

Statistical Computing with R: Masters in Data Sciences 503, S29 First Batch, SMS, TU, 2021

Shital Bhandary

Associate Professor

Statistics/Bio-statistics, Demography and Public Health Informatics

Patan Academy of Health Sciences, Lalitpur, Nepal

Faculty, Data Analysis and Decision Modeling, MBA, Pokhara University, Nepal

Faculty, FAIMER Fellowship in Health Professions Education, India/USA.

Review Preview:

- R script
- R markdown
- Project in R studio
- R notebook
- Working with Project
- Version control of Project

Chapter 8 (Workflow:projects) of the course text book: R for Data Science

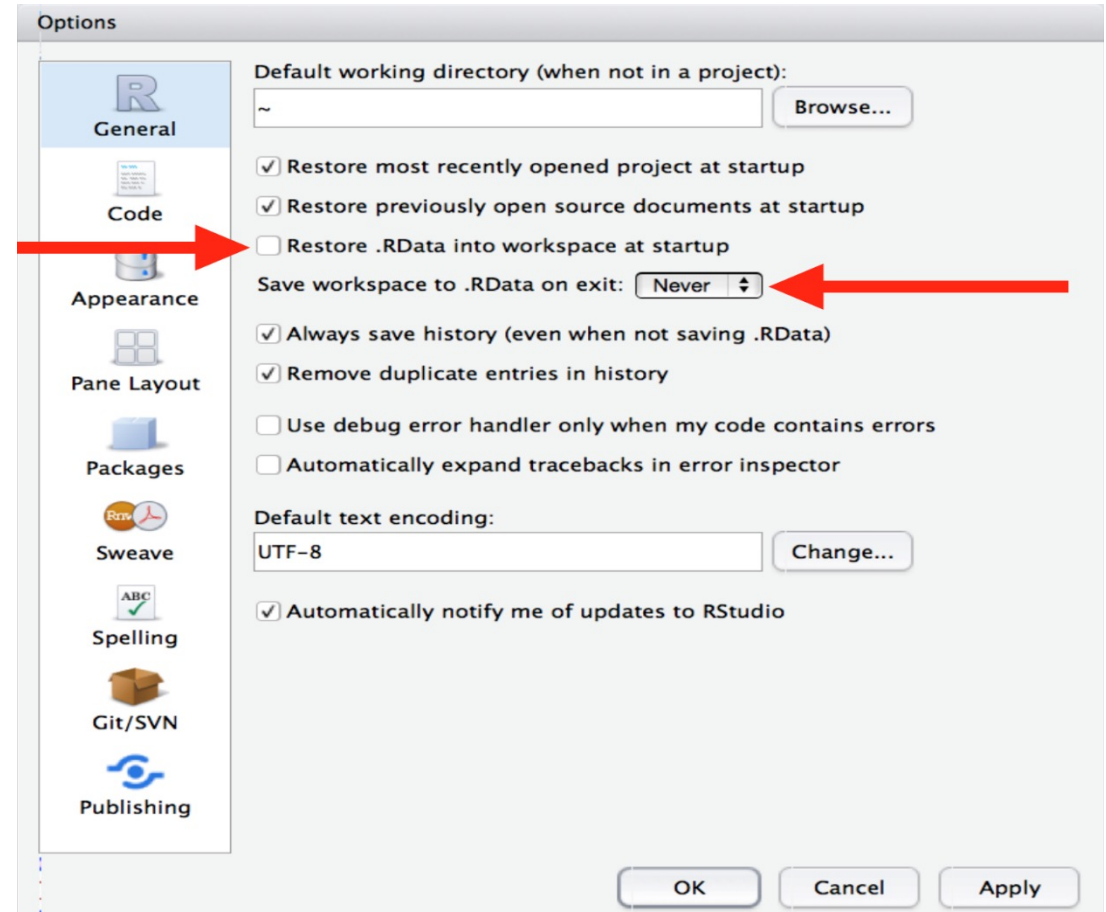
- One day you will need to quit R, **go do something else and return to your analysis** the next day.
- One day you will be **working on multiple analyses simultaneously that all use R** and you want to keep them separate.
- One day you will need to **bring data from the outside world into R and send numerical results and figures from R back out** into the world.
- To handle these real life situations, you need to make two decisions:
 - What about your analysis is “real”, i.e. what will you save as your lasting record of what happened?
 - Where does your analysis “live”?

Chapter 8 (Workflow:projects) of the course text book: R for Data Science

- As a **beginning R user**, it's **OK** to consider your environment (i.e. the objects listed in the environment pane) “real”.
- However, **in the long run**, you'll be much better off **if you consider your R scripts as “real”**.
- With your R scripts (and your data files), **you can recreate the environment**.
- It's much harder to recreate your R scripts from your environment!
- You'll either have to **retype a lot of code from memory** (making mistakes all the way) or **you'll have to carefully mine your R history**.

