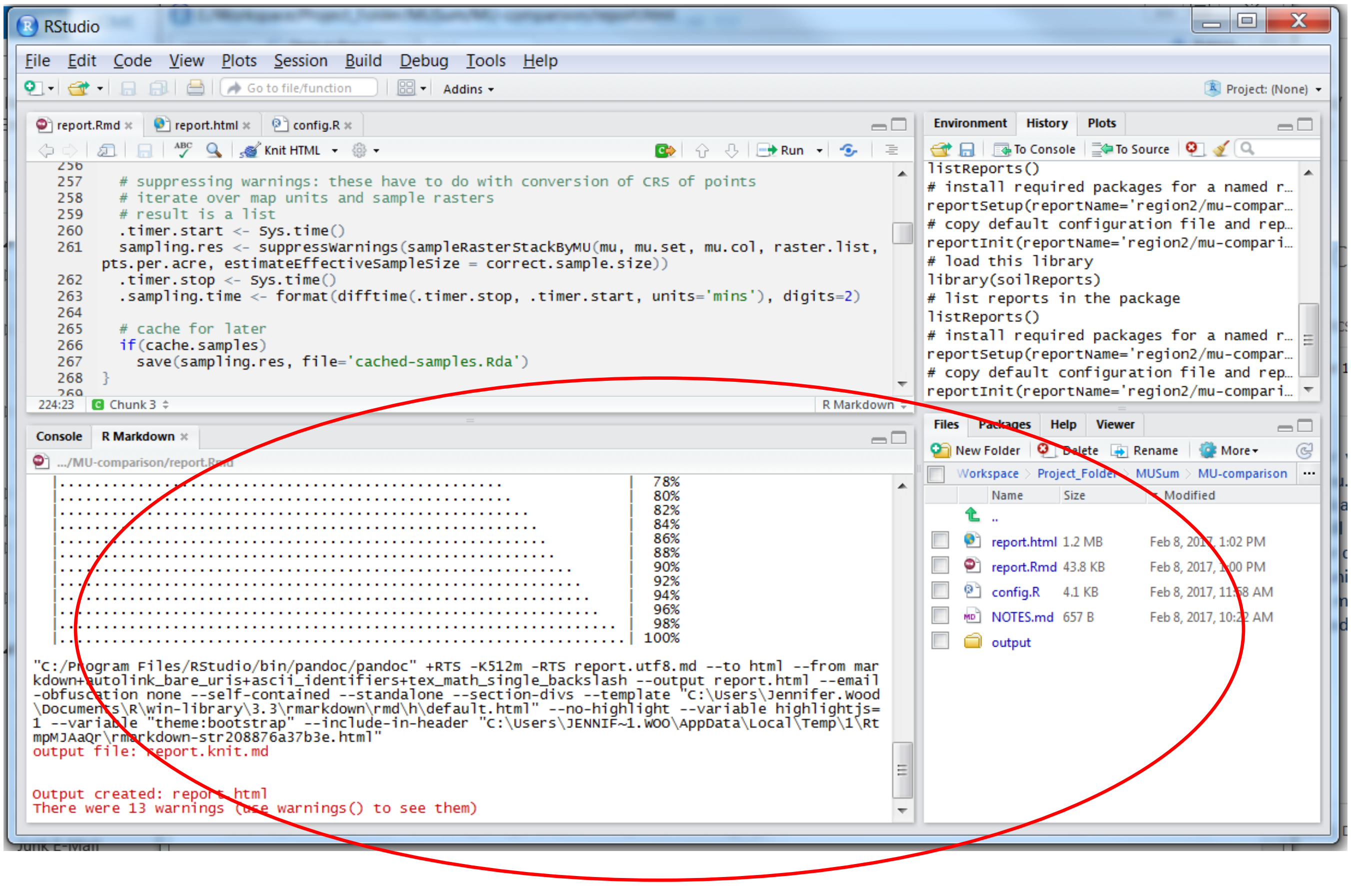
You can use the File window menu bar to Delete, Rename, Copy, and Move files in R Studio



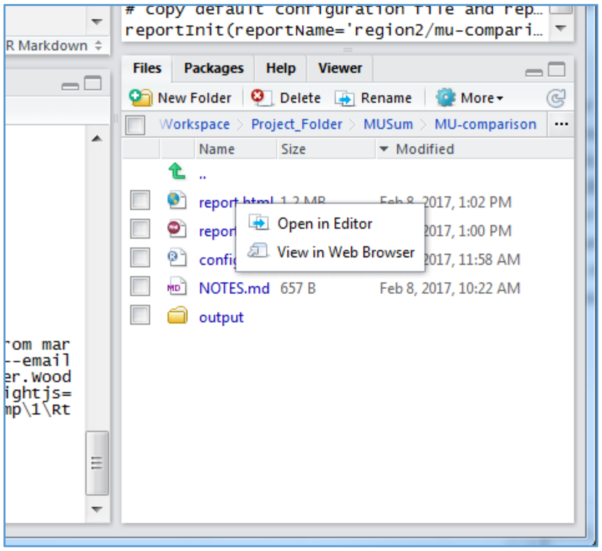
Inspect the report – see [Appendix 3](#) for example html file.

The report might take several minutes or more to run, depending on the size of the area you are analyzing and the size of the DEM and derivatives. A new tab in the lower left Console window, called R Markdown, will appear. The report will start running and produce a list of percentages that represent how far along the report has run. Depending on the size of the project, it might sit there a while and look like it is not doing anything. When it is done, the screen will look similar to the following. You can ignore the red warnings text.

An html output document is generated and will automatically be saved in the MU-comparison folder, with an .html file extension.



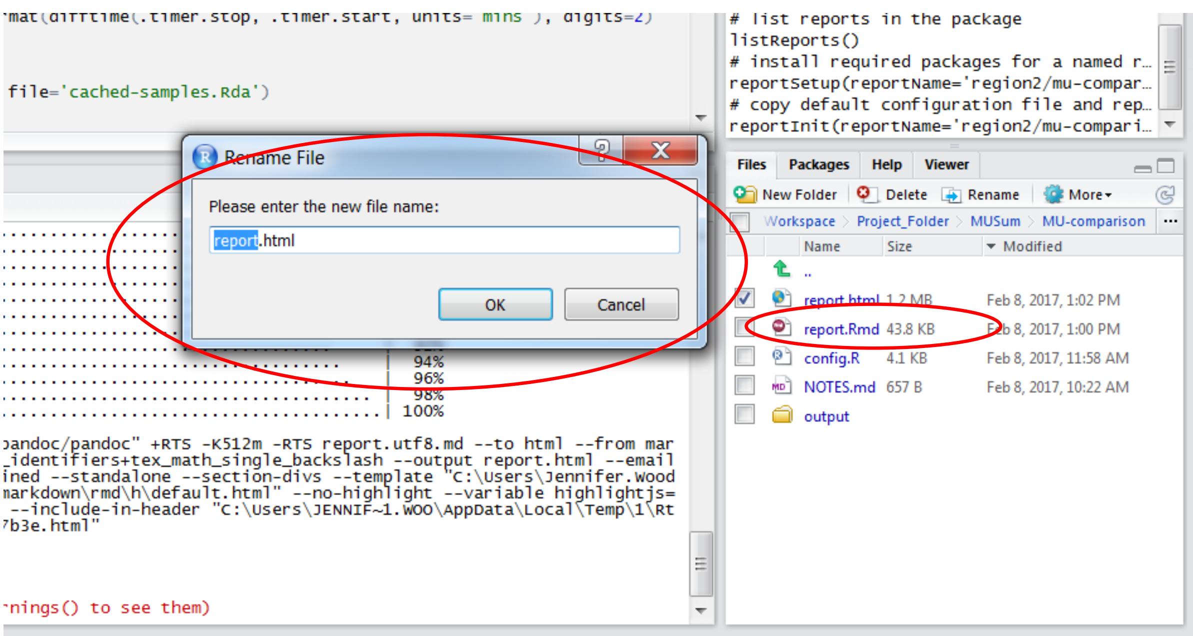
To inspect the report, click on the report.html document in the MU-comparison folder, and choose View in Web Browser. If your default browser is Internet Explorer, click on the "allow blocked content" warning. Mozilla Firefox has the advantage of being able to bring an individual report graph to view in a separate window, in a larger size by right clicking and selecting the appropriate view option.



Once you have taken a look at the html document (see discussion below and Appendix 3 for an example), you can close it. You might want to save that output for future reference because if you run the Report.Rmd file again in that same working directory, it will overwrite the file in the MU-comparison folder.

