MLRA COMPARISON REPORT 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

MLRA COMPARISON REPORT TUTORIAL

This is a tutorial for a follow up report to the Map Unit Summary Report tool that has the same functionality, except that it compares polygons at the MLRA scale. It includes access to R script that downloads the files you need – four pre-sampled raster files that make it convenient to compare coarse environmental and selected soils data across MLRAs within the continental United States (CONUS).

The rasters sampled to provide the data files used in this report are: 800m cell sized PRISM data and derivatives, elevation and a Geomorphon product derived from the 30 m DEM data, and the National Land Cover Data (NLCD), 800m gridded soil properties from SSURGO and STATSGO. These are coarse enough data to facilitate reasonable computing time, and matched to the coarse scale (1:250,000) scale of the MLRA polygons.

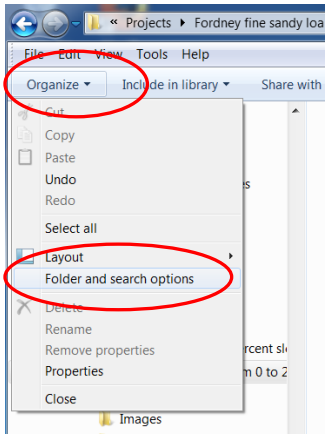
The prerequisite to this tutorial is reading the document titled “Map Unit Summary Report – Part 1 – Background” and following the instructions to run the MUSum tool in the document titled “Map Unit Summary Report - Part 2 – Instructions”.

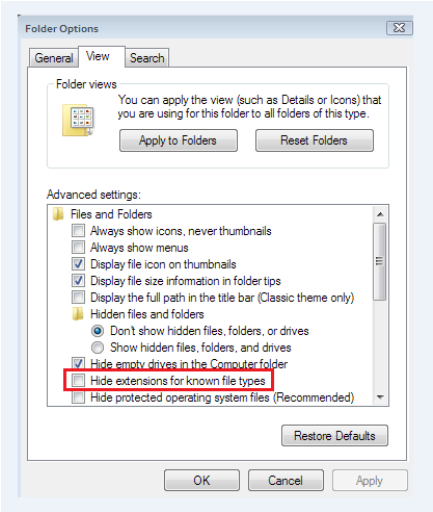
Tip: Make this document easier to use by checking the Navigation Pane box under the View menu

Step 1. If you have not done this already, set up R Studio and your Windows folder options

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.

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. Then, in the View tab, uncheck the box to hide file extensions



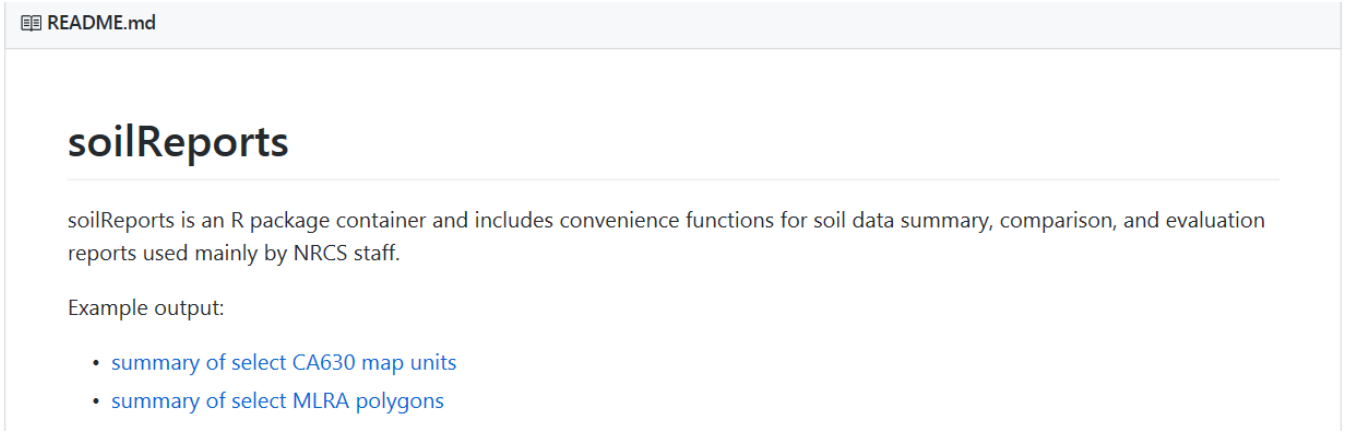


Step 2. Get the report and required pre-sampled raster files.

All R files are obtained by running lines of script found on the NCSS GitHub, at this location:

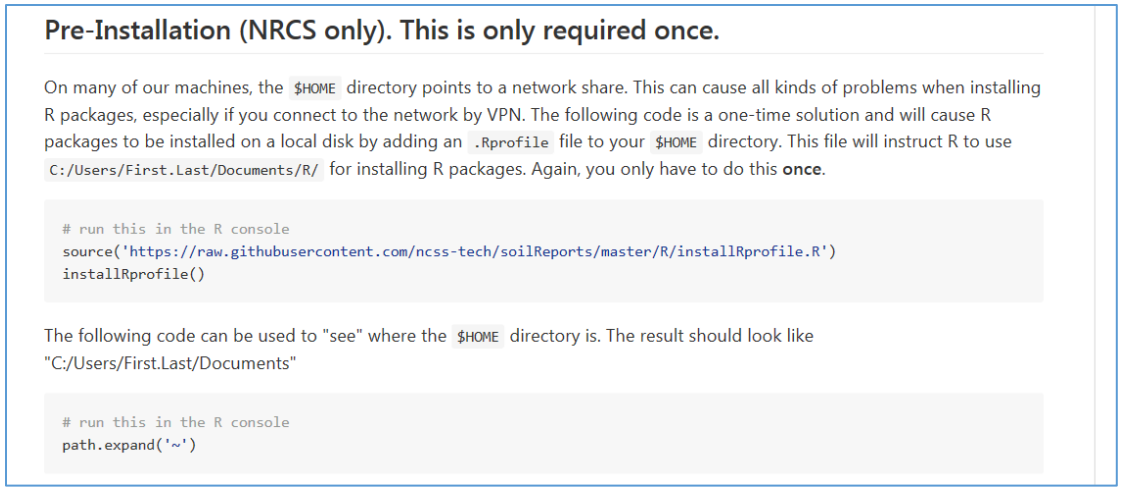
https://github.com/ncss-tech/soilReports

Scroll down to the section titled soilReports



Open R Studio if it is not open already.

If you have been using the MUSum Report, or have already done this for your computer, you can skip this step. 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.



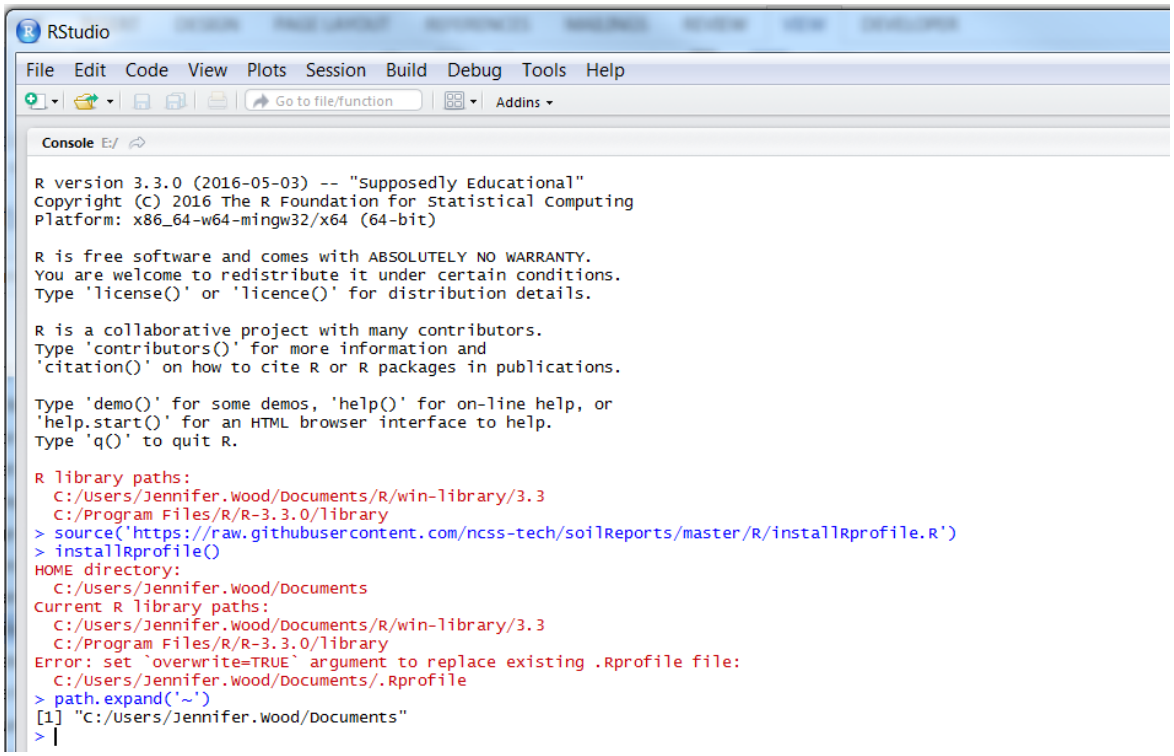
Pre-Installation (NRCS only). This is only required once.