Chapter 8 (Workflow:projects) of the course text book: R for Data Science

- To foster this behaviour, it is highly recommended to instruct RStudio not to preserve your workspace between sessions.
- It can be done using global options and unchecking the “restore .Rdata into workspace at startup” and setting “never” save workspace to .Rdata on exit as shown in the screenshot:



Chapter 8 (Workflow:projects) of the course text book: R for Data Science

- This will **cause you some short-term pain**, because now when you restart RStudio it will not remember the results of the code that you ran last time.
- But this **short-term pain will save you long-term agony** because it forces you to capture all important interactions in your code.
- **There's nothing worse than discovering three months after the fact that you've only stored the results of an important calculation in your workspace, not the calculation itself in your code.**
- Therefore we need "Project" and R Studio provides this feature to us!

R Studio Project:

<https://support.rstudio.com/hc/en-us/articles/200526207-Using-RStudio-Projects>

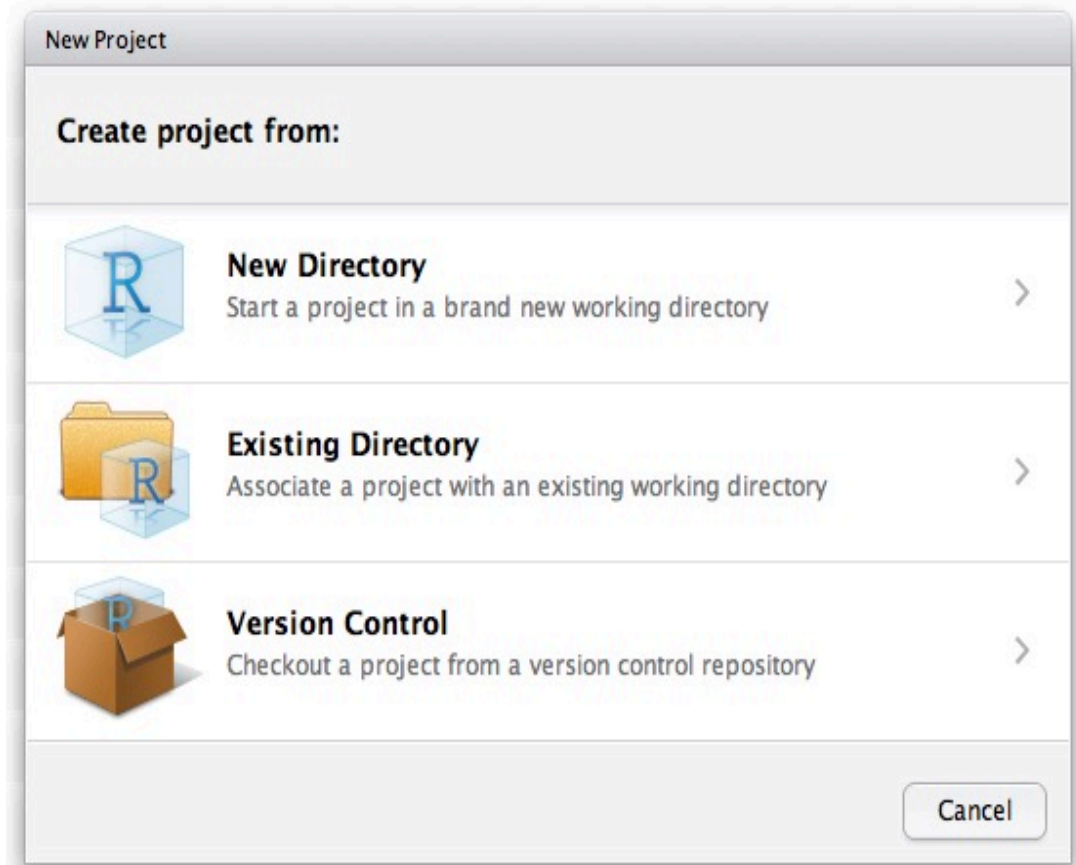
- R experts keep all the files **associated with a project together** — input data, R scripts, analytical results, figures.
- This is such a wise and common practice that RStudio has built-in support for this via **projects**.
- RStudio projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents.

