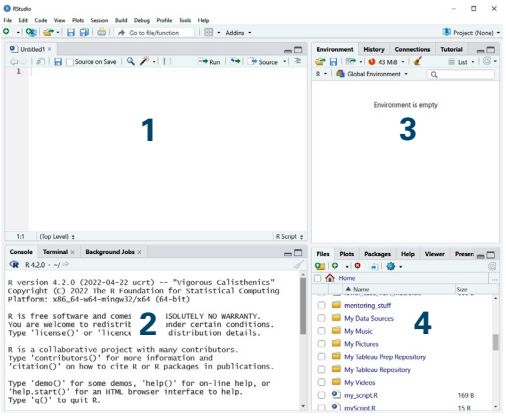
Spatial Data Science with R: Overview of the RStudio Interface

RStudio 2022.07.1

Welcome to the Data Science Task Sheet Series. This series supplements the Iowa State University Extension and Outreach Geospatial Technology Training Program's workshops and short courses by providing quick and easy instructions for performing a variety of mapping, data science, analysis, and visualization tasks.

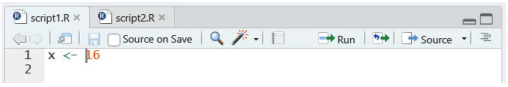
RStudio (recently re-branded as Posit) is an integrated development environment (IDE) for programming in both the R and Python programming languages. RStudio combines a number of different tools into a single application: source code editor, debugger, versioning control, environment and package managers, and more. This task sheet will give an overview of the user interface and some key features of RStudio. Installing R and RStudio are covered in GISTP 0021 - Getting Started with R and R Studio. Although RStudio works with Python, this series is focused on the R programming language.



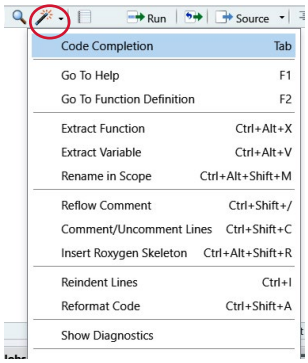
The first time RStudio is launched after installation, the application window will be divided into three panes; when editing code, a fourth pane will appear. Each pane can contain multiple tabs providing extra functionality. Pane locations and included tabs are fully customizable.

1. Source Pane

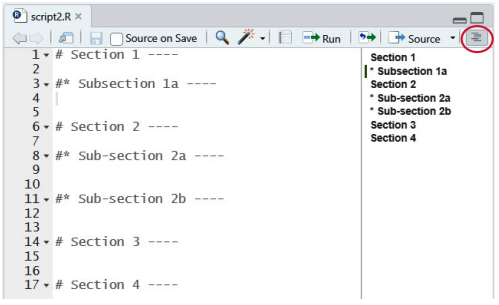
The **Source** pane is where you will write and edit most of your R code. The buttons below the tabs are tools that give you additional functionality when working in the **Source** pane. If you have multiple files open, the arrow buttons allow you to toggle between tabs in the **Source** pane. The **Show in New Window** button will pop the active tab to a free-floating window. The **Save** button will save changes made to the active tab. Checking the **Source on Save** box Source on Save will run the script in its entirety and add any functions in the script to the global environment each time the script is saved. The **Find/Replace** button provides text search and replacement.



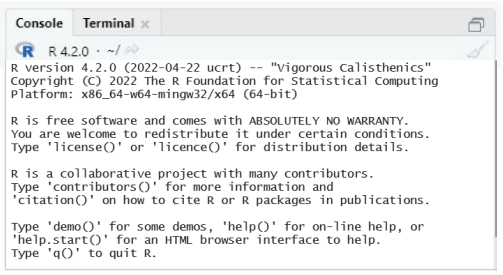
The **Code Tools** button provides quick access to a variety of actions useful during script development. The **Compile Report** button will render an R script into an HTML document containing code, comments, and output. The **Run** button Run will execute either the line of code at the cursor location or a block of highlighted code. The **Re-run code** button will re-execute the last command, regardless of cursor position or highlighted code.