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('~')
```



This step is 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)
```

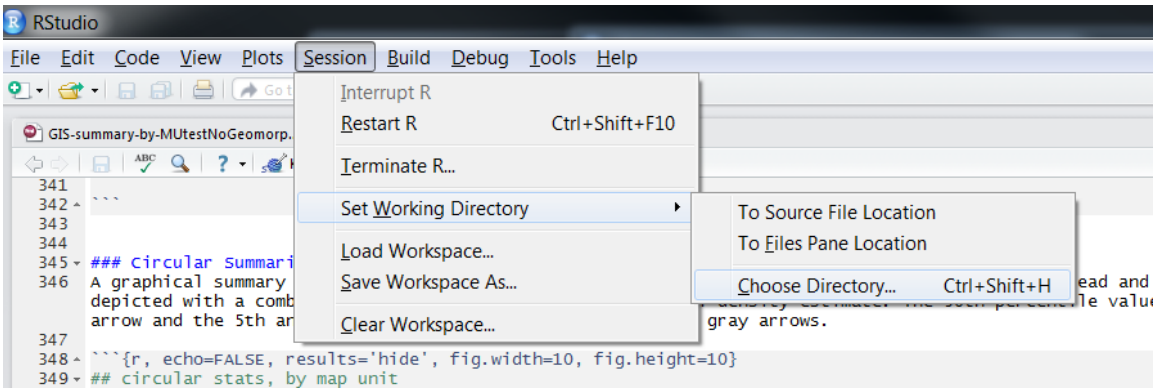
```
> install.packages('devtools', dep=TRUE)
Installing package into 'C:/Users/Jennifer.Wood/Documents/R/win-library/3.3'
(as 'lib' is unspecified)
Warning in install.packages :
  dependency 'BiocInstaller' is not available
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/devtools_1.12.0.zip'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

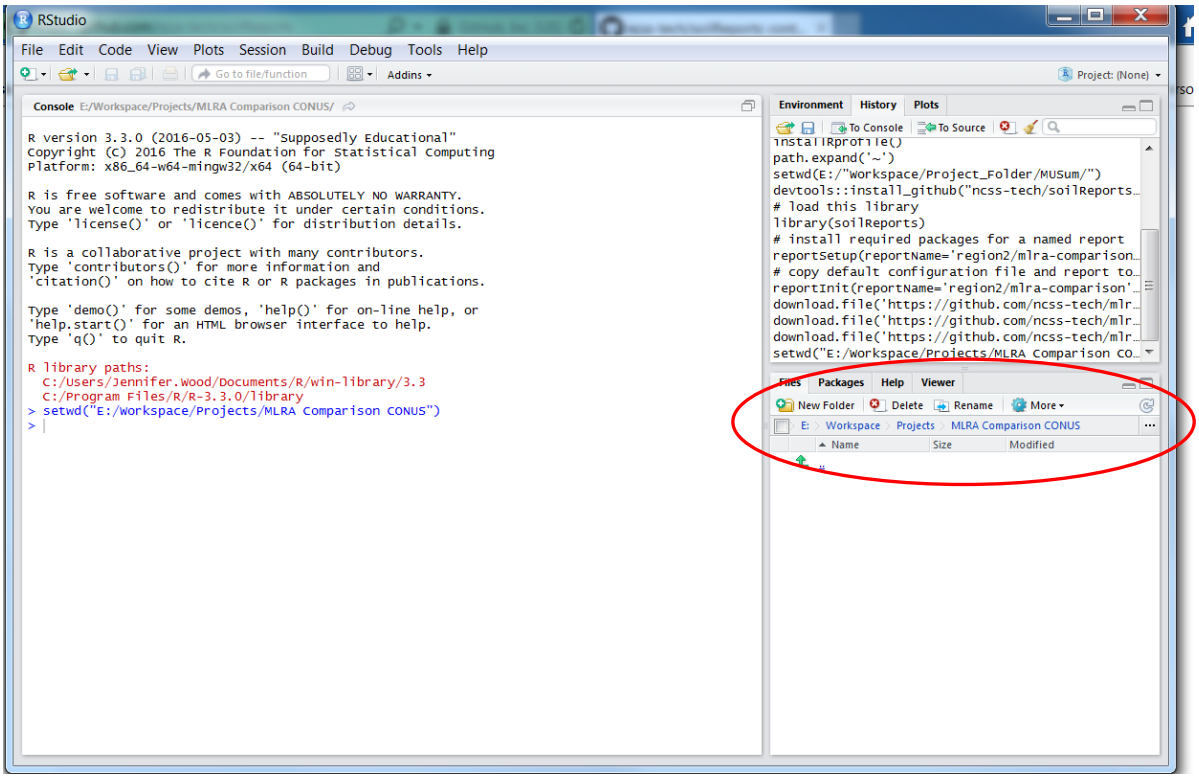
Set the working directory

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

Set the working directory to a folder titled something like 'MLRA Comparison CONUS'.

See Appendix 2 for an example file stucture.





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.

Get the report needed to compare MLRAs in the CONUS. You only need to do this once or if you need a new copy of the report

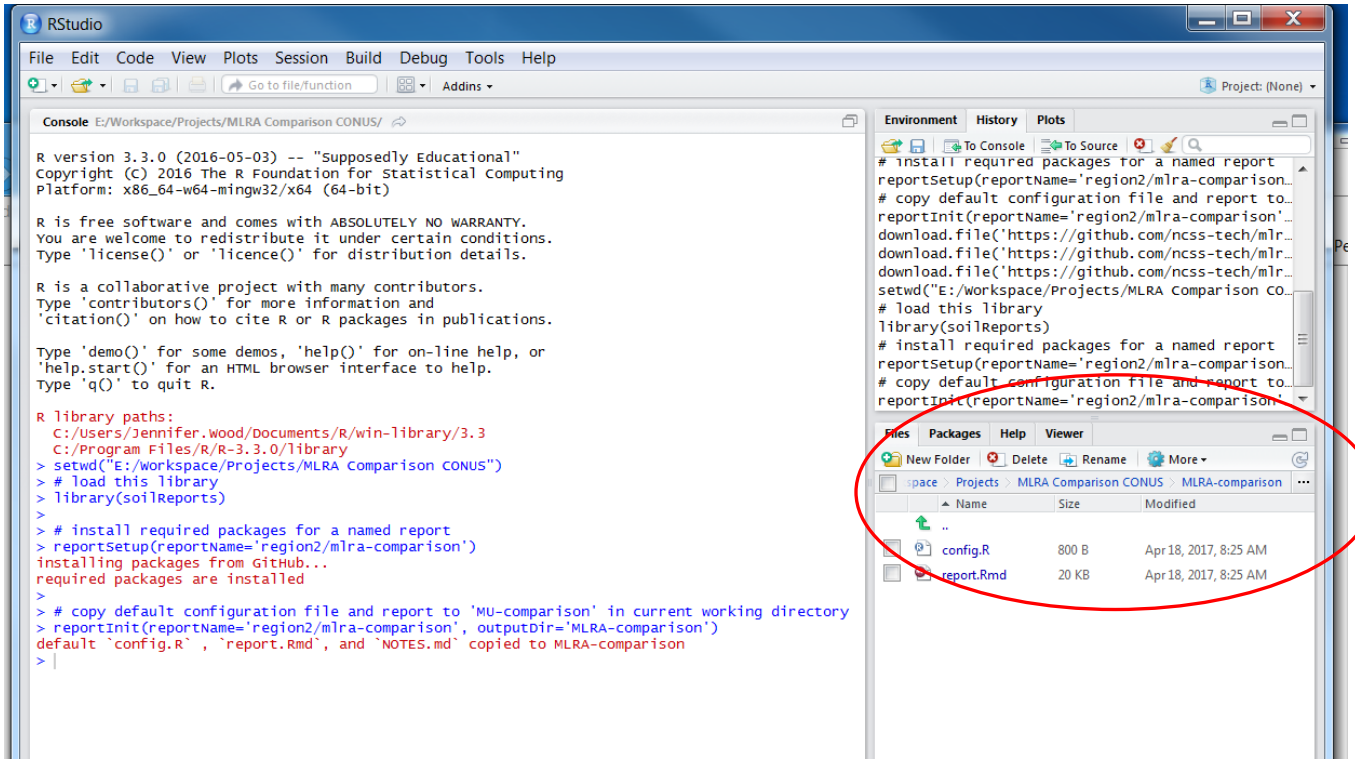
In the soilReports GitHub page, scroll down to the section titled “MLRA Comparison/Summary Report” towards the bottom of the page. Copy the chunk of script in the grey block and paste it into your R console window, and press enter.

MLRA Comparison/Summary Report.

This report was designed to assist with comparisons between MLRA concepts using a pre-made raster sample database. You will need to put these database files into the same folder as `report.Rmd`. MLRA selection is done within `config.R`. Contact Dylan Beaudette (dylan.beaudette at ca.usda.gov) for questions or comments.