#Creating Projects

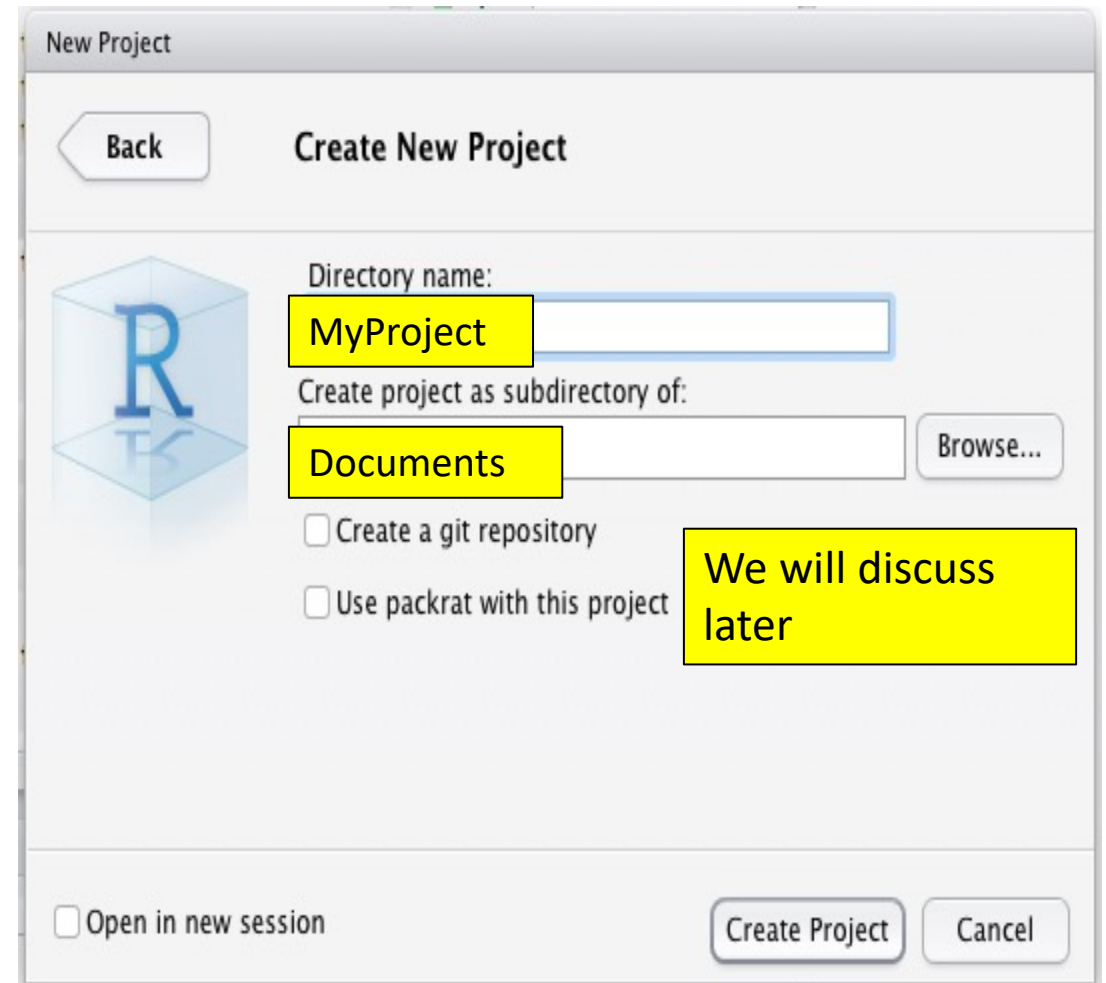
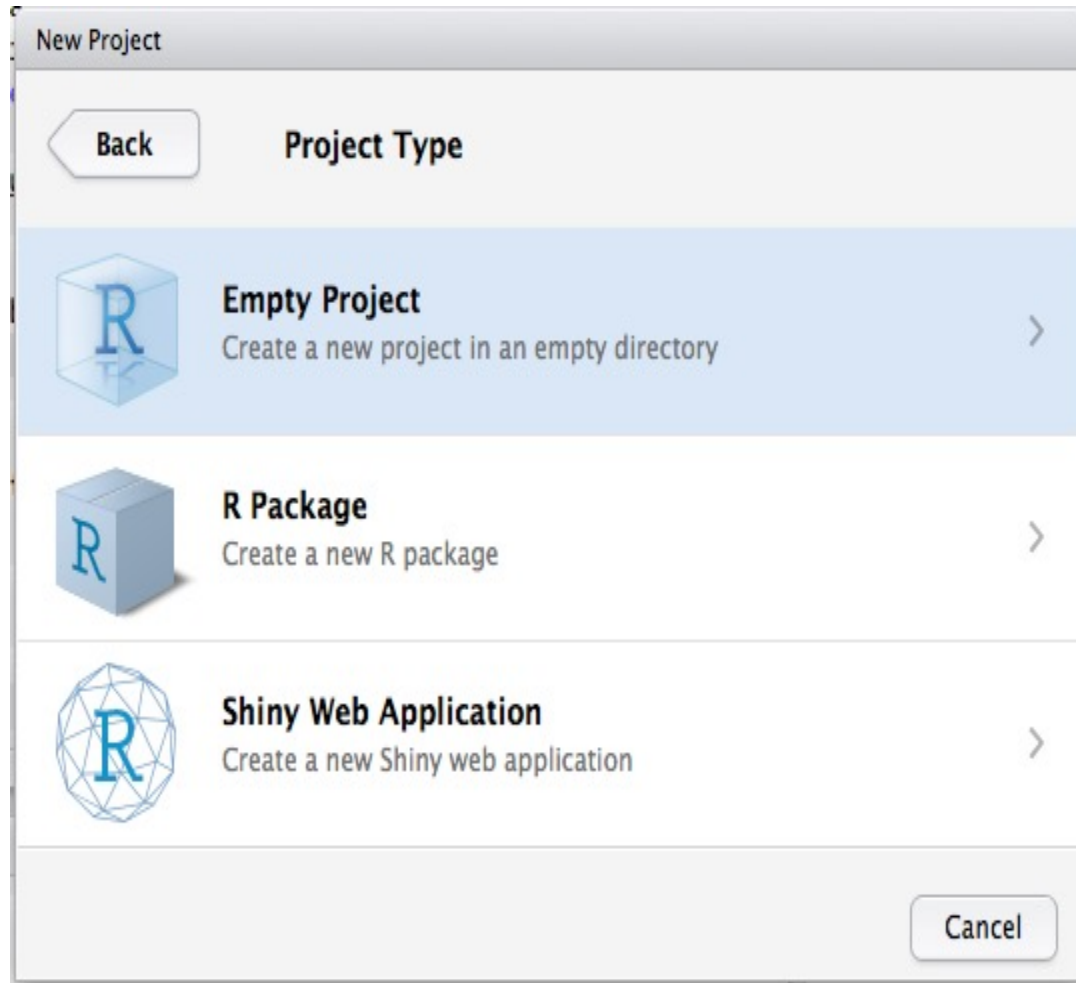
- RStudio projects are associated with R working directories.
- You can create an RStudio project:
 - In a brand new directory
 - In an existing directory where you already have R code and data
 - By cloning a version control (git or subversion) repository