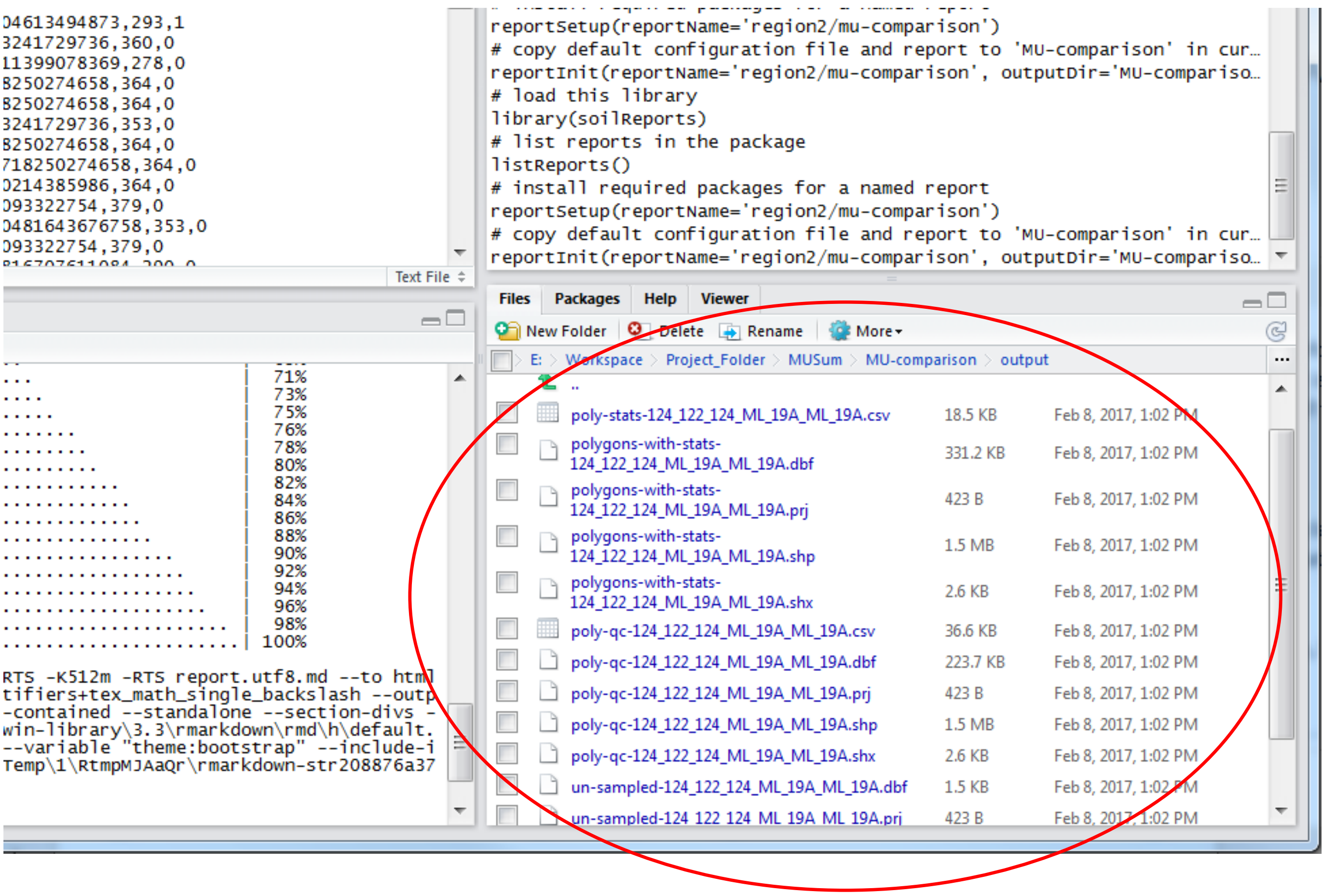
To rename, first close the html file. Then you can check the box next to the html file, click on the Rename button in the File window menu, and change the name in the dialog box that pops up.

You can also rename and move these files around in Windows Explorer.



Inspect the data files in the output folder

The report will also create a .csv file of 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 in ArcMap, R, or other applications. There is an explanation of what these files contain in the html report document.



Using the report output data

<u>Simplest example</u>

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:

Elevation (m)

MUSYM	Q5	Q10	Q25	Q50	Q75	Q90	Q95
122	1232	1233	1257	1296	1299	1307	1337
124	1231	1232	1233	1239	1297	1299	1300
124_ML	1295	1295	1296	1298	1299	1301	1302
19A	1235	1238	1243	1250	1264	1281	1295
19A_ML	1234	1236	1241	1249	1263	1280	1292

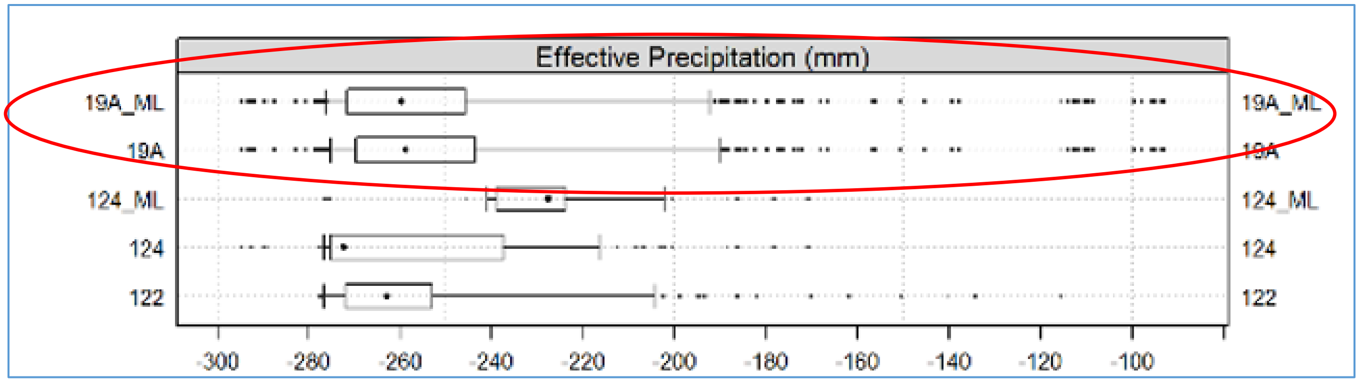
<u>Map unit comparison example</u>

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.

Note: If you decide that there are two sets of delineations that you need to combine, then you would follow this procedure: Create a new polygon file with those two sets merged and labeled with one symbol, rerun the report on the merged and relabeled polygons, and populate the database based on the new data output from the combined polygons.

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 19A_ML would be used to populate NASIS.



Elevation (m)

MUSYM	Q5	Q10	Q25	Q50	Q75	Q90	Q95
122	1232	1233	1257	1296	1299	1307	1337
124	1231	1232	1233	1239	1297	1299	1300
124_ML	1295	1295	1296	1298	1299	1301	1302
19A	1235	1238	1243	1250	1264	1281	1295
19A_ML	1234	1236	1241	1249	1263	1280	1292

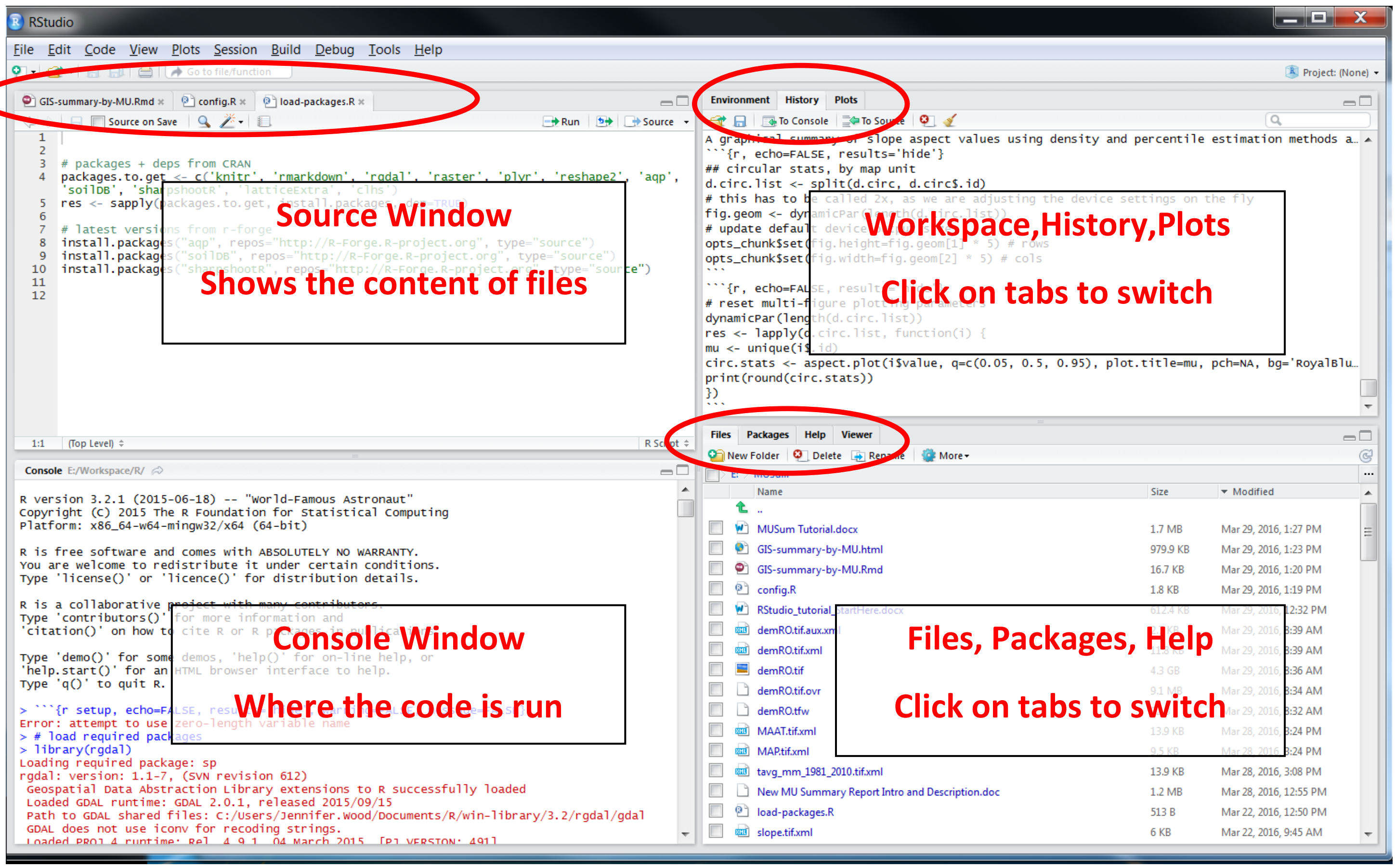
After inspecting the output, you may decide that delineations should be combined in different ways. In that case, you could create a new map unit polygon file with different sets of delineations belonging to newly configured map units. The instruction for how to do that are available from the Regional office. 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