```
# load this library
library(soilReports)

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

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

You will now see a folder named MLRA-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.



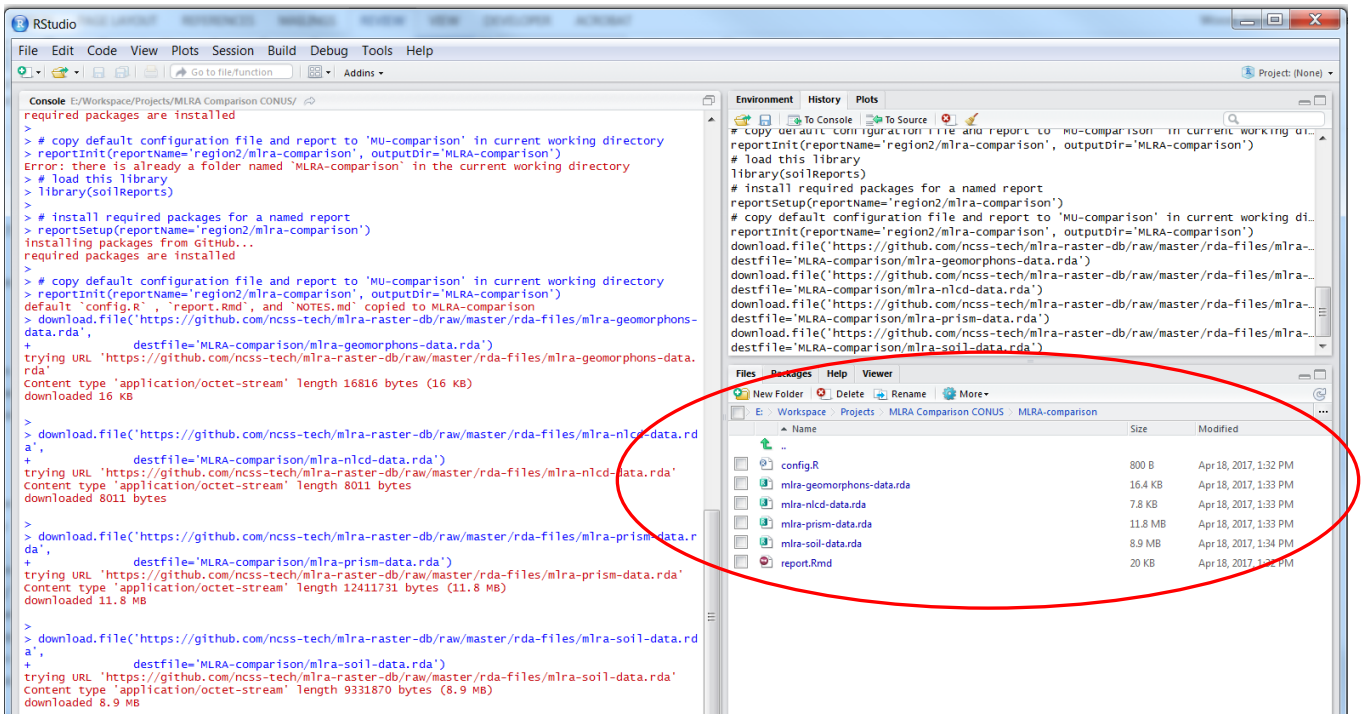
Get the 4 pre-sampled raster files from the GitHub

In the [soilReports GitHub page], scroll down to the second chunk of script in the section "MLRA Comparison/Summary Report" towards the bottom of the page. Copy the chunk of script in the grey block and paste it into your R console window, and press enter.

Download the three raster sample databases into the directory created above.

```
download.file('https://github.com/ncss-tech/mlra-raster-db/raw/master/rda-files/mlra-geomorphons-data.rda',
destfile='MLRA-comparison/mlra-geomorphons-data.rda')

download.file('https://github.com/ncss-tech/mlra-raster-db/raw/master/rda-files/mlra-nlcd-data.rda',
destfile='MLRA-comparison/mlra-nlcd-data.rda')

download.file('https://github.com/ncss-tech/mlra-raster-db/raw/master/rda-files/mlra-prism-data.rda',
destfile='MLRA-comparison/mlra-prism-data.rda')

download.file('https://github.com/ncss-tech/mlra-raster-db/raw/master/rda-files/mlra-soil-data.rda',
destfile='MLRA-comparison/mlra-soil-data.rda')
```

The report will place these files into the MLRA-comparison folder



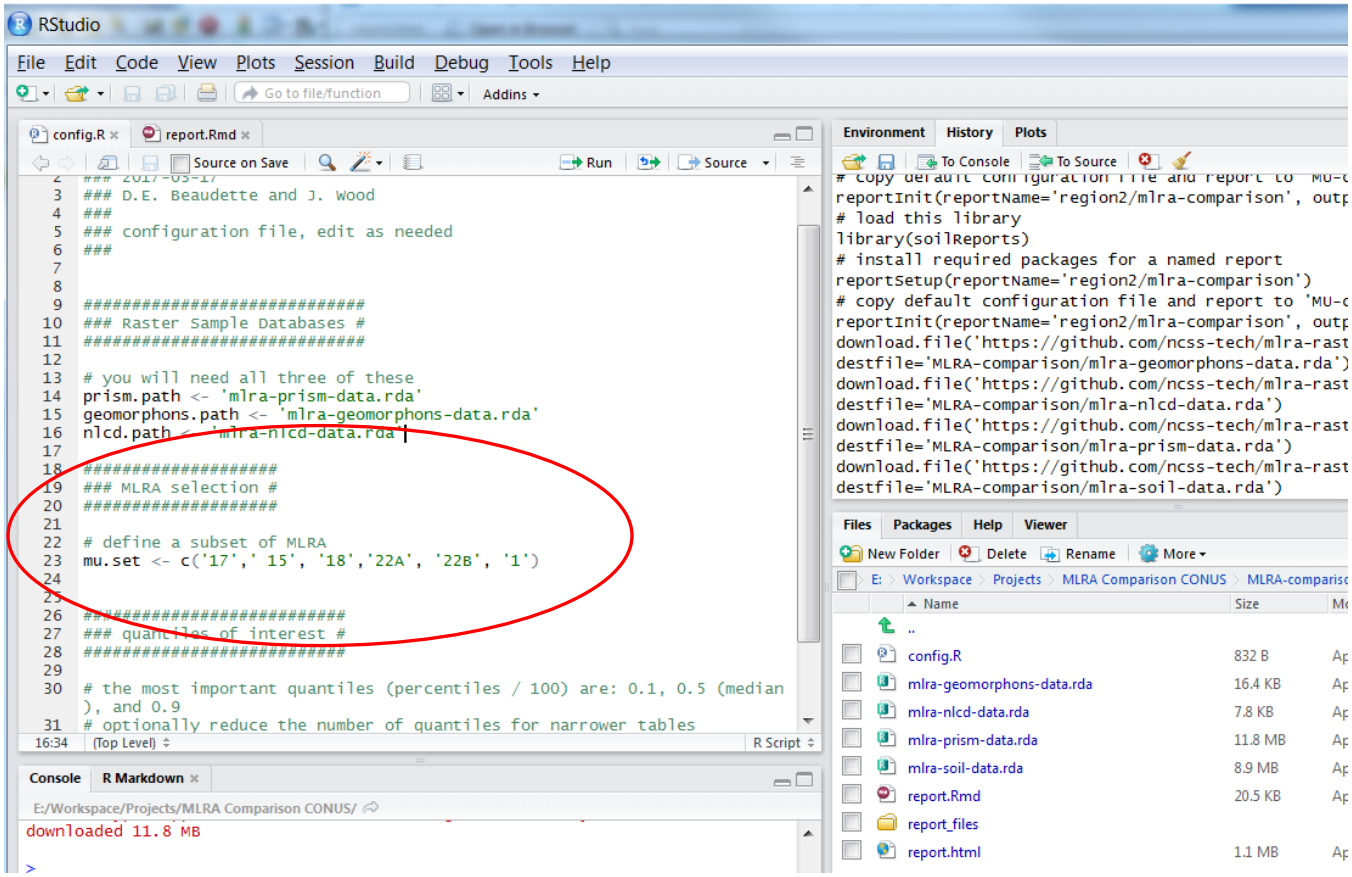
Step 3. Open the config.R files and specify the MLRAs of interest

Open the file titled ‘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 the section titled ###MLRA selection### around line 21, and edit the list. Because this report compares a group of MLRAs together, limit the list to about 6-8..