R Studio project: Let's do it

- To create a new project in the RStudio IDE, use the **Create Project** command
- It is available on the Projects menu and on the global toolbar:



Working with Project: Let's do it



Working with Project:

- When a new project is created RStudio:
- Creates a project file (**with an .Rproj extension e.g. MyProject.Rproj**) within the project directory.
- This file contains various project options (discussed below) and can also be used as a shortcut for opening the project directly from the file system.
- Creates a **hidden directory** (named **.Rproj.user**) where project-specific temporary files (e.g. auto-saved source documents, window-state, etc.) are stored. This directory is also automatically added to **.Rbuildignore, .gitignore**, etc. if required.
- Loads the project into RStudio and display its name in the Projects toolbar (**which is located on the far right side of the main toolbar**)

Working with Project:

- Call your project “MyProject” and think carefully about which *subdirectory* you put the project in.
- If you don’t store it somewhere sensible, it will be hard to find it in the future!
- Once this process is complete, you’ll get a new RStudio project.
- Check that the “home” directory of your project is the current working directory with:
 - `getwd()`
 - "C:/Users/Dell/Documents/MyProject"

Working with Project:

- Now **enter the following commands** in the script editor, and **save the file**, calling it “diamonds.R” **inside/in the project directory**.
- Quit Rstudio & Inspect the folder associated with your project — notice the MyProject.**Rproj** file!
- Next, **run the complete script** which will save a PDF and CSV file into **your project directory**.

```
#diamonds.R
```

- library(tidyverse)
- ggplot(diamonds, aes(carat, price)) +
- geom_hex()
- ggsave("diamonds.pdf")
- write_csv(diamonds, "diamonds.csv")

I got this in Dell Optiplex 3010: Windows 10

MyProject - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

diamonds.R x

Package tidyverse required but is not installed. [Install](#) [Don't Show Again](#)

```
1 #diamonds.R
2
3 library(tidyverse)
4 ggplot(diamonds, aes(carat, price)) +
5   geom_hex()
6 ggsave("diamonds.pdf")
7 write_csv(diamonds, "diamonds.csv")
8
```

tidyverse was installed a priori but it is asking to install for this project!
When I clicked Install, it says “installing diamond.R” dependencies!

Environment History Connections Tutorial

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > MyProject

	Name	Size	Modified
	..		
	MyProject.Rproj	218 B	Jan 10, 2022, 11:29 AM
	diamonds.R	150 B	Jan 10, 2022, 11:30 AM

Console Terminal x Jobs x

~/MyProject/

R is free software and comes with absolutely no warranty.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

ggsave asked me to install “hexbin” package as it was required for the geom_hex layer although it was also installed already!

Finally, I got this:

The screenshot displays the RStudio interface for a project named 'MyProject'. The main editor window shows a script named 'diamonds.R' with the following R code:

```
1 #diamonds.R
2
3 library(tidyverse)
4 ggplot(diamonds, aes(carat, price)) +
5   geom_hex()
6 ggsave("diamonds.pdf")
7 write_csv(diamonds, "diamonds.csv")
8
```

The Environment pane on the right shows the 'Global Environment' with the message 'Environment is empty'.