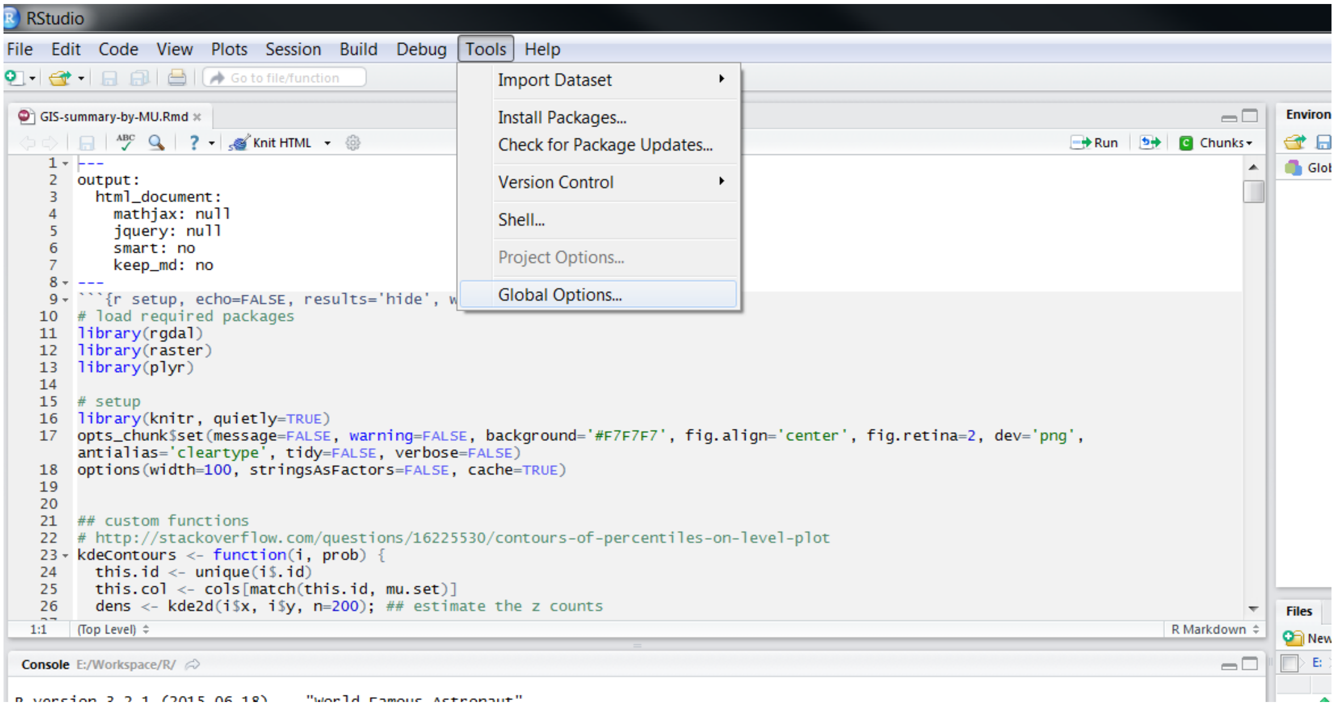
Check that R Studio is installed on your computer. Request the installation from IT, if not.
Open R Studio: Start – All Programs - Rstudio. The default display for R studio has four windows. We recommend a particular layout that can be specified from the Tools-Options menu. Check that the Options settings 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

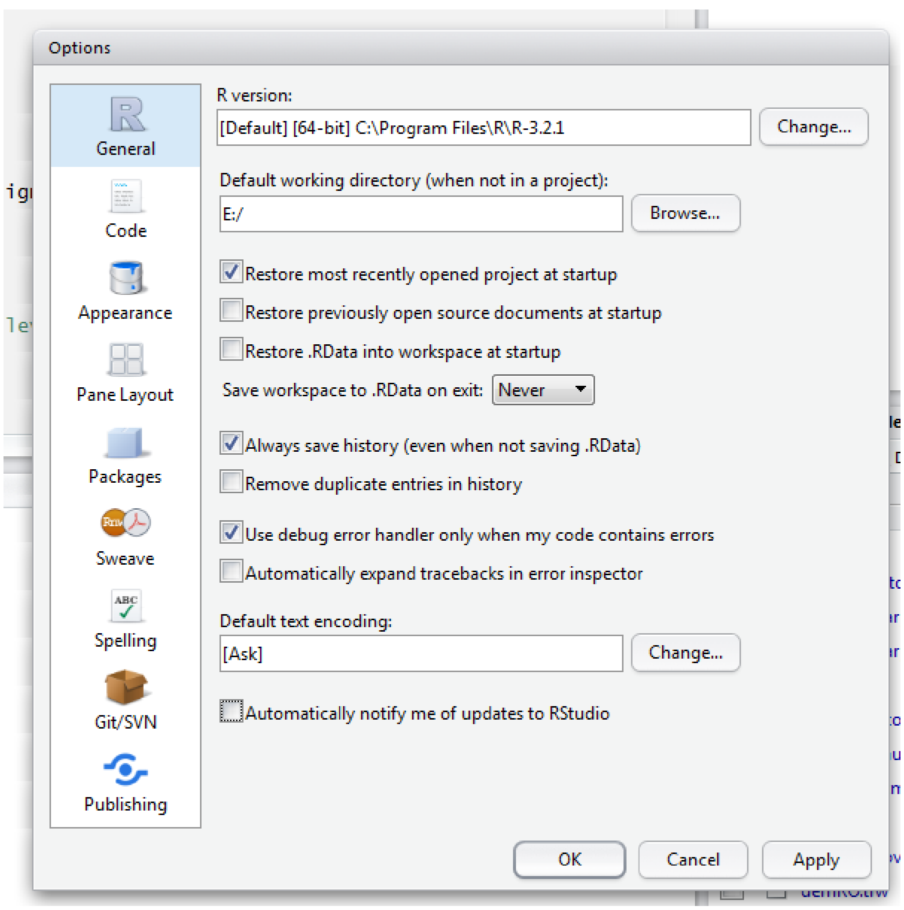
This is how your R Studio screen will be arranged after you set up your R Studio environment



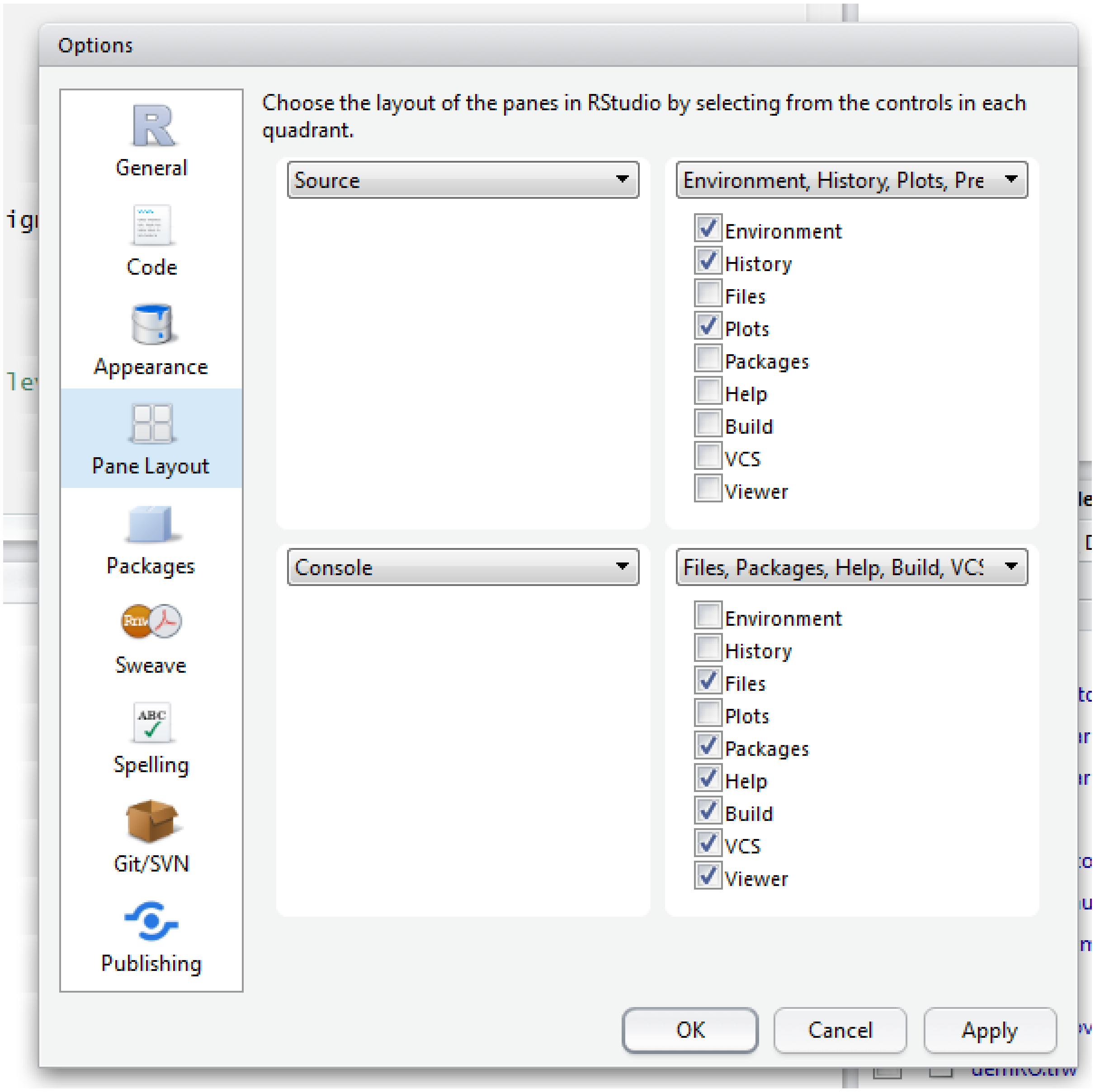
To set up your screen like the above screenshot, click Tools- Global Options