Tip: It can be useful to include an MLRA that is very different, for purposes of drawing attention to the scales in the output report

After editing the list of MLRAs, save the config.R file by clicking on File-Save or pressing Ctrl-S.

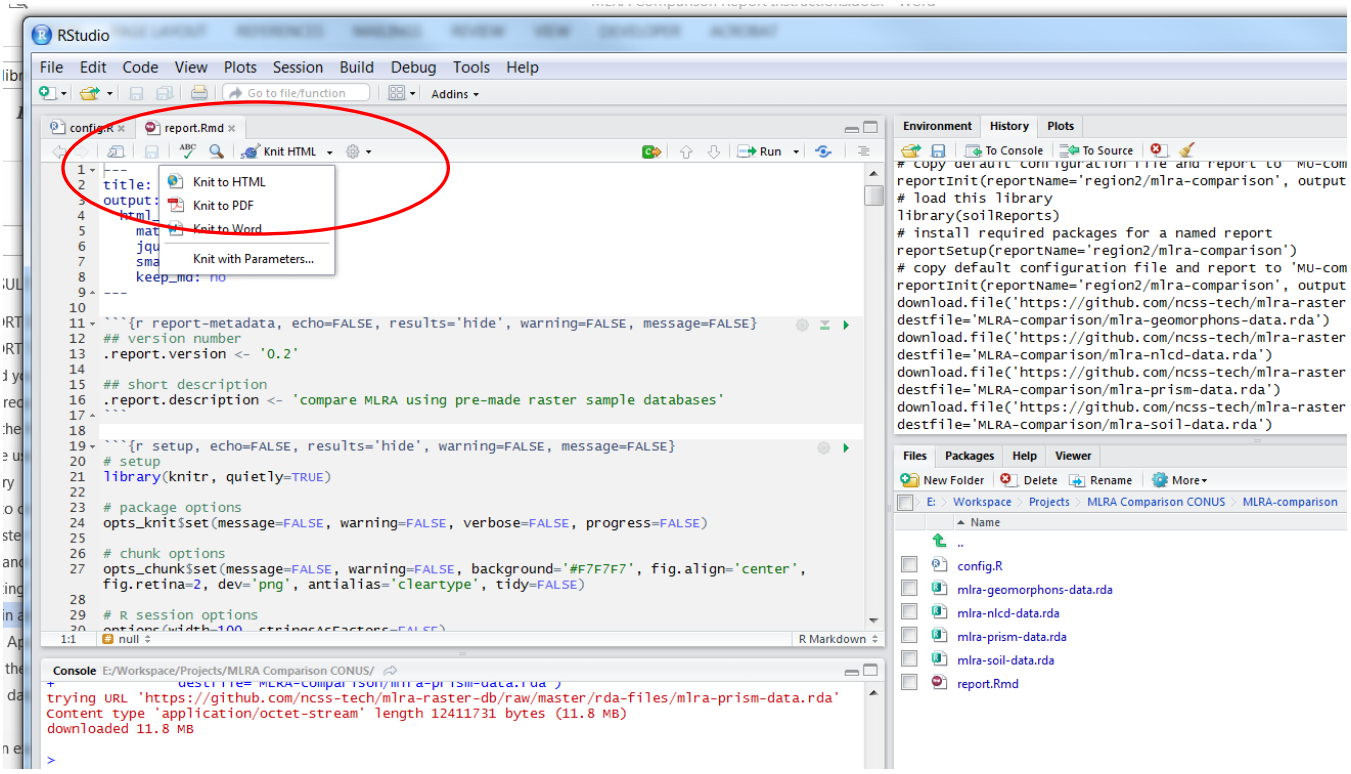


Step 4. Run the Report and evaluate the results

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

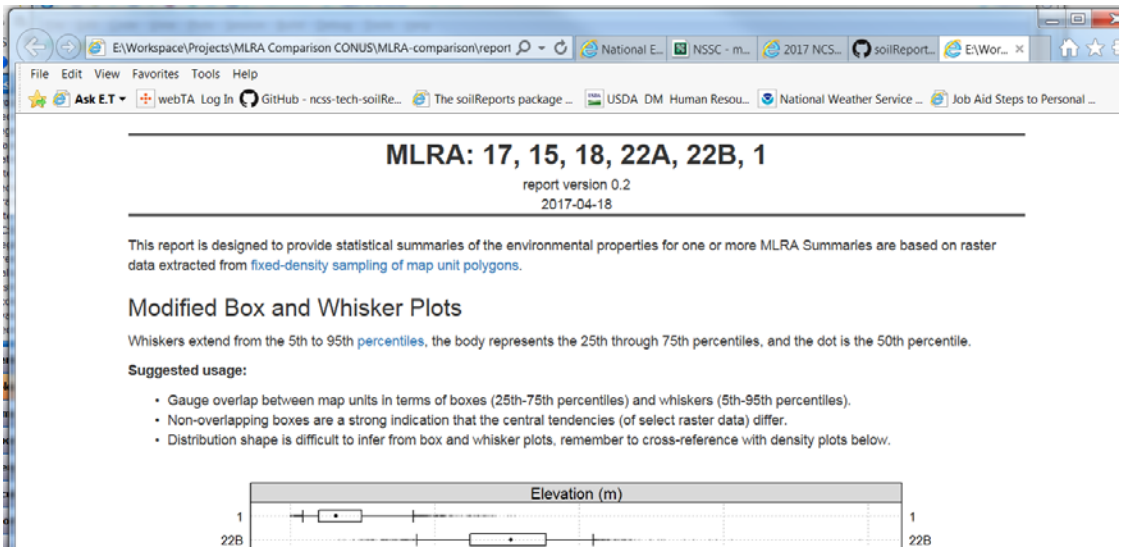
Open the file titled 'Report.Rmd', by either clicking File-Open or by clicking on the file in the File Window. It will open in the Source Window.

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.



The report will not take long to run. 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. When it is done, there might be some red warnings text that you can ignore.

An html file will be generated and will open automatically. A file titled 'report.html' will be generated in the MLRA-comparison folder.

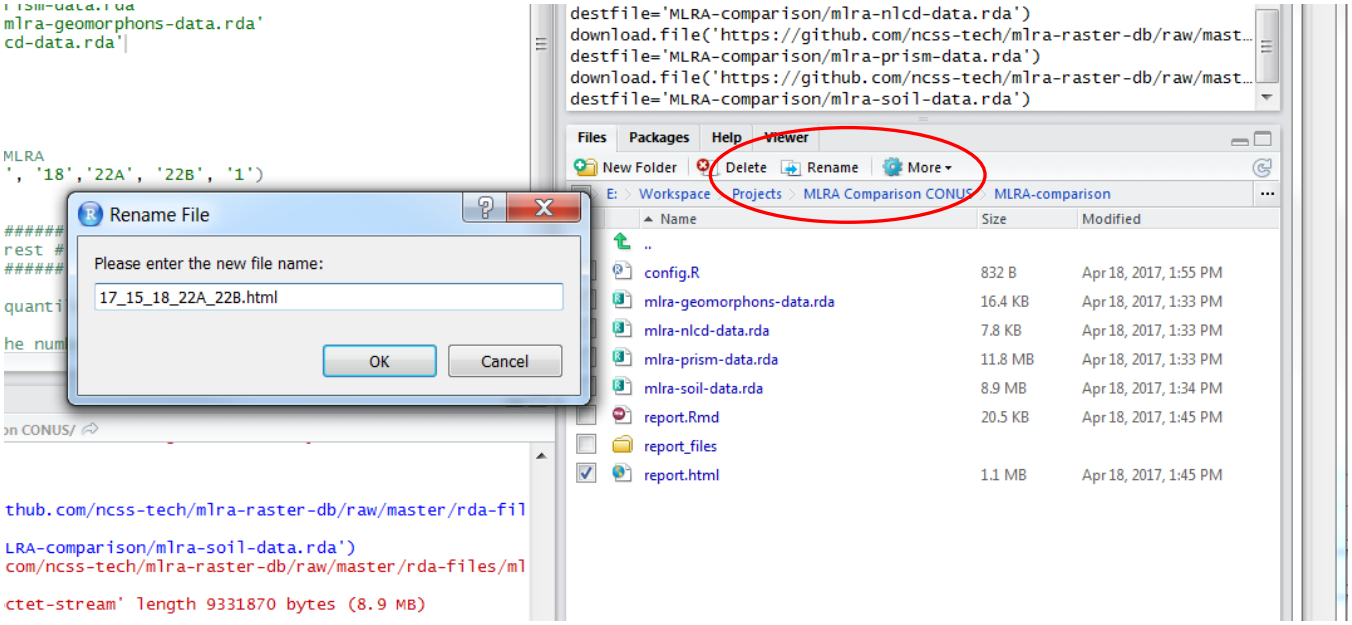


Running the report again and output file management

To run the report with a different set of MLRAs to compare, edit the list of MLRA's in the config.R files, then save the config.R file by going to File-Save or pressing Ctrl-S.