The Files pane at the bottom right shows the project structure:

Name	Size	Modified
..		
MyProject.Rproj	218 B	Jan 10, 2022, 11:29 AM
diamonds.R	150 B	Jan 10, 2022, 11:30 AM
diamonds.csv	2.3 MB	Jan 10, 2022, 11:34 AM
diamonds.pdf	12.5 KB	Jan 10, 2022, 11:44 AM

The Console pane at the bottom shows the output of the script execution:

```
~/MyProject/
Selection: 1
Installing package into 'C:/Users/Dell/Documents/R/win-library/4.0'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/hexbin_1.28.2.zip'
Content type 'application/zip' length 1474974 bytes (1.4 MB)
downloaded 1.4 MB

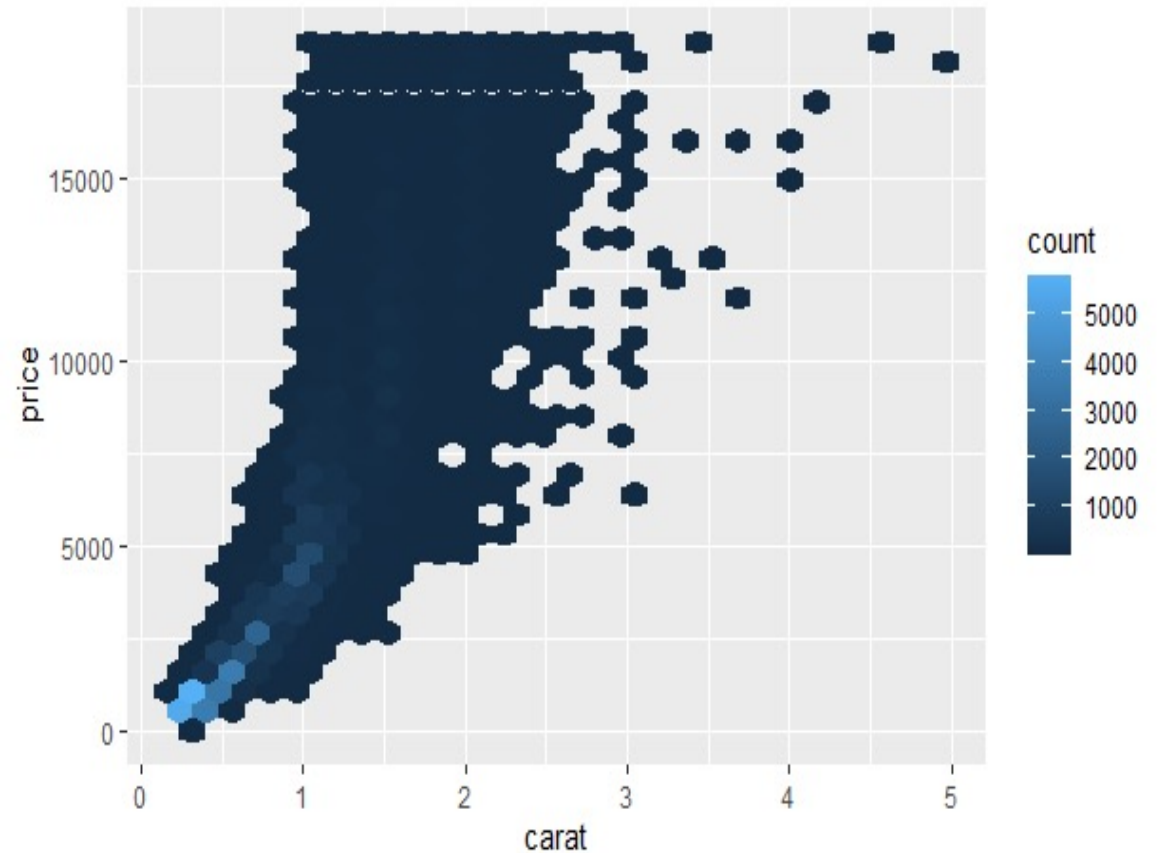
package 'hexbin' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/Dell/AppData/Local/Temp/RtmpkbnKiv/downloaded_packages
> write_csv(diamonds, "diamonds.csv")
> ggsave("diamonds.pdf")
Saving 6.4 x 3.78 in image
>
```

The plot: Saved as PDF and also seen in the “plots” tab

#It was produced due to this code:

- `library(tidyverse)`
- `ggplot(diamonds, aes(carat, price)) +`
- `geom_hex()`



Working with Project:

Opening Projects

- There are several ways to open a project:
- Using the **Open Project** command (also available from both the Projects menu and the Projects toolbar) to browse for and select an existing project file
 - e.g. MyProject.Rproj
- Selecting a project from the list of most recently opened projects (also available from both the Projects menu and toolbar).
- Double-clicking on the project file within the system shell (e.g. Windows Explorer, OSX Finder, etc.).