The **Source** button Source allows you to add your code to the global environment; you can "source" a script and all of its functions will be usable in a different script during the same R session. The **Show Document Outline** button will open a panel next to the tab showing a document outline, provided you have marked-up your code using the outline format. This is useful when you are working with long, complex scripts. *Note: To create outline sections in a script, place 4 hypens or hash marks at the end of a comment. Add an asterisk and space after the leading hash mark to create a subsection.*

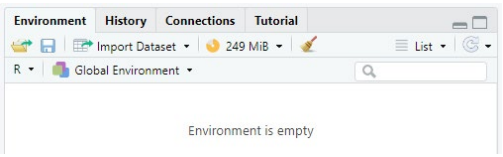


2. Console Pane

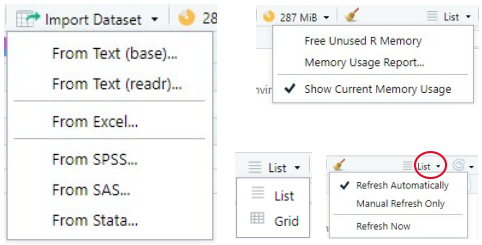


The **Console** pane is typically located in the bottom left position and contains two default tabs, **Console** and **Terminal**. The **Console** tab offers a command-line interface to the R environment. Outputs of scripts run from the **Source** pane will also display here. The **Terminal** tab provides command line access to your computer through a shell of your choice: bash, Windows command prompt or PowerShell, or a custom shell.

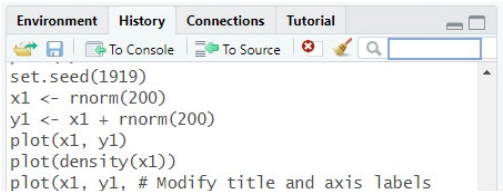
3. Environment Pane



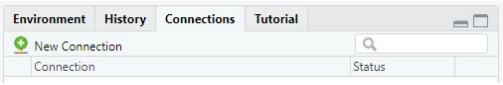
The **Environment** pane is typically located in the upper right quadrant and contains four default tabs: **Environment**, **History**, **Connections**, and **Tutorial**. The **Environment** tab shows variables and data objects loaded in the current workspace. Objects in the workspace, such as variables and data frames, can be saved from one session and loaded in a later session. The **Import Dataset** button allows you to import data into R from CSV files, Microsoft Excel, and other statistics applications. The **Memory** button provides a summary on memory usage. The **Clear** button allows you to "clean up" your environment by removing variables and data from the current workspace. The **List** button opens a drop-down menu that allows you to view items in your environment as a list or a grid. The **Refresh** button allows you to select between automatic and manual refresh of objects in your environment.



The **History** tab contains a list every command issued in your R environment. You can load or save a history file using the load and save buttons. If you click on a line in the **History** tab to select it, you can send that line to the **Console** or to a script by clicking To Console or To Source. The last three buttons in the bar allow you to remove a selected item, clear all items, or search for items in the **History** tab.

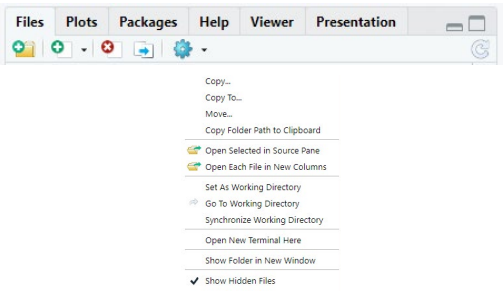


The **Connections** tab allows you to connect to external data sources, typically a database or CSV file. The type of data sources available to you will depend on the packages installed in your R environment. *Note: Creating connections is outside of the scope of this task sheet.*



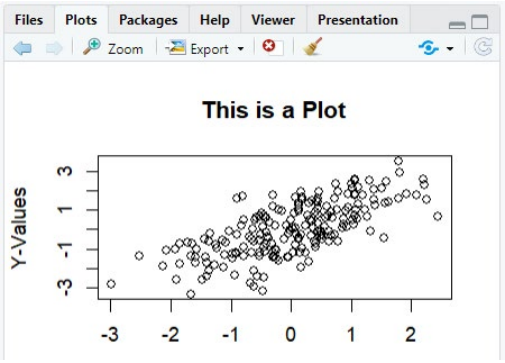
The **Tutorial** tab allows you to run interactive tutorials from within RStudio. Beginner tutorials are provided through the `learnr` package. Following some of these built-in tutorials may provide a good introduction to programming in R.