If you run the report again and haven't renamed the previous output, it will overwrite the report.html file in the MLRA-comparison folder.

Save the output file if you want to save the results for future reverence. You can use the File window menu bar to Delete, Rename, Copy, and Move files in R Studio. Click on the box next to the file to mark it for Renaming or another operation. You can also rename and move these files around in Windows Explorer.

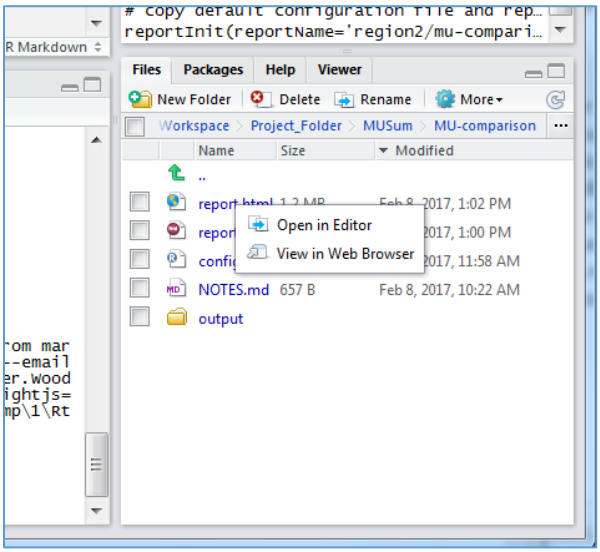


Inspect the report – see [Appendix 3](#) for example html file.

This output is similar to the ouput generated for map units in the Map Unit Summary report, except using polygons from the 1:250,000 scale MLRA shape file as well as coarse raster data. Thus, the output is necessarily a coarse look at the differences in various environmental data sets among the MLRAs specified. All the usual caveats about understanding the sources, accuracy, and precision of the underlying data apply here.

If you close the report after it first opens, and want to inspect the report again, 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.



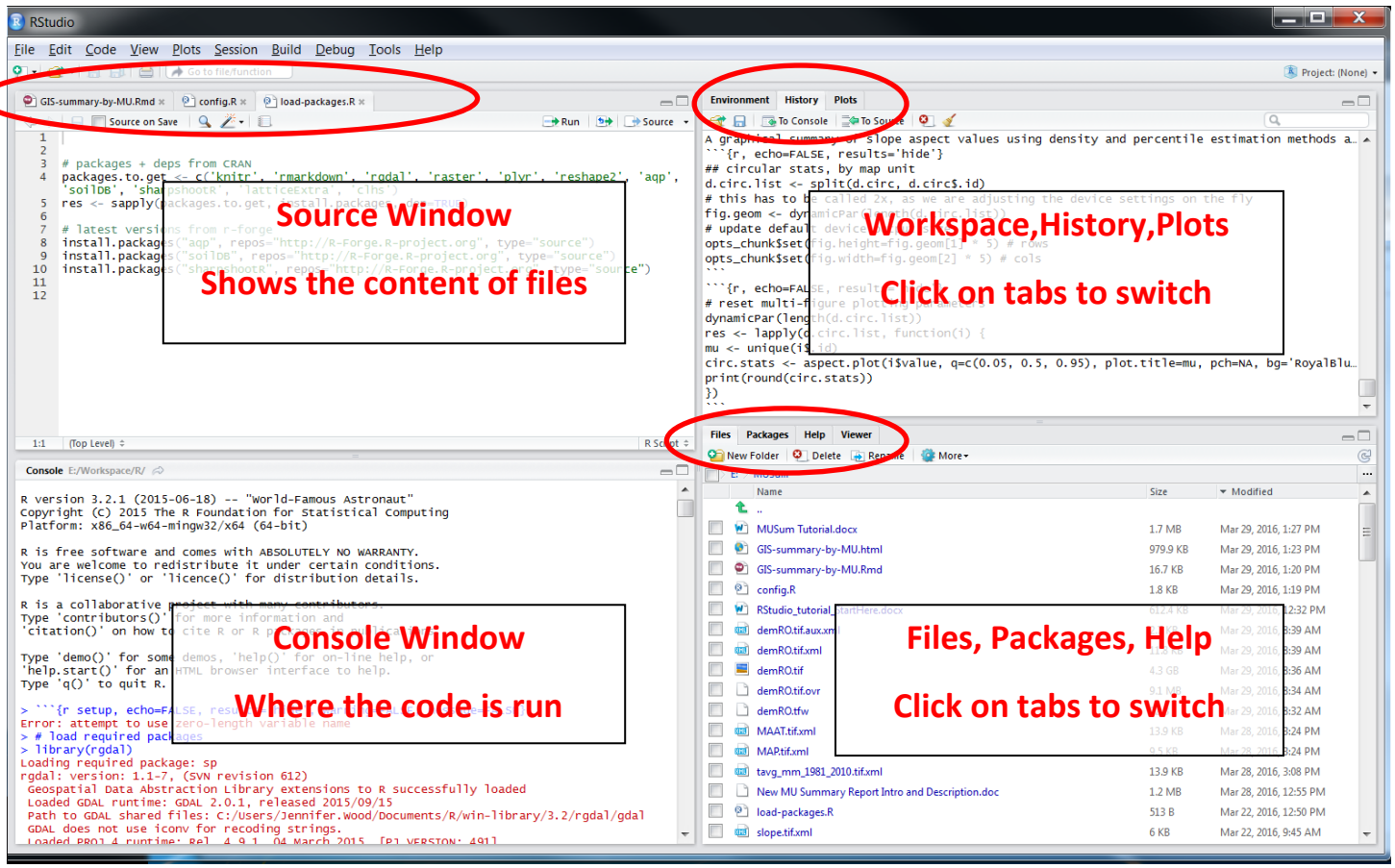
.

Appendix 1 - RStudio Tutorial - Getting started

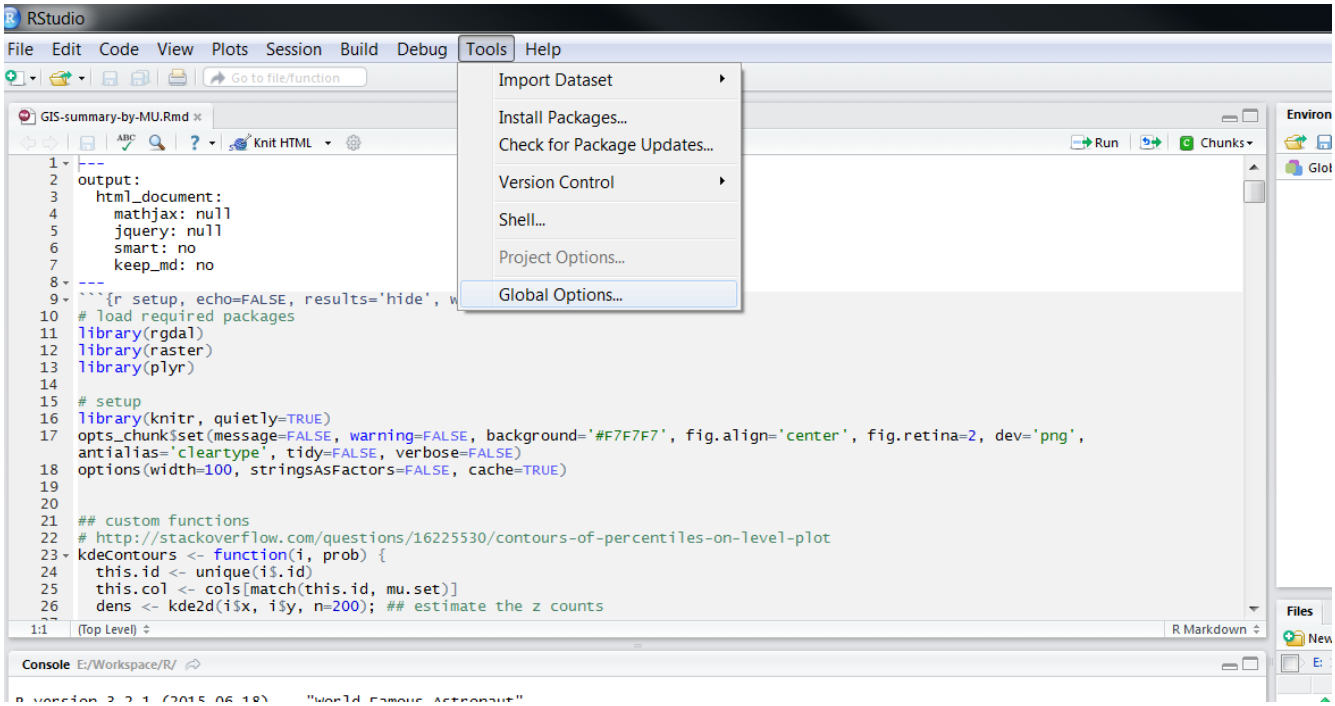
Check that R Studio is installed on your computer. Request the installation from IT, if not.
Open R Studio: Start – All Programs - Rstudio. The default display for R studio has four windows. We recommend a particular layout that can be specified from the Tools-Options menu. Check that the Options settings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