Working with Project:

- When a project is opened within RStudio the following actions are taken:
- A new R session (process) is started
- The **.Rprofile** file in the project's main directory (if any) **is sourced by R**
- The **.RData** file in the project's main directory **is loaded** (if project options indicate that it should be loaded).
- The **.Rhistory** file in the project's **main directory is loaded** into the RStudio History pane (and used for Console Up/Down arrow command history).
- The **current working directory is set to the project directory.**
- Previously edited source documents are restored into editor tabs
- Other **RStudio settings** (e.g. active tabs, splitter positions, etc.) **are restored to where they were the last time the project was closed.**

Working with Project:

Quitting a Project

- When you are within a project and choose to either Quit, close the project, or open another project the following actions are taken:
- **.RData and/or .Rhistory are written to the project directory** (if current options indicate they should be)
- The list of open source documents is saved (so it can be restored next time the project is opened)
- Other RStudio settings (as described above) are saved.
- The R session is terminated.

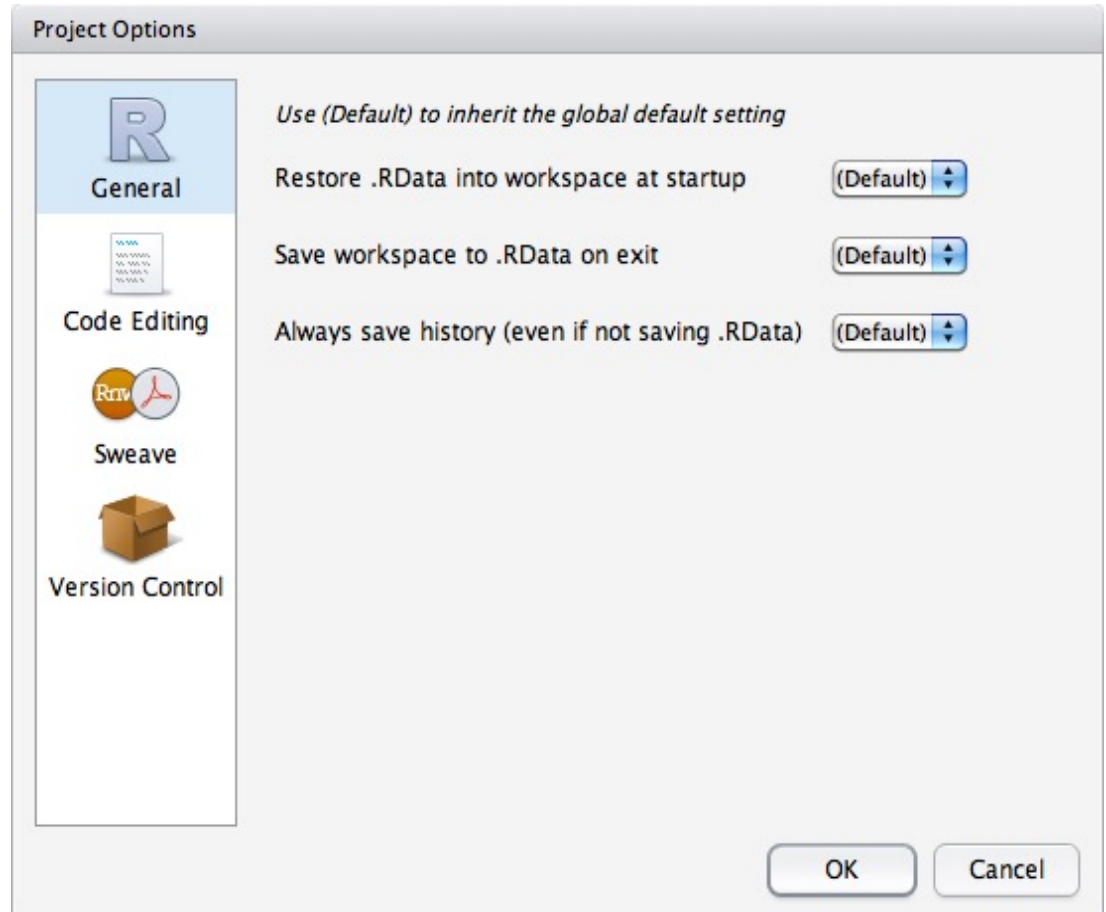
Working with Project:

Working with Multiple Projects at Once

- You can work with more than one RStudio project at a time by simply opening each project in its own instance of RStudio.
- There are two ways to accomplish this:
 - Use the **Open Project in New Window** command located on the Project menu.
 - Opening **multiple project files via the system shell** (i.e. double-clicking on the project file).