4. Files / Outputs Pane

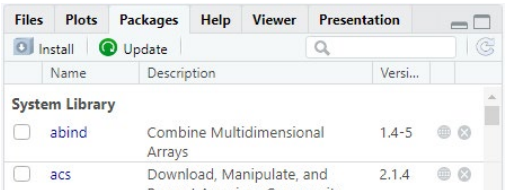


The **Files** pane (sometimes called the **Outputs** pane) is typically located in the bottom right corner. The **Files** tab offers simple file management capabilities. From this tab, you can create new folders New Folder, create a variety of new file types New Blank File, and delete Delete or rename Rename existing files. The list of available commands is expanded under the **More** dropdown menu More.

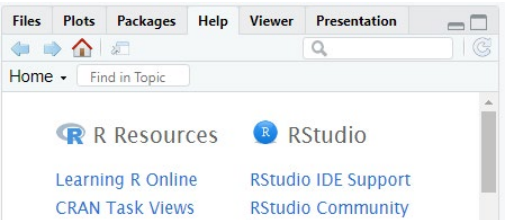
The **Plots** tab will display visualizations created in R. You can switch between plots using the blue arrows. The **zoom** button will enlarge your plot in a new window. The **export** button will let you save the output in a variety of formats and sizes. The two clear image buttons will remove a single plot from the **Plots** tab or "clean up" all the plots, respectively. The **Publish** button allows you to share your work on RStudio Connect or RPubs. *Note: Setting up and linking accounts from those services is outside the scope of this task sheet.*



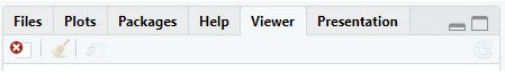
The **Packages** tab provides a list of packages available in your R environment; packages loaded and available for use will have a check mark in the box next to their names. The **Install** and **Update** buttons allow you to install new packages or update existing packages to their latest version. See GISTP 0023 - Installing Packages in RStudio to learn more.



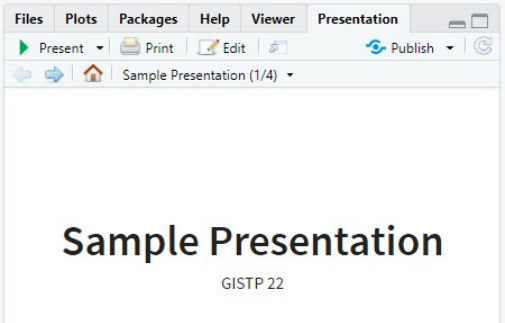
The **Help** tab displays package documention. You can search for topics using the search bar or type a question mark followed by a function name in the **Console** to bring up the documentation or vignette page, e.g., `?mean`, `?class`, `?str`.



The **Viewer** tab will display HTML output produced by a script. Like other tabs, you can remove one or more items from the **Viewer** tab using and show the content in a new window.

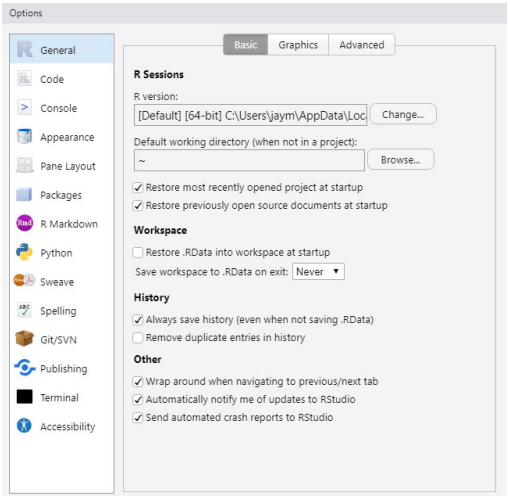


The **Presentation** tab will display rendered presentations made with packages such as `beamer`, `ioslides`, and `reveal.js`. The **Present** button will return the presentation back to the first slide. The **Print** button will open the presentation in a browser, allowing you to print slides. The **Edit** button will open the presentation script in the **Source** pane. The **pop-out** button will display the presentation in a new browser window. You can publish your finished presentation to RStudio Connect or RPubs using the **Publish** button. The second row of buttons allow you to navigate between slides, return to the first slide, or select a specific slide to view using a drop-down menu Sample Presentation (1/4).