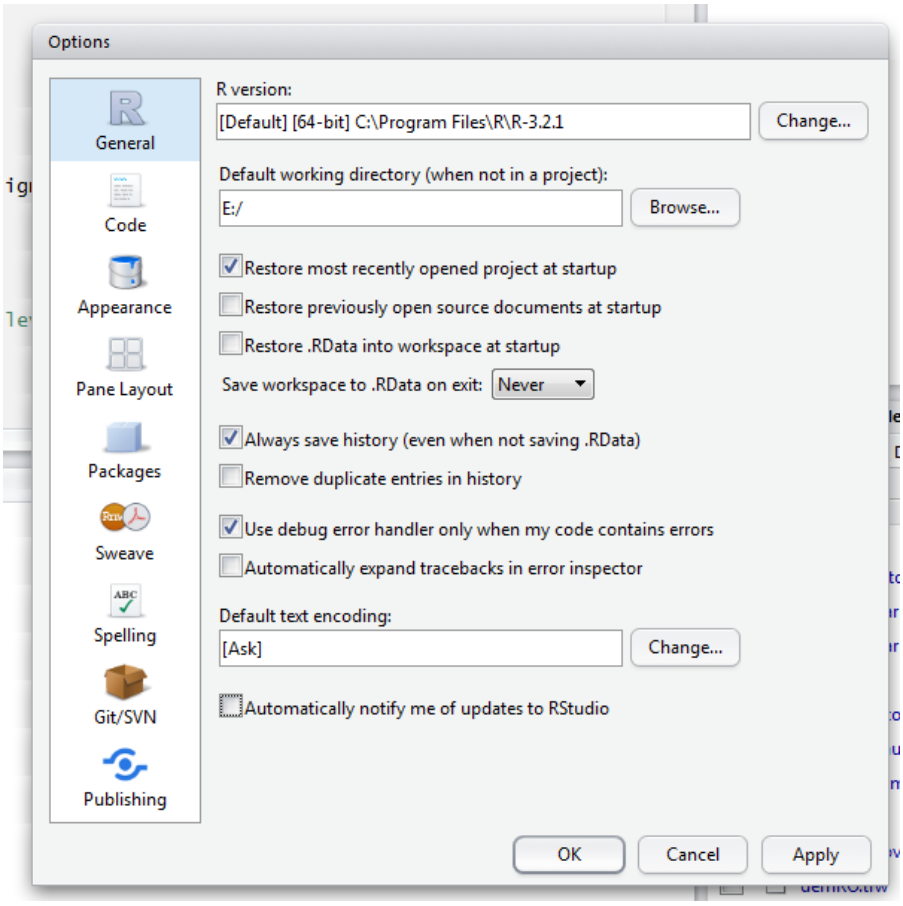
This is how your R Studio screen will be arranged after you set up your R Studio environment



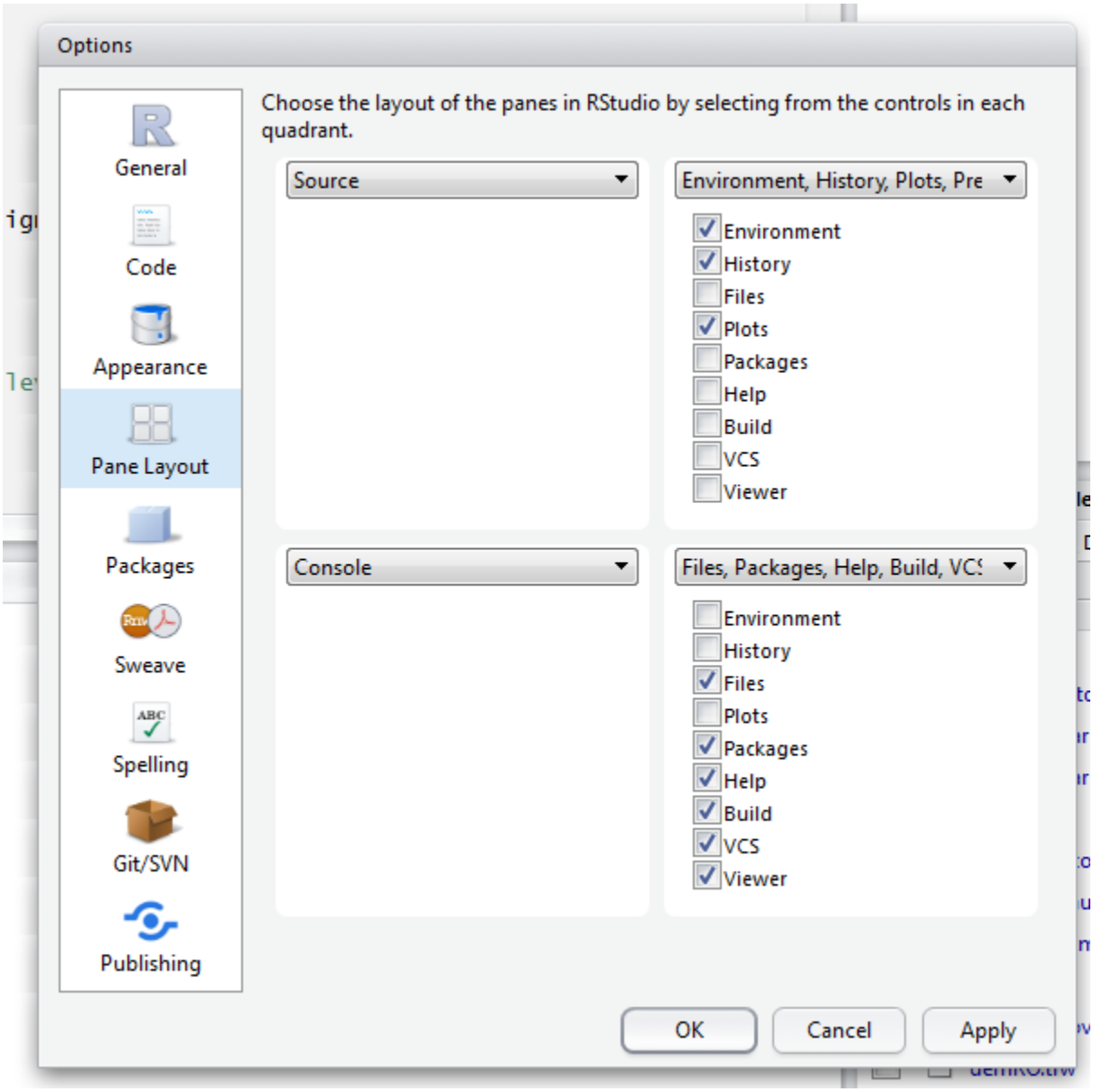
To set up your screen like the above screenshot, click Tools- Global Options



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 – Example File Structure

```
Directory Structure
¦--Local disk
¦--Example of local disk folder structure
¦   ¦--E:Workspace/Projects/MLRA Comparison CONUS
                                                Set this as working directory in R Studio
¦       ¦--MLRA-comparison/  This folder is automatically created when you get the report
¦           ¦--¦--config.R                The script from GitHub will put this file here
¦           ¦          This file tells the report where to look for the pre-sampled data
¦           ¦--¦--report.Rmd              The script from GitHub will put this file here
¦           ¦                          This is the report you will “knit” to run the report
¦           ¦(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
¦           ¦--report_files/              Folder automatically created by the report.Rmd
¦           ¦         .png files of the figures in the html doc are placed in this folder
```