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

```
Directory Structure
¦--Files and folders provided to you on the MSSO shared data drive, as stored in Davis
¦   ¦--geodata/project data/
¦       ¦--MuSum_10m_MSSO/                                       10m DEM and derivatives
¦       ¦--MUSum_10m_SSR2/                                       10m DEM and derivatives
¦       ¦--MUSum_Curvature/                                         10 curvature classes
¦   ¦--geodata/elevation/derivatives/
¦       ¦--¦       ¦--Geomorphons/                                 10m landform elements
¦   ¦--geodata/climate/PRISM/
¦       ¦--PRISM800m/                                                  800m climate data
¦   ¦--geodata/land_use_land_cover/                          NLCD 2011 cropped to Region 2

¦--Local disk
¦--Report runs faster from copies on local disk
¦--Example of local disk folder structure for above folders and files listed above
¦   ¦--E:Workspace/geodata
¦       ¦--MUSum_10m_MSSO/                                       10m DEM and derivatives
¦       ¦--MUSum_10m_SSR2/                                       10m DEM and derivatives
¦       ¦--MUSum_Curvature/                                        10m curvature classes
¦       ¦--Geomorphons/                                            10m landform elements
¦       ¦--PRISM_800m/                                                 800m climate data
¦       ¦--land_use_land_cover/                              NLCD 2011 cropped to Region 2
¦--For files that you provide, please use the Project folders as follows
¦   ¦--Project folders
¦   ¦--[SDJR|MLRA Project Name/]
¦       ¦--Correlation/
¦       ¦--Images/
¦       ¦--Lab/
¦       ¦--Maps/
¦           ¦--map unit polygon data (.shp, .gdb)
¦       ¦--MUDs/
¦       ¦--Pedons/
¦       ¦--Vegetation/
¦       ¦--MUSum/                                Set this as working directory in R Studio
¦           ¦--map unit polygon data   Copy input map unit polygon data here (.shp, .gdb)
¦           ¦--MU-comparison/               The script from GitHub will create this folder
¦           ¦--¦--config.R                  The script from GitHub will put this file here
¦           ¦       This file tells the report where to look for raster and polygon data
¦           ¦Once you run the GitHub script, you can copy to subsequent MUSum project folders
¦           ¦--¦--report.Rmd                The script from GitHub will put this file here
¦           ¦                            This is the report you will “knit” to run the report
¦           ¦Once you run the GitHub script, you can copy to subsequent MUSum project folders
¦           ¦(If you are notified that the report changed,you will need to get a new copy)
¦           ¦--report.html       Html report automatically placed in report.Rmd location
¦           ¦--output/                                Automatically created by the report.Rmd
¦           ¦                 .csv, .shp output files automatically placed in this folder
```

Appendix 3 – Example html report output file

Map units (MUSYM): 124, 122, 124_ML, 19A_ML, 19A

report version 2.1-2
2017-02-08

This report is designed to provide statistical summaries of the environmental properties for one or more map units. Summaries are based on raster data extracted from fixed-density sampling of map unit polygons. Please see the document titled *R-Based Map Unit Summary Report Introduction and Description* for background and setup.

Map Unit Polygon Data Source

MU.Polygons	File.or.Feature
E:/Workspace/Project_Folder/MUSum	MUs_for_analysis

Raster Data Sources

Variable	File	inMemory	ContainsMU	Moran.I
Mean Annual Air Temperature (degrees C)	E:/Workspace/geodata/PRISM_800m/final_MAAT_800m.tif	TRUE	TRUE	0
Mean Annual Precipitation (mm)	E:/Workspace/geodata/PRISM_800m/final_MAP_mm_800m.tif	TRUE	TRUE	0
Effective Precipitation (mm)	E:/Workspace/geodata/PRISM_800m/effective_precipitation_800m.tif	TRUE	TRUE	0
Frost-Free Days	E:/Workspace/geodata/PRISM_800m/ffd_mean_800m.tif	TRUE	TRUE	0
Growing Degree Days (degrees C)	E:/Workspace/geodata/PRISM_800m/gdd_mean_800m.tif	TRUE	TRUE	0
Elevation (m)	E:/Workspace/geodata/DEM_KLM_int_AEA.tif	FALSE	TRUE	0
Slope Gradient (%)	E:/Workspace/geodata/Slope_KLM_int_AEA.tif	FALSE	TRUE	0
Geomorphon Landforms	E:/Workspace/geodata/Geomorphons/forms10_region2.tif	FALSE	TRUE	0
Curvature Classes	E:/Workspace/geodata/MU_Curvature/curvature_classes_10_class_region2.tif	FALSE	TRUE	0
NLCD 2011	E:/Workspace/geodata/land_use_land_cover/nlcd_2011_cropped.tif	FALSE	TRUE	0
Slope Aspect (degrees)	E:/Workspace/geodata/Aspect_KLM_int_AEACopy.tif	FALSE	TRUE	0

Area Summaries

Consider increasing the sampling density (**1 points/ac.**) in `config.R` if there are unsampled polygons. Note that the actual sampling density will always be slightly lower than the target sampling density defined in `config.R` .

Map Unit Acreage by Polygon

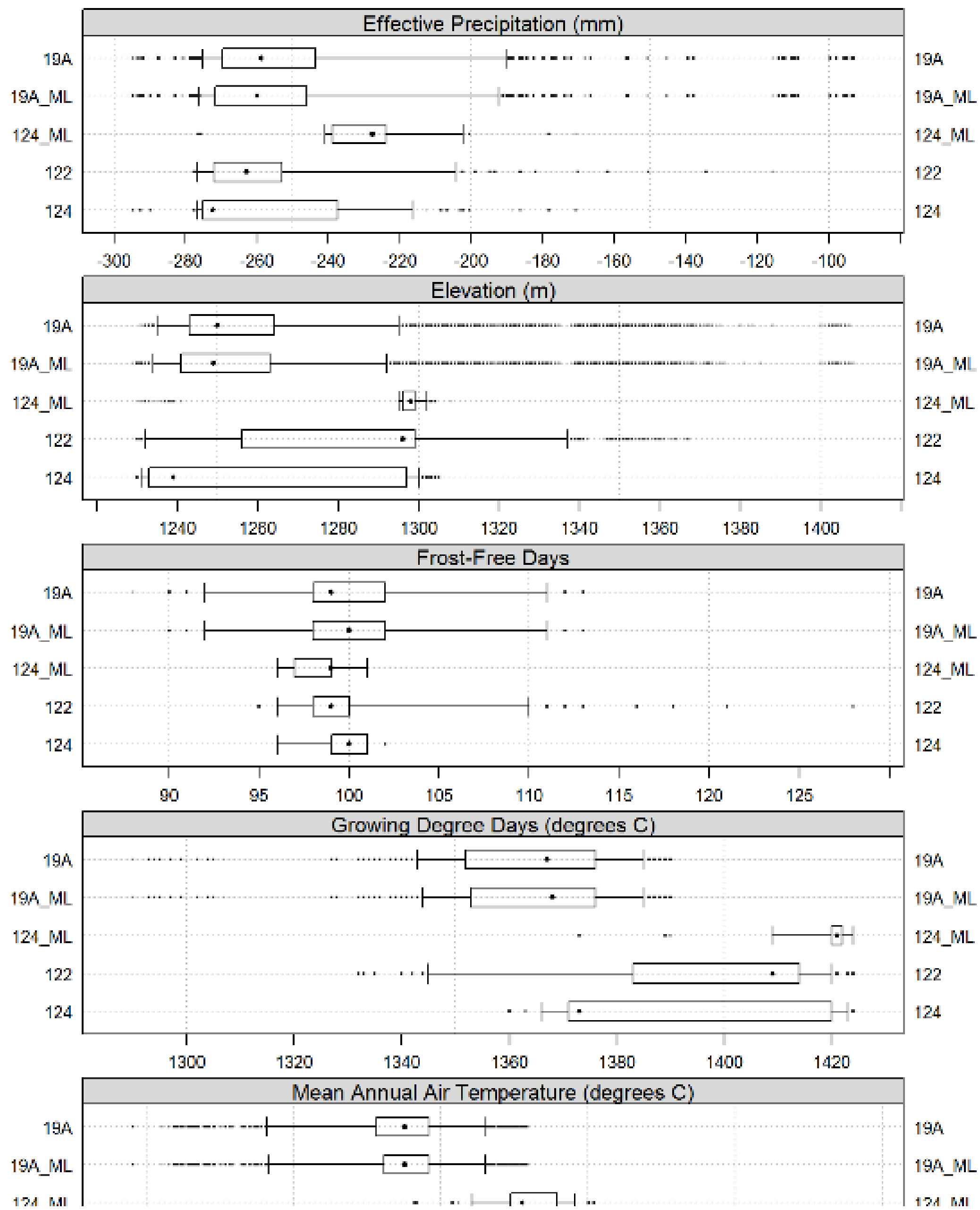
MUSYM	Min	Q5	Q25	Median	Q75	Q95	Max	Total Area	Samples	Polygons	Polygons Not Sampled	Mean Sample Dens.
124	3	3	8	23	95	813	1810	3740	3029	18	0	0.81
122	12	16	46	79	158	1316	3618	7329	6307	24	0	0.86
124_ML	3	3	6	12	31	461	638	1238	1062	14	0	0.86
19A_ML	1	6	17	43	188	872	5442	33442	27844	134	1	0.83
19A	1	6	17	41	170	841	5442	30907	25778	129	1	0.83

Modified Box and Whisker Plots

Whiskers extend from the 5th to 95th percentiles, the body represents the 25th through 75th percentiles, and the dot is the 50th percentile. Notches (if enabled) represent an approximate confidence interval around the median, adjusted for spatial autocorrelation. Overlapping notches suggest that median values are not significantly different. This feature can be enabled by setting `correct.sample.size=TRUE` in `config.R`.