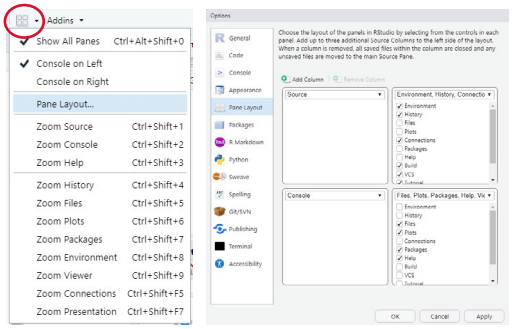
5. Other RStudio Features

As you become more proficient with RStudio, you may want to customize how the program works and looks. The options window for RStudio can be found under **Tools -> Global Options...**. This window provides access to all of the settings and options controlling how RStudio looks and operates.

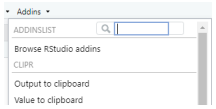


Some useful sections include **Code**, where you can customize RStudio keyboard shortcuts to mach other popular editors, and adjust code auto-completion settings; **Appearance**, where you can choose between a variety of included user-interface themes; **Pane Layout**, where you can customize where individual tabs are located; and **Git/SVN** where you can enable version control for your RStudio Projects.

Returning to the toolbar beneath the main application window, the **Workspace Panes** button , allows you to quickly switch between the default multi-pane view or filling the application window with the contents of a single tab. The **Pane Layout...** option will open up the **Global Options** window and allow you to fully customize the contents of each pane.

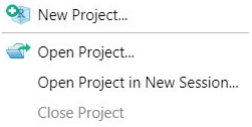


The **Addins** drop-down Addins, next to the **Panes** button, provides access to any addins currently installed in RStudio An addin is a special type of package that provides extended functionality to RStudio. Different addins allow you to build advanced visualizations then output the code (esquisse), paste the contents of the computer clipboard to a script as comment lines (pasteAsComment), or search for and install addins (addinslist).

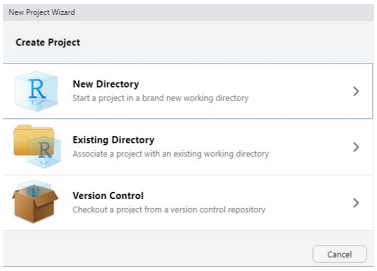


6. RStudio Projects

An RStudio project is a folder containing a project file (.Rproj) and associated scripts, data files, helper files, and sub-directories. The **Project** button Project: (None) (located above the **Environment** pane) allows you to create a new project, close the current project or open an existing project from a list of recent projects kept by RStudio. You can also create a new project by clicking the button on th application toolbar.

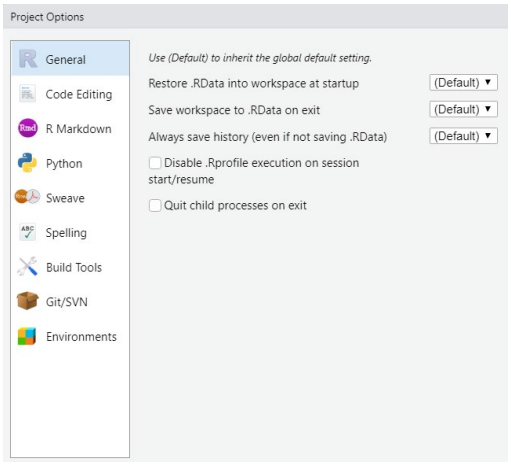


When creating a new project, you can create a new folder, choose an existing folder, or retrieve code from a version control system. The project folder will be considered the working directory (sometimes called the root or top-level directory) for your project. In addition to your own scripts and data, an RStudio project folder contains hidden directories containing temporary files and project settings.



When working with an RStudio proejct, it is good practice to keep your data and outputs within the project folder or sub-directories. By doing so, all file paths can be written relative to the project folder. Additionally, by keeping all related assets inside the project directory structure, you can share your projects between computers and colleagues.

Projects can have their own custom settings, overriding the default global options. If you have a project open in RStudio, **Tools -> Project Options...** will bring up the settings panel for your project. You may notice the similarities between this option screen and the global options screen.



Contact: Jay Maxwell, Data Analyst, and Professor Christopher J. Seeger, PLA, GISP can be reached at geospatial@iastate.edu. Additional task sheets and information about the Geospatial Technology and Spatial Data Science Programs are available at www.extension.iastate.edu/communities/gis.