Appendix 3 – Example html report output file

MLRA: 17, 15, 18, 22A, 22B, 1

report version 0.2
2017-04-18

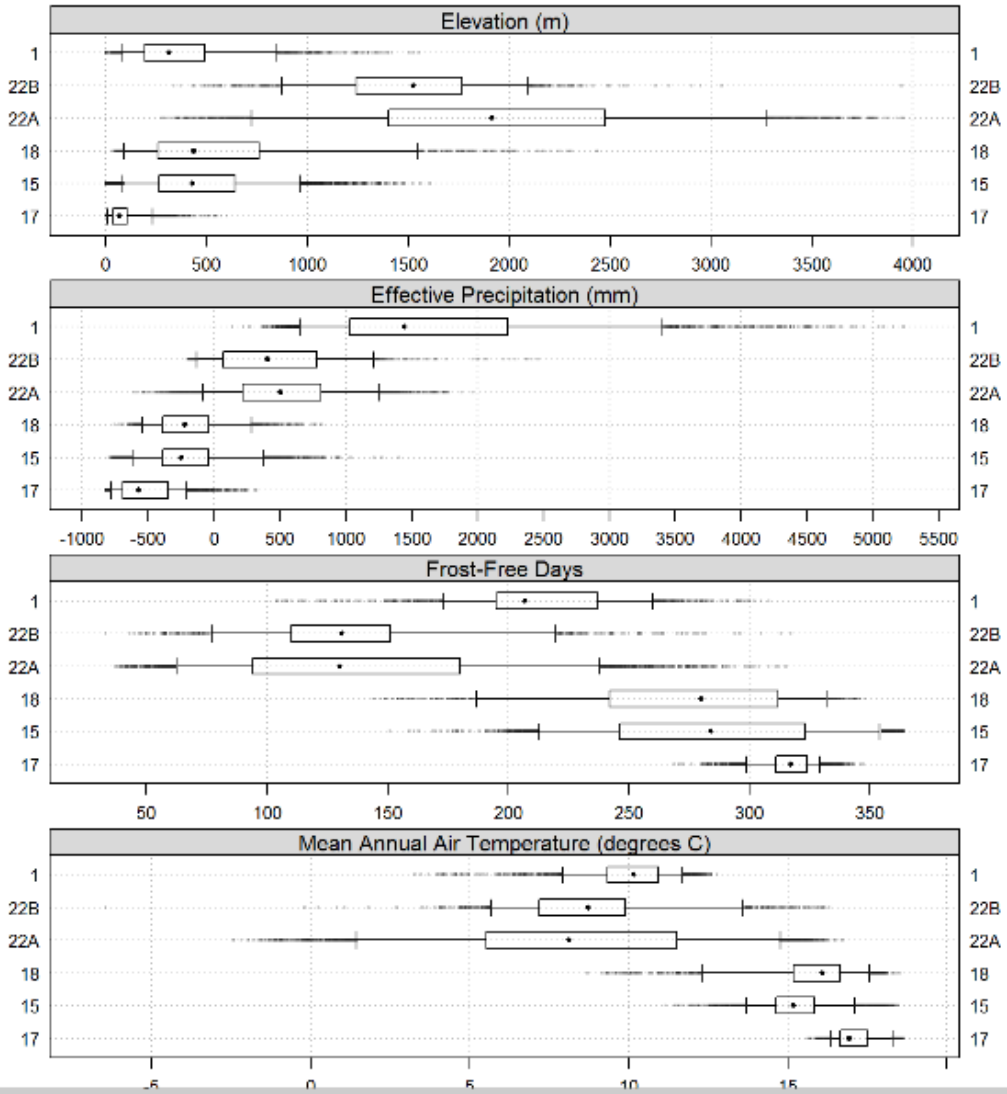
This report is designed to provide statistical summaries of the environmental properties for one or more MLRA Summaries are based on raster data extracted from fixed-density sampling of map unit polygons.

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.

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.

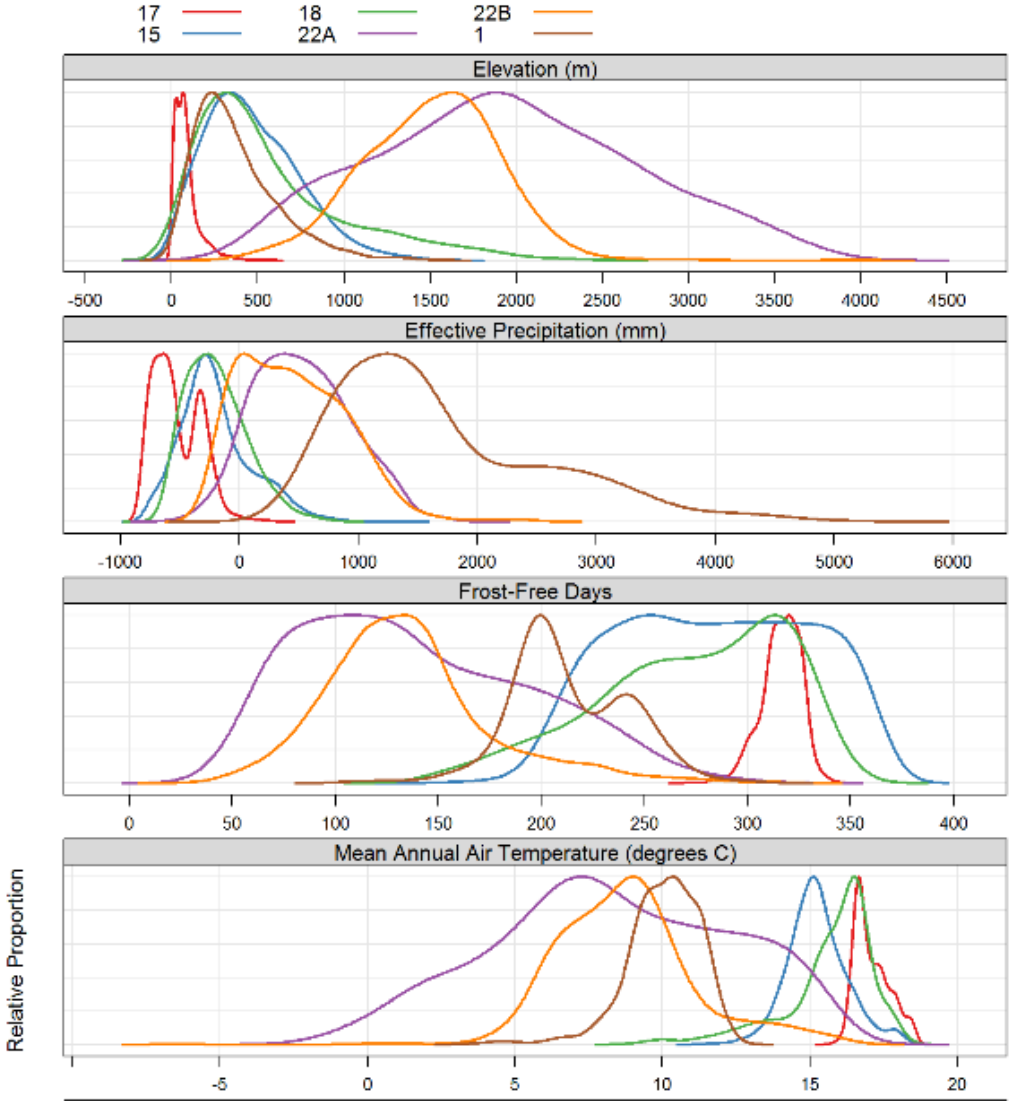


Density Plots

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. Each data set (ID / variable) are rescaled to {0,1} so that the y-axis can be interpreted as the "relative proportion of samples".

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"



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

MLRA	Elevation (m)	Effective Precipitation (mm)	Frost-Free Days	Mean Annual Air Temperature (degrees C)	Mean Annual Precipitation (mm)	Growing Degree Days (degrees C)	Fraction of Annual PPT as Rain
17	68	-572.26	317	16.94	319	2737	99
15	429	-248.00	284	15.19	525	2431	99
18	436	-223.40	280	16.10	600	2563	99
22A	1913	503.97	130	8.11	1073	1280	77
22B	1526	402.91	131	8.74	969	1361	83
1	314	1447.20	207	10.16	2125	1169	96

Elevation (m)

MLRA	Q5	Q10	Q25	Q50	Q75	Q90	Q95
17	9.00	15.0	33	68	105.00	176.0	231.0
15	85.05	142.0	265	429	641.00	839.0	964.0
18	91.00	138.0	259	436	760.00	1254.4	1548.2
22A	723.55	912.1	1402	1913	2475.25	3006.9	3279.0
22B	872.80	1016.8	1241	1526	1768.00	1965.8	2096.8
1	84.00	115.0	194	314	491.00	704.0	848.6

Effective Precipitation (mm)

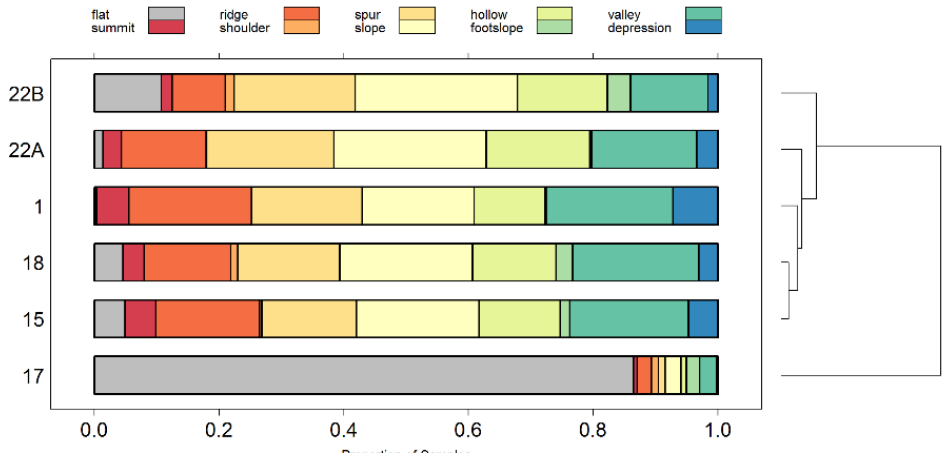
MLRA	Q5	Q10	Q25	Q50	Q75	Q90	Q95
17	-781.41	-761.27	-690.34	-572.26	-347.98	-267.31	-205.48
15	-620.91	-549.01	-396.52	-248.00	-44.47	255.73	378.47
18	-548.05	-495.68	-392.04	-223.40	-36.85	153.54	283.39
22A	-85.90	29.78	217.71	503.97	811.09	1098.16	1254.87
22B	-132.34	-82.54	72.04	402.91	779.46	1044.87	1207.39
1	651.15	784.40	1027.07	1447.20	2229.45	3033.70	3402.45

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.