Suggested usage:

- Gauge overlap between map units in terms of boxes (25th-75th percentiles) and whiskers (5th-95th percentiles).
- Non-overlapping boxes are a strong indication that the central tendencies (of select raster data) differ.
- Distribution shape is difficult to infer from box and whisker plots, remember to cross-reference with density plots below.

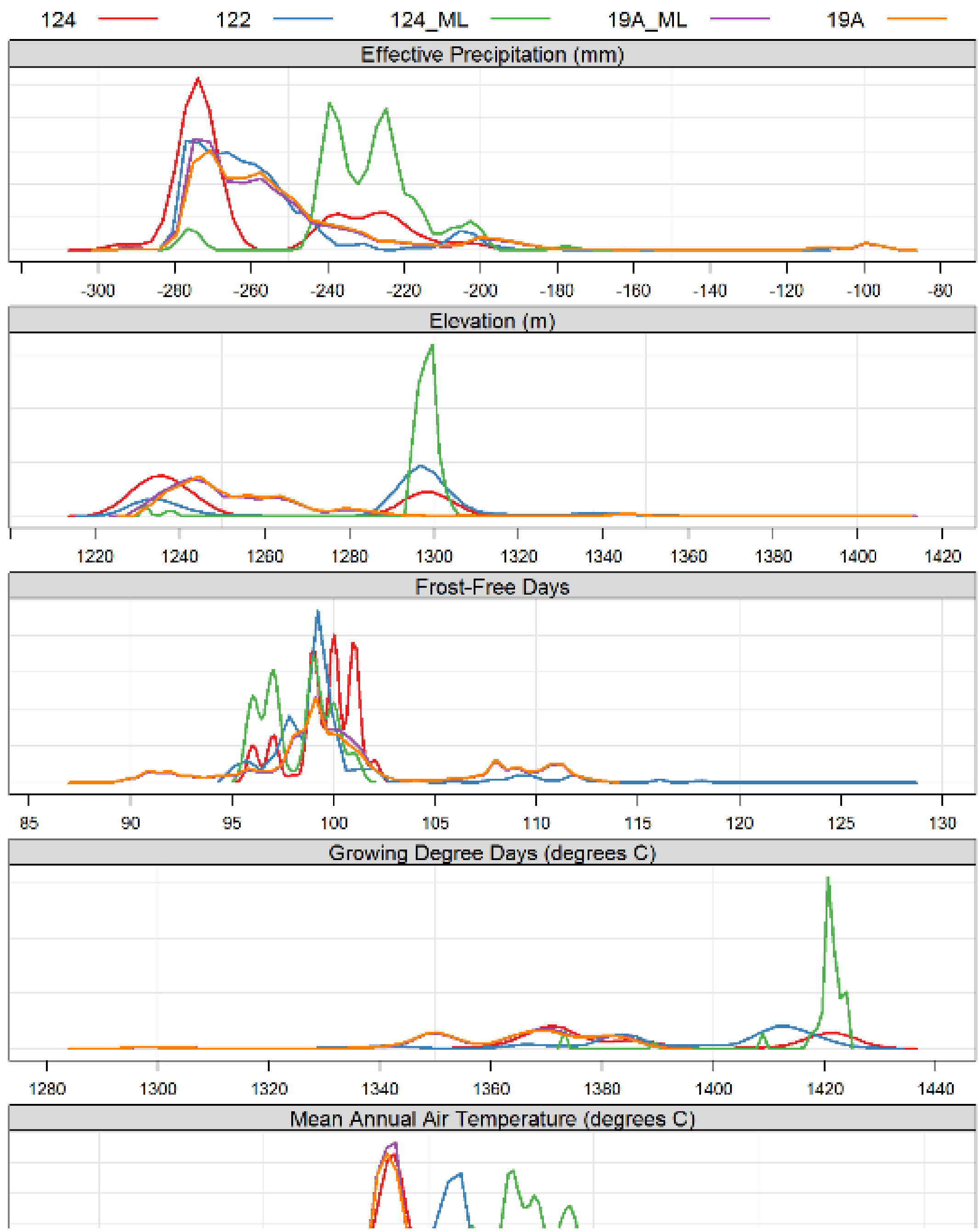


Density Plots

These plots are a smooth alternative (denisty estimation) to the classic "binned" (histogram) approach to visualizing distributions. Peaks correspond to values that are most frequent within a data set.

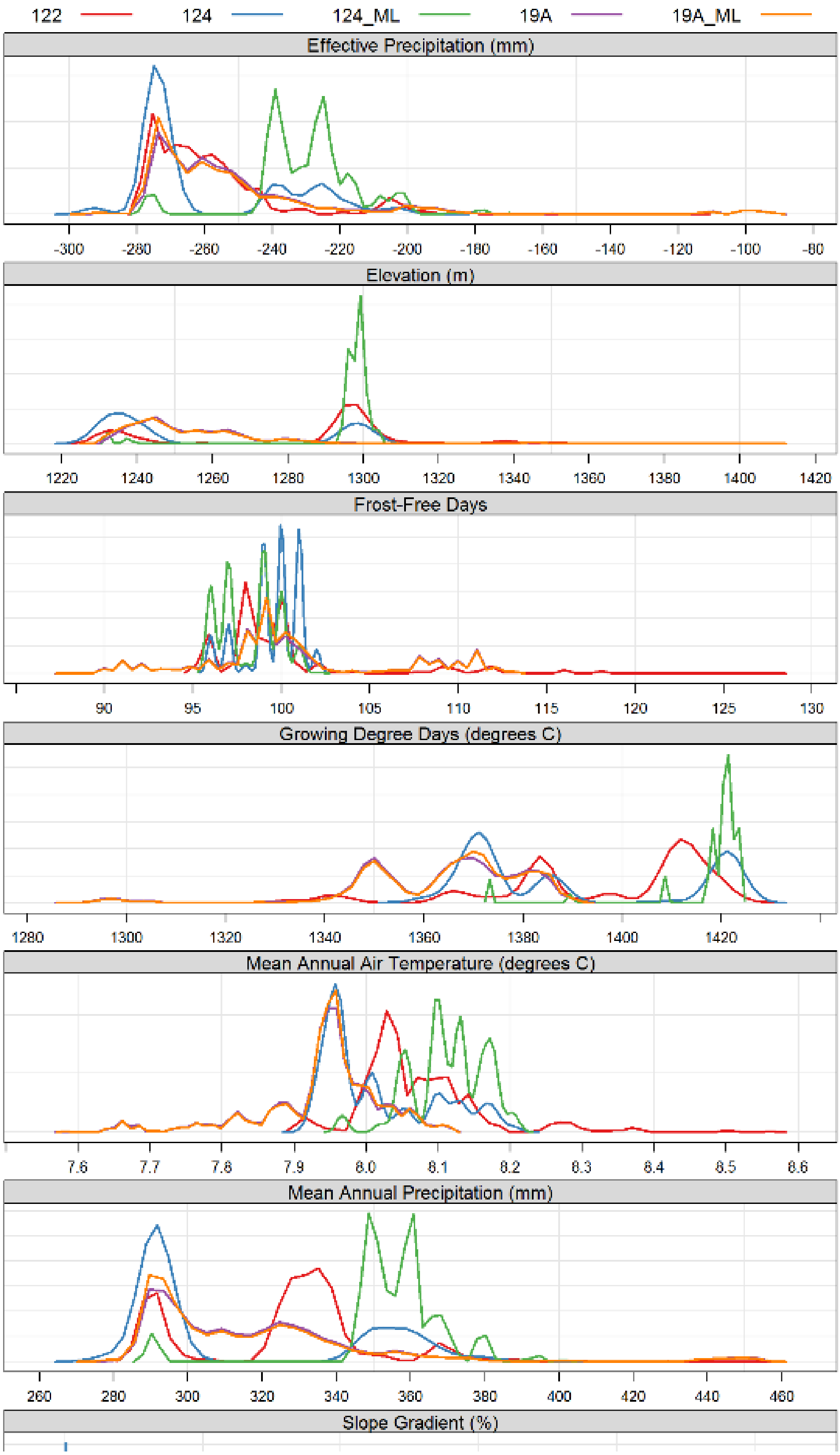
Suggested usage:

- Density plots depict a more detailed summary of distribution shape.
- When making comparisons, be sure to look for:
- multiple peaks
- narrow peaks vs. wide "mounds"
- short vs. long "tails"



Density Plots

These plots are a smooth alternative (denisty estimation) to the classic "binned"" (histogram) approach to visualizing distributions. Peaks coorrospond to values that are most frequent within a dataset.



Tabular Summaries

Table of select percentiles, by variable. In these tables, headings like "Q5" can be interpreted as the the "5th percentile"; 5% of the data are less than this value. The 50th percentile ("Q50") is the median.