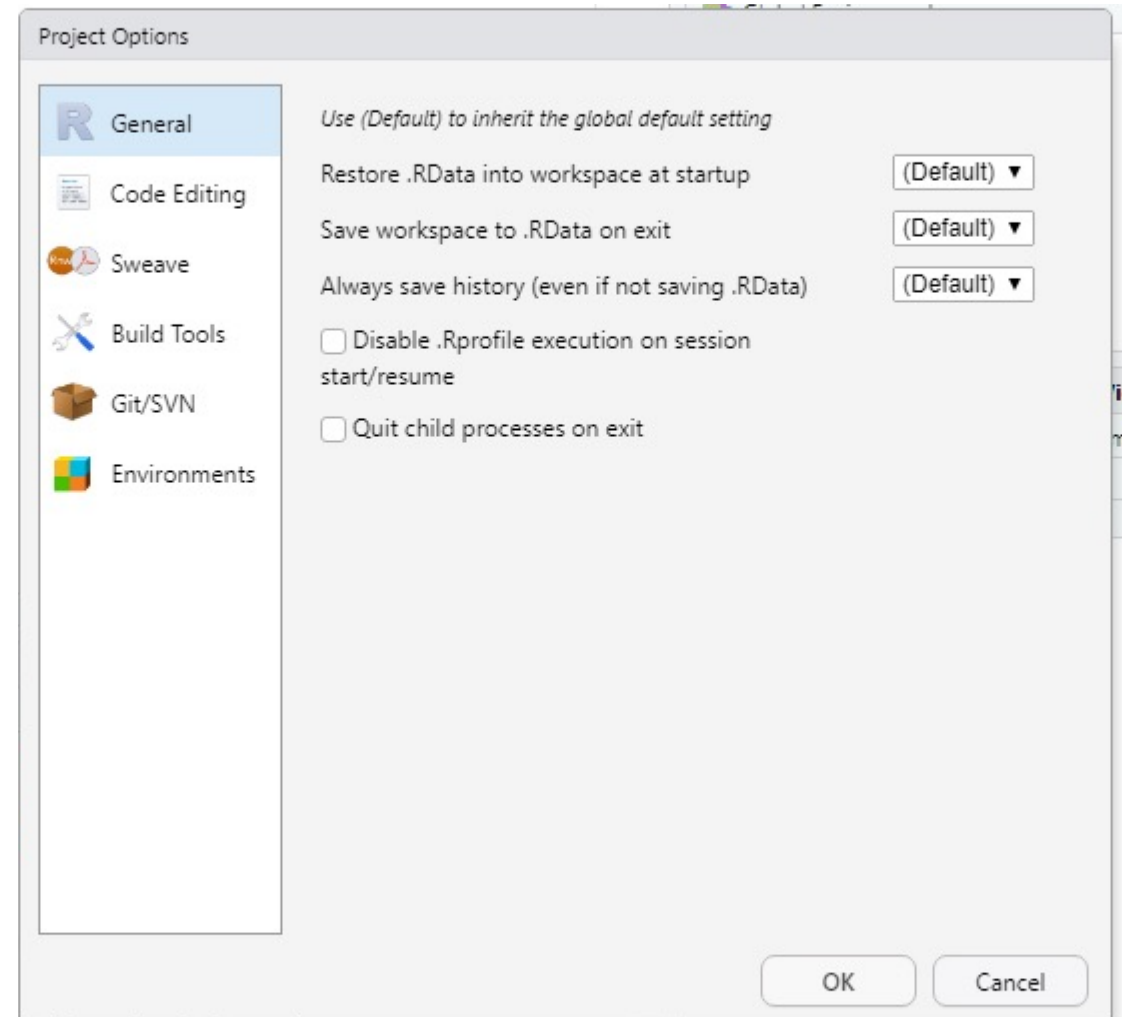
Working with Project:

- Tools → Project Options
- There are several options that can be set on a per-project basis to customize the behavior of RStudio. You can edit these options using the **Project Options** command on the **Project** menu:



You can explore each of the options seen:

- **General** (*Do not change!*)
- Code Editing
- Sweave
- Build Tools
- Git/SVN
- Environments
 - packrat would appear if it was selected while creating the project



Working with Project:

- Example: Code Editing
- **Index R source files** — Determines whether R source files within the project directory are indexed for code navigation (i.e. go to file/function, go to function definition).
- Normally this should remain enabled, however if you have a project directory with thousands of files and are concerned about the overhead of monitoring and indexing them you can disable indexing here.
- **Insert spaces for tab** — Determine whether the tab key inserts multiple spaces rather than a tab character (soft tabs). Configure the number of spaces per soft-tab.
- **Text encoding** — Specify the default text encoding for source files.
- Note that source files which don't match the default encoding can still be opened correctly using the **File : Reopen with Encoding** menu command.

Working with Project:

#Project options: Sweave

PDF Generation

- Weave Rnw (or Sweave) file using: Sweave
- Typeset LaTeX into PDF using: pdfLaTeX