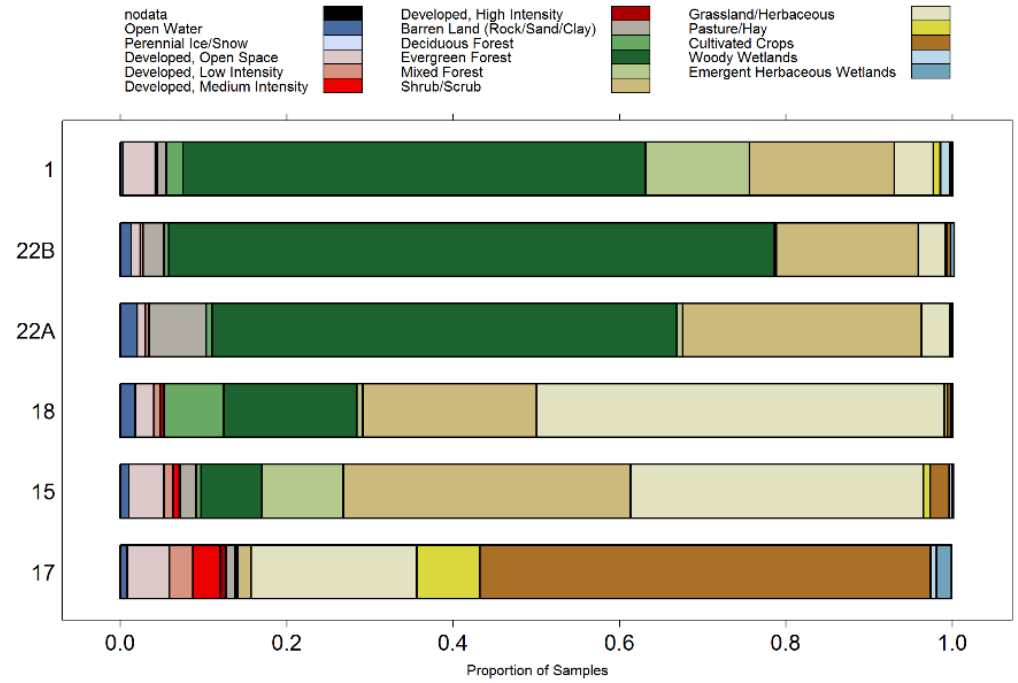
Suggested usage:

- 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.



Landcover Summary

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



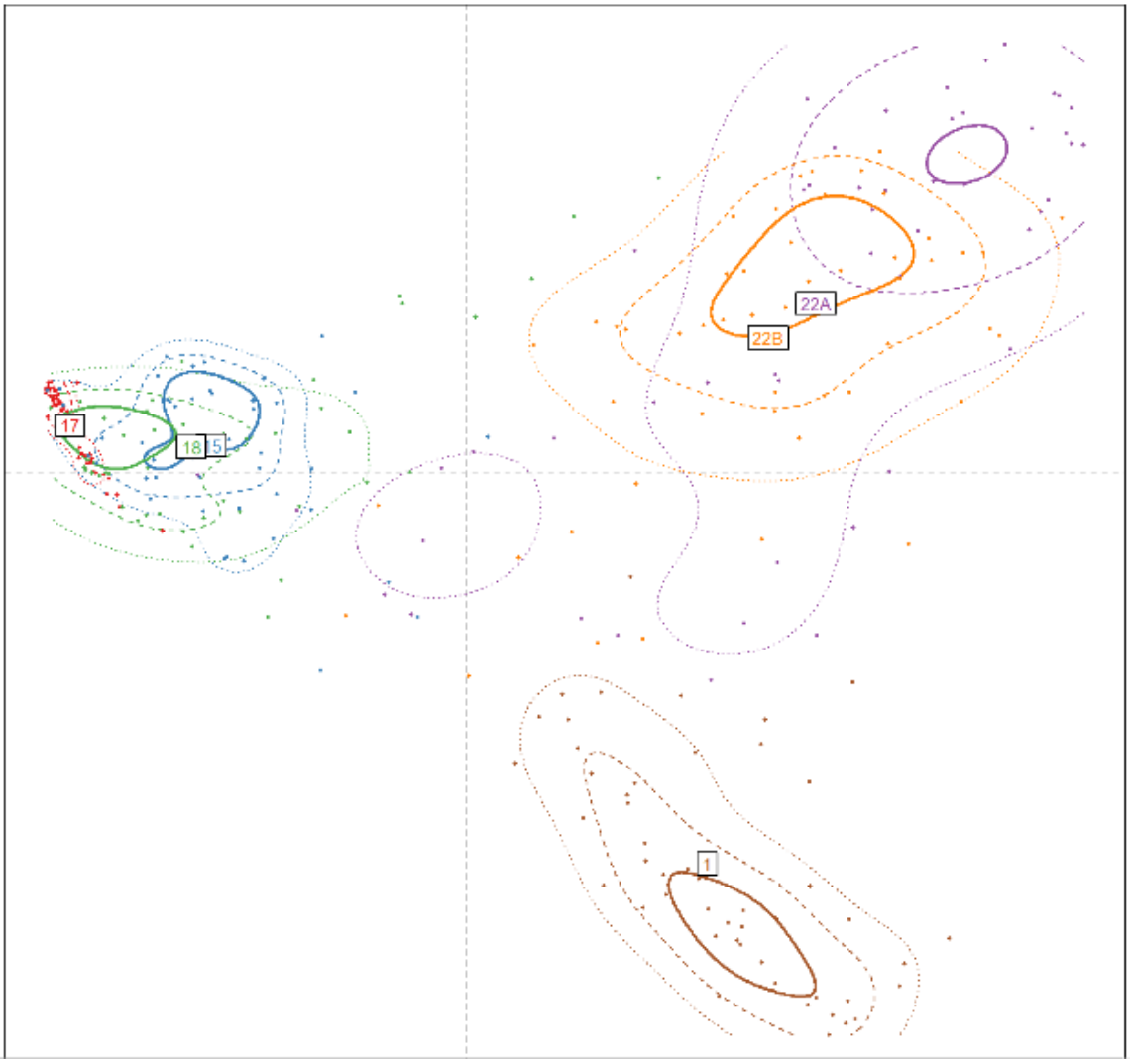
Multivariate Summary

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

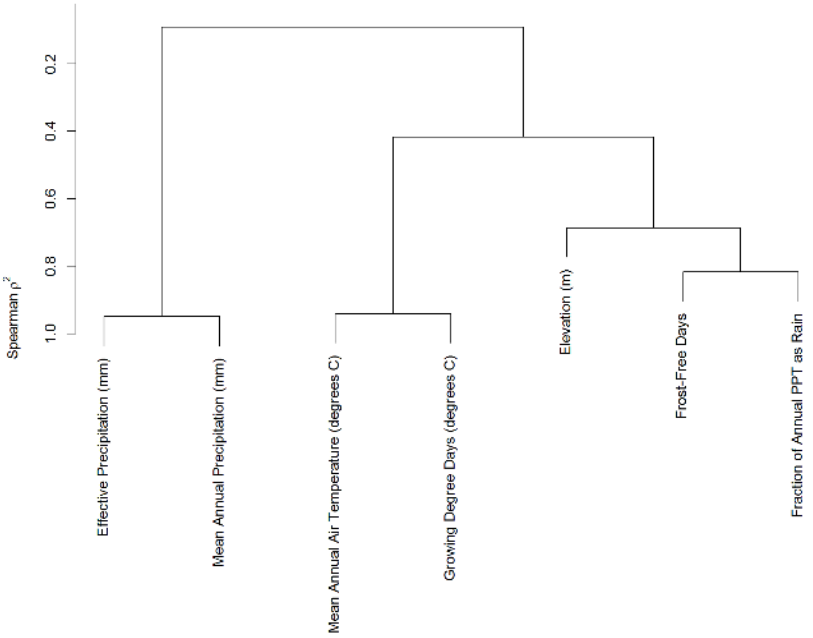


Raster Data Correlation

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.