Median Values

MUSYM	Effective Precipitation (mm)	Elevation (m)	Frost-Free Days	Growing Degree Days (degrees C)	Mean Annual Air Temperature (degrees C)	Mean Annual Precipitation (mm)	Slope Gradient (%)
124	-272.38	1239	100	1373	7.97	292	0
122	-263.06	1296	99	1409	8.05	330	0
124_ML	-227.69	1298	99	1421	8.11	357	0
19A_ML	-259.88	1249	100	1368	7.95	308	0
19A	-258.85	1250	99	1367	7.95	312	0

Effective Precipitation (mm)

MUSYM	Q5	Q10	Q25	Q50	Q75	Q90	Q95
124	-276.76	-276.47	-275.36	-272.38	-237.40	-223.87	-216.43
122	-276.71	-276.45	-272.05	-263.06	-253.02	-238.87	-204.32
124_ML	-241.08	-241.08	-238.87	-227.69	-223.87	-208.53	-202.39
19A_ML	-276.07	-274.58	-271.63	-259.88	-246.00	-214.34	-192.28
19A	-275.23	-274.17	-269.65	-258.85	-243.71	-206.72	-189.96

Elevation (m)

MUSYM	Q5	Q10	Q25	Q50	Q75	Q90	Q95
124	1231	1232	1233	1239	1297	1299	1300
122	1232	1233	1256	1296	1299	1306	1337
124_ML	1295	1295	1296	1298	1299	1301	1302
19A_ML	1234	1236	1241	1249	1263	1280	1292
19A	1235	1238	1243	1250	1264	1281	1295

Frost-Free Days

MUSYM	Q5	Q10	Q25	Q50	Q75	Q90	Q95
124	96	97	99	100	101	101	101
122	96	96	98	99	100	102	110
124_ML	96	96	97	99	99	100	101
19A_ML	92	95	98	100	102	110	111
19A	92	95	98	99	102	110	111

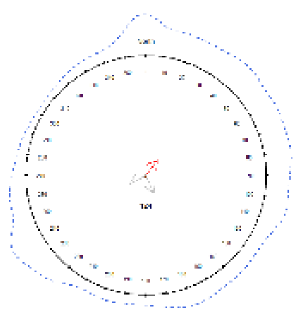
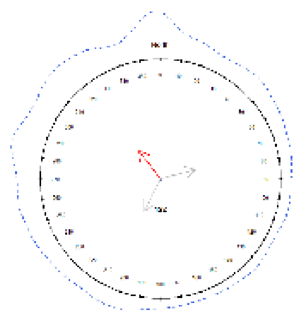
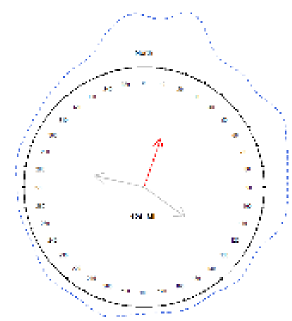
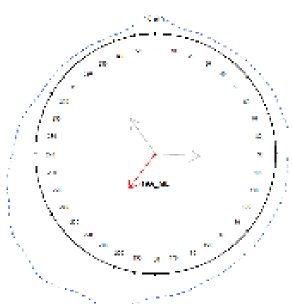
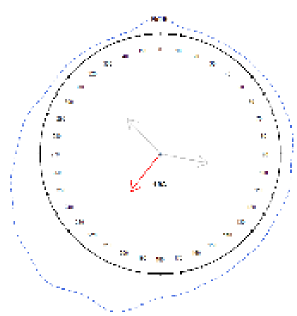
Slope Aspect

A graphical summary of slope aspect values using density and percentile estimation methods adapted to circular data. Spread and central tendency are depicted with a combination of (circular) kernel density estimate (dashed blue lines) and arrows. The 50th percentile value is shown with a red arrow and the 10th and 90th percentile values are shown with gray arrows. Arrow length is proportional to the strength of directionality.

Suggested usage:

- Check circular density estimates for peaks: this suggests directionality.
- These summaries are meaningless when slope values are less than approximately 3%.
- These summaries are (mostly) meaningless when arrow lengths are short; e.g. low directionality.

MUSYM	10%	50%	90%
124	157	35	241
122	75	322	207
124_ML	126	19	283
19A_ML	324	218	93
19A	317	219	101

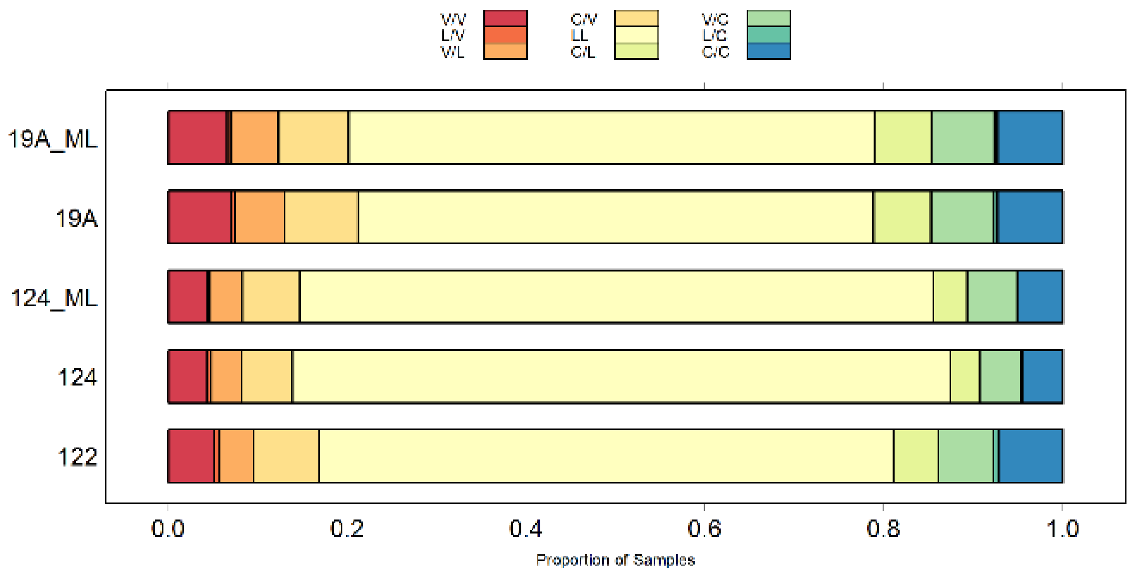






Slope Shape (Curvature) Summary

The classes were generated using a 5x5 moving window, from a regional 30m or 10m, integer DEM. The precision may be limited, use with caution. See instructions for using your own (higher resolution) curvature classification raster.

Suggested usage:

- Use the graphical summary to identify patterns, then consult the tabular representation for specifics.
- The conventions used here are "C" = concave, "L" = linear, and "V" = convex; "down slope" / "across slope".
- Window size has a significant impact on reported curvature classes; larger windows = more generalization.
- Curvature class and colors are aligned with an idealized *shedding → accumulating* hydrologic gradient.



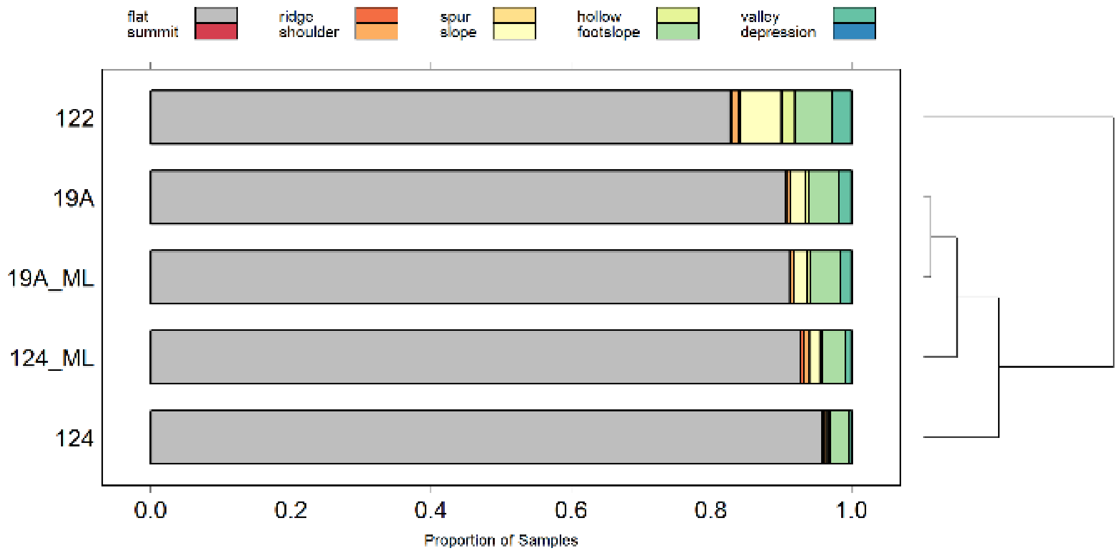
	V/V	L/V	V/L	C/V	LL	C/L	V/C	L/C	C/C
124	0.04	0.00	0.03	0.06	0.74	0.03	0.04	0.00	0.05
122	0.05	0.01	0.04	0.07	0.64	0.05	0.06	0.01	0.07
124_ML	0.04	0.00	0.04	0.06	0.71	0.04	0.06	0.00	0.05
19A_ML	0.07	0.00	0.05	0.08	0.59	0.06	0.07	0.00	0.07
19A	0.07	0.00	0.06	0.08	0.58	0.06	0.07	0.00	0.07

Geomorphon Landform Classification

Proportion of samples within each map unit that correspond to 1 of 10 possible landform positions, as generated via geomorphon algorithm. Landform classification by this method is scale-invariant and is therefore not affected by computational window size selection.

Suggested usage:

- Use the graphical summary to identify patterns, then consult the tabular representation for specifics.
- "Flat" is based on a 3% slope threshold.
- Map units are organized (in the figure) according to the similarity, computed from proportions of each landform position.
- The dendrogram on the right side of the figure describes relative similarity. "Lower branch height" (e.g. closer to the right-hand side of the figure) denotes more similar landform positions.
- Landform class labels and colors are aligned with an idealized *shedding* → *accumulating* hydrologic gradient.



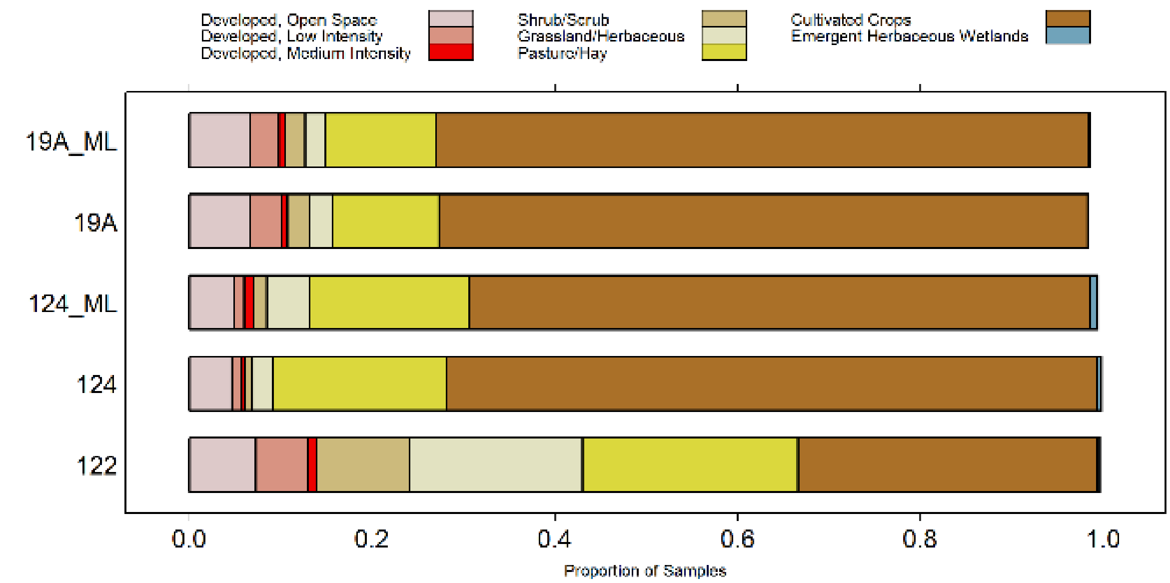
	flat	summit	ridge	shoulder	spur	slope	hollow	footslope	valley	depression
124	0.96	0	0	0.00	0	0.00	0.00	0.03	0.00	0
122	0.83	0	0	0.01	0	0.06	0.02	0.05	0.03	0
124_ML	0.93	0	0	0.01	0	0.02	0.00	0.03	0.01	0
19A_ML	0.91	0	0	0.01	0	0.02	0.00	0.04	0.02	0
19A	0.91	0	0	0.01	0	0.02	0.00	0.04	0.02	0

Landform "signatures": these are created from the top 75% fraction of (sampled) landform classes, in decreasing order.

	landform signature
124	flat
122	flat
124_ML	flat
19A_ML	flat
19A	flat

Landcover Summary

These values are from the 2011 NLCD (30m) database.



	Developed, Open Space	Developed, Low Intensity	Developed, Medium Intensity	Shrub/Scrub	Grassland/Herbaceous	Pasture/Hay	Cultivated Crops	Emergent Herbaceous Wetlands
124	0.05	0.01	0.00	0.01	0.02	0.19	0.71	0.00
122	0.07	0.06	0.01	0.10	0.19	0.24	0.33	0.00
124_ML	0.05	0.01	0.01	0.01	0.05	0.17	0.68	0.01
19A_ML	0.07	0.03	0.01	0.02	0.02	0.12	0.71	0.00
19A	0.07	0.03	0.01	0.02	0.02	0.12	0.71	0.00

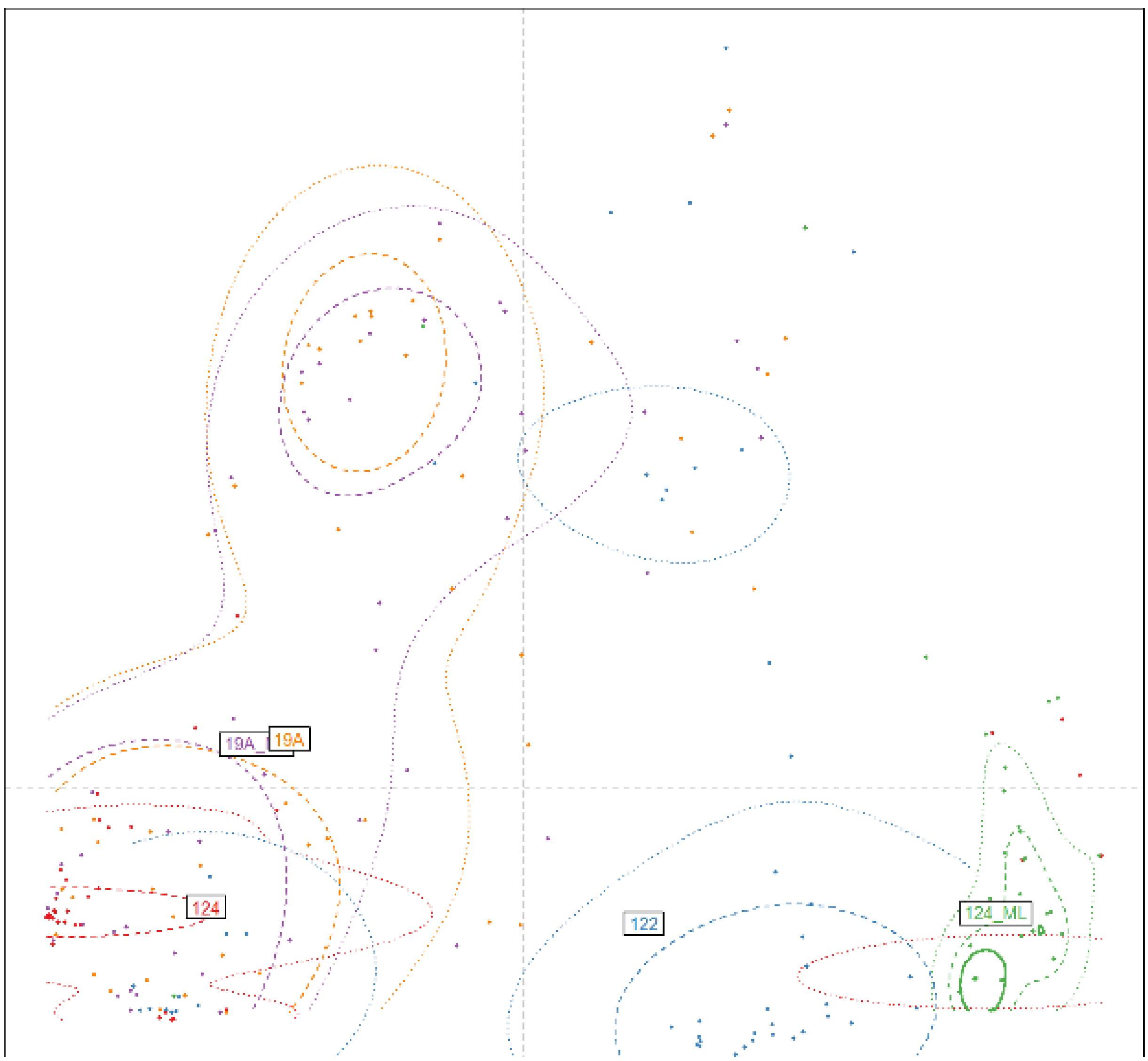
Multivariate Summary

This plot displays the similarity of the map units across the set of environmental variables used in this report. The contours contain 75% (dotted line), 50% (dashed line), and 25% (solid line) of the points in an optimal 2D projection of multivariate data space. Data from map units with more than 1,000 samples are (sub-sampled via cLHS). Map units with very low variation in environmental variables can result in tightly clustered points in the 2D projection. It is not possible to generate a multivariate summary when any sampled variable (e.g. slope) has a near-zero variance. See this chapter, from the new *Statistics for Soil Scientists* NEDS course, for an soils-specific introduction to these concepts.

Suggested usage:

- The relative position of points and contours are meaningful; absolute position will vary each time the report is run.
- Colors match those used in the density plots above. Be sure to cross-reference this figure with density plots.
- Look for "diffuse" vs. "concentrated" clusters: these suggest relatively broadly vs. narrowly defined map unit concepts.
- Multiple, disconnected contours (per map unit) could indicate errors or small map unit separated by large distances. Check for multiple peaks in the associated density plots.
- Nesting of clusters (e.g. smaller cluster contained by larger cluster) suggests superset/subset relationships.
- Overlap is proportional to similarity.

Ordination of Raster Samples (cLHS Subset) with 25%, 50%, 75% Density Contours

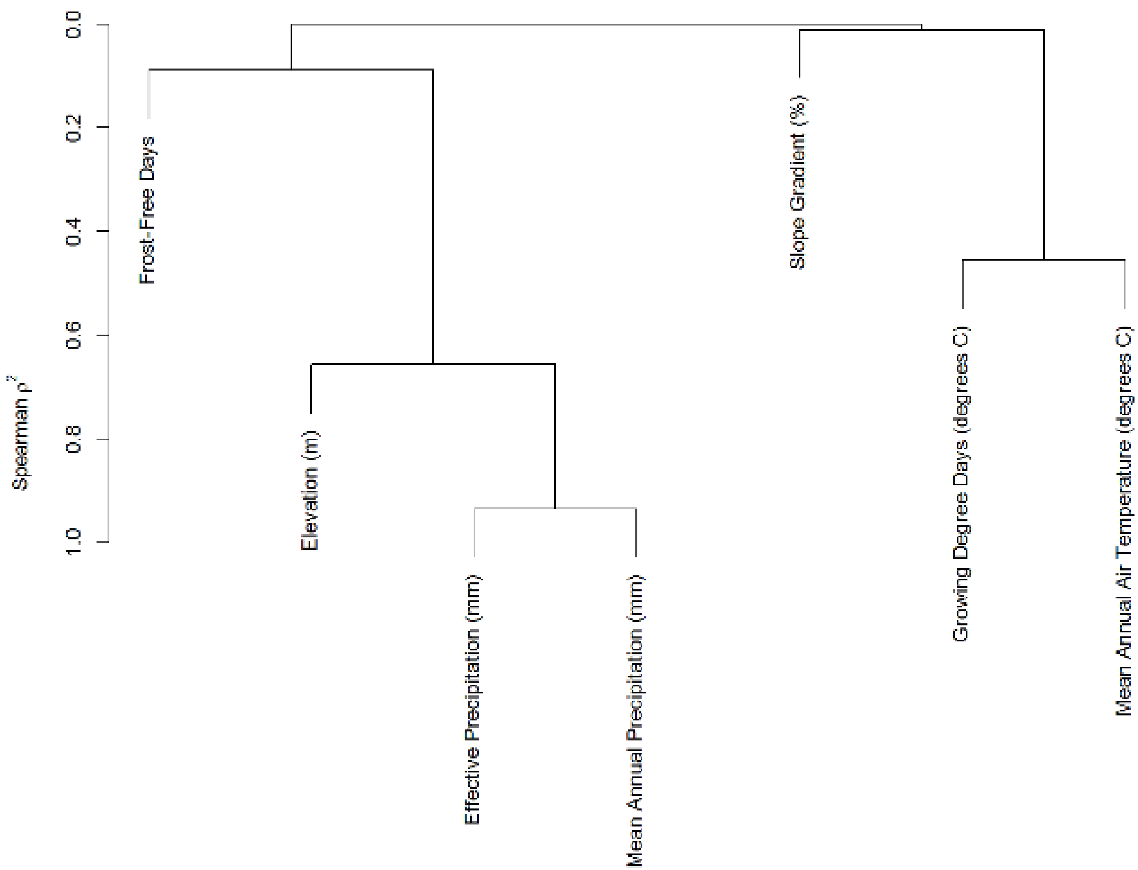


Raster Data Correlation

The following figure highlights shared information among raster data sources based on Spearman's Ranked Correlation coefficient. Branch height is associated with the degree of shared information between raster data.

Suggested usage:

- Look for clustered sets of raster data: typically PRISM-derived and elevation data are closely correlated.
- Highly correlated raster data sources reduce the reliability of the "raster data importance" figure.

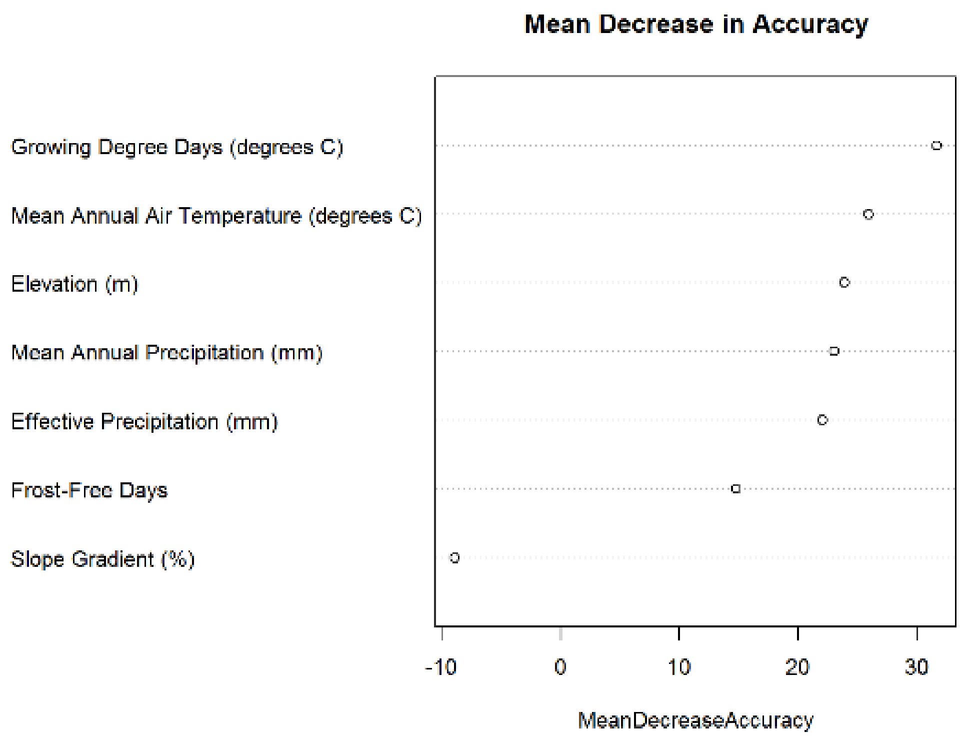


Raster Data Importance

The following figure ranks raster data sources in terms of how accurately each can be used to discriminate between map unit concepts.

Suggested usage:

- Map unit concepts are more consistently predicted (by supervised classification) using those raster data sources with relatively larger "Mean Decrease in Accuracy" values.
- Highly correlated raster data sources will "compete" for positions in this figure. For example, if *elevation* and *mean annual air temperature* are highly correlated, then their respective "importance" values are interchangeable.



Polygon Summaries

A shapefile is generated each time a report is run ("polygons-with-stats-XXX" where "XXX" is the set of map units symbols listed in `config.R`) that contains several useful summaries, computed by polygon. Polygons are uniquely identified by the `pID` column. Median raster values are given, with data source names abbreviated to conform to the limitations of DBF files:

pID	MUSYM	EstmtMASTC	AnnlBmRdMJ	EffctvPrcp	Elevationm	FrostFrDys	GrwngDgrDC	MnAnnlArTC	MnAnnlPrcp	SlopeGrdnt
1	7011	17.75	67118.27	-413.72	139.0	319	2656	16.58	443	5
2	7011	16.66	67126.19	-231.08	298.0	293	2583	16.20	612	2
3	7089	15.46	56270.39	-84.02	321.5	297	2557	16.17	747	34
4	7011	17.02	66833.37	-270.86	242.0	306	2629	16.41	588	2

There are several columns containing the proportions of each landform element (geomorphons algorithm), the most likely ("ml_landfrm") landform element, and the Shannon entropy associated with landform proportions. The Shannon entropy value can be used to judge the relative "landform purity" of a delineation: smaller values are associated with more homogeneous delineations. Equal proportions of all landform elements (within a polygon) would result in a Shannon entropy value of 1.

pID	MUSYM	flat	summit	ridge	shoulder	spur	slope	hollow	footslope	valley	depression	ml_landfrm	shannon_h
1	7011	0.00	0.01	0.03	0.00	0.05	0.12	0.19	0.04	0.52	0.03	valley	0.44
2	7011	0.34	0.00	0.02	0.02	0.05	0.15	0.04	0.20	0.17	0.00	flat	0.62
3	7089	0.00	0.04	0.08	0.00	0.32	0.45	0.08	0.00	0.02	0.02	slope	0.54
4	7011	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.08	0.85	0.00	valley	0.17

In the case of un-sampled polygons (very small delineations or too low sampling density), an additional shapefile will be saved in the output folder with a prefix of "un-sampled-". This file contains those polygons that were not allocated any sampling points and thus not included in the report summaries.

Polygon Quality Control

A shapefile is generated each time a report is run ("poly-qc-XXX" where "XXX" is the set of map units symbols listed in `config.R`) that contains the proportion of samples outside the 5-95% percentile range. In the attribute table there is one column per raster data source and one row per map unit delineation.

The 5-95% percentile range for each map unit is derived from the samples across all polygons with the corresponding map unit symbol. Proportions of samples outside the range within individual polygons are given for each (continuous) raster data source. Data source names are abbreviated to conform to the limitations of DBF files. Polygons are uniquely identified by the `pID` column.

Assuming one has sufficient polygons and samples to characterize the data distribution, and that the data are roughly normally distributed, one would expect that 10% of samples across the extent of a particular map unit will fall outside the 5-95% percentile range. Individual delineations that have more than 10-15% of samples outside the range for one or more raster data sources may need to be investigated to see that they fit the map unit concept. It is expected that some delineations will occur at the margins of the map unit extent and therefore may have higher proportions outside the range. Expert judgement is required to determine whether action should be taken to resolve any potentially problematic delineations.

pID	MUSYM	EffctvPrcp	Elevationm	FrostFrDys	GrwngDgrDC	MnAnnlArTC	MnAnnlPrcp	SlopeGrdnt
1	7089	0.00	0	0.05	0.03	0.05	0.00	0.00
2	7089	0.00	0	0.01	0.00	0.00	0.00	0.00
3	7011	0.05	0	0.05	0.01	0.01	0.04	0.03

This document is based on `sharpshootR` version 1.2.
Report configuration and source code are hosted on GitHub.
Sampling time: 1.7 mins