PDF Preview

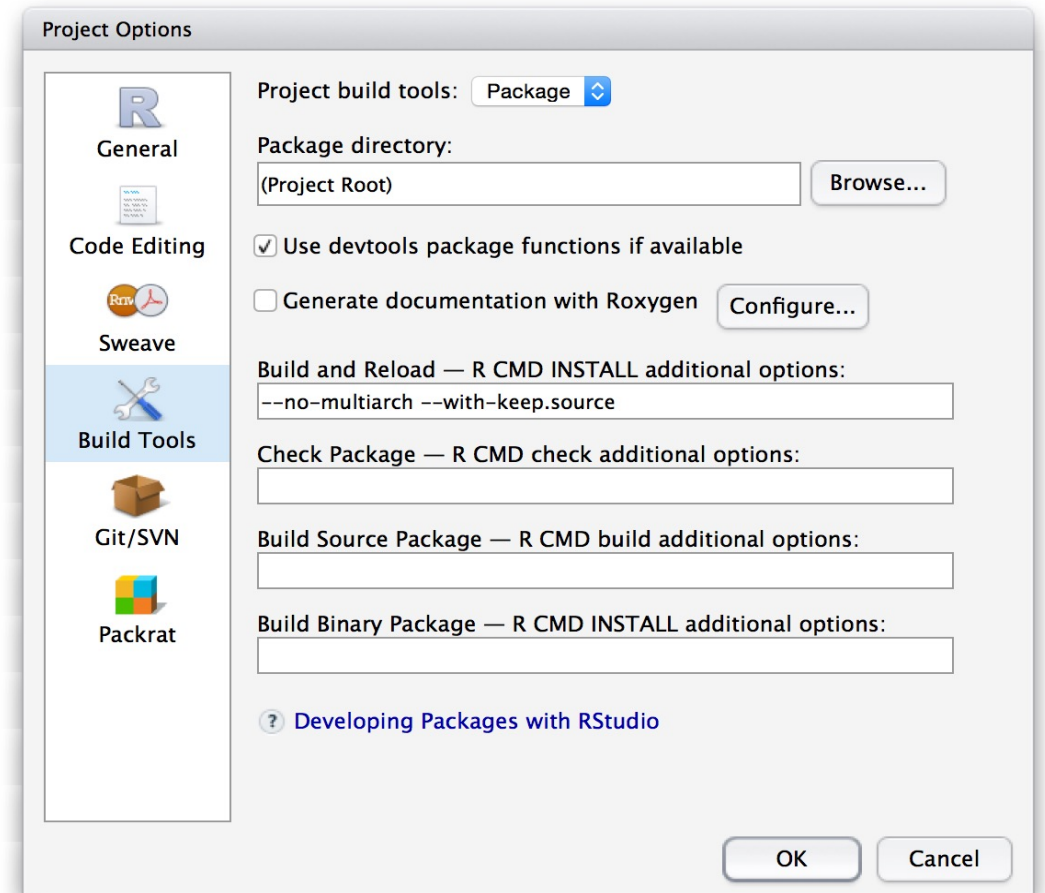
- Compile PDF root document:

- More here:
<https://support.rstudio.com/hc/en-us/articles/200532257?version=1.3.1093&mode=desktop>
- <https://support.rstudio.com/hc/en-us/articles/200532247?version=1.3.1093&mode=desktop>
- <https://support.rstudio.com/hc/en-us/articles/200486298?version=1.3.1093&mode=desktop>

Working with Project:

#Project options: Build Tools

- There are three R package build commands used by the package development tools in the RStudio IDE:
 - R CMD check
 - R CMD build
 - R CMD INSTALL
- It's possible to customize the options passed to each of these commands using **Project Options : Build Tools**



Working with Project:

- **Version control system** - Specify the version control system to use with this project.
- Note that **RStudio automatically detects** the presence of version control for projects by scanning for a .git or .svn directory.
- **Therefore it isn't normally necessary to change this setting.**
- You may want to change the setting for the following reasons:
 - You have both a .git and .svn directory within the project and wish to specify **which version control system RStudio should bind to.**
 - You have no version control setup for the project and you **want to add a local git repository** (equivalent to executing git init from project root directory)
- **Origin** — Read-only display of the remote origin (if any) for the project version control repository

More here: <https://support.rstudio.com/hc/en-us/articles/200532077?version=1.3.1093&mode=desktop>

Version Control with Git and SVN in R Studio:

<https://support.rstudio.com/hc/en-us/articles/200532077?version=1.3.1093&mode=desktop>

- Prior to using RStudio's version control features you will **need to ensure that you have Git and/or Subversion installed on your system.**
- Git: Prior to using Git with RStudio you should install it using the appropriate method for your platform:
 - **Windows & OS X:** <http://git-scm.com/downloads>
 - **Debian/Ubuntu:** `sudo apt-get install git-core`
 - **Same is true for using sub-versions!**
- Once you've installed your preferred Version Control system, you'll need to activate it on your system by following these steps:
 - Go to **Global Options** (from the Tools menu)
 - Click **Git/SVN**
 - Click **Enable version control interface for RStudio projects**
 - If necessary, enter the path for your Git or SVN executable where provided.
 - You can also create or add your RSA key for SSH if necessary.

Working with Project:

- Project options: Environments
- Rstudio uses the “renv” package to give your projects their own privately managed library, making your R code more isolated, portable and reproducible
- Click/check: “Use “renv” with this project” to use this feature!

More here:

<https://rstudio.github.io/renv/>

#How to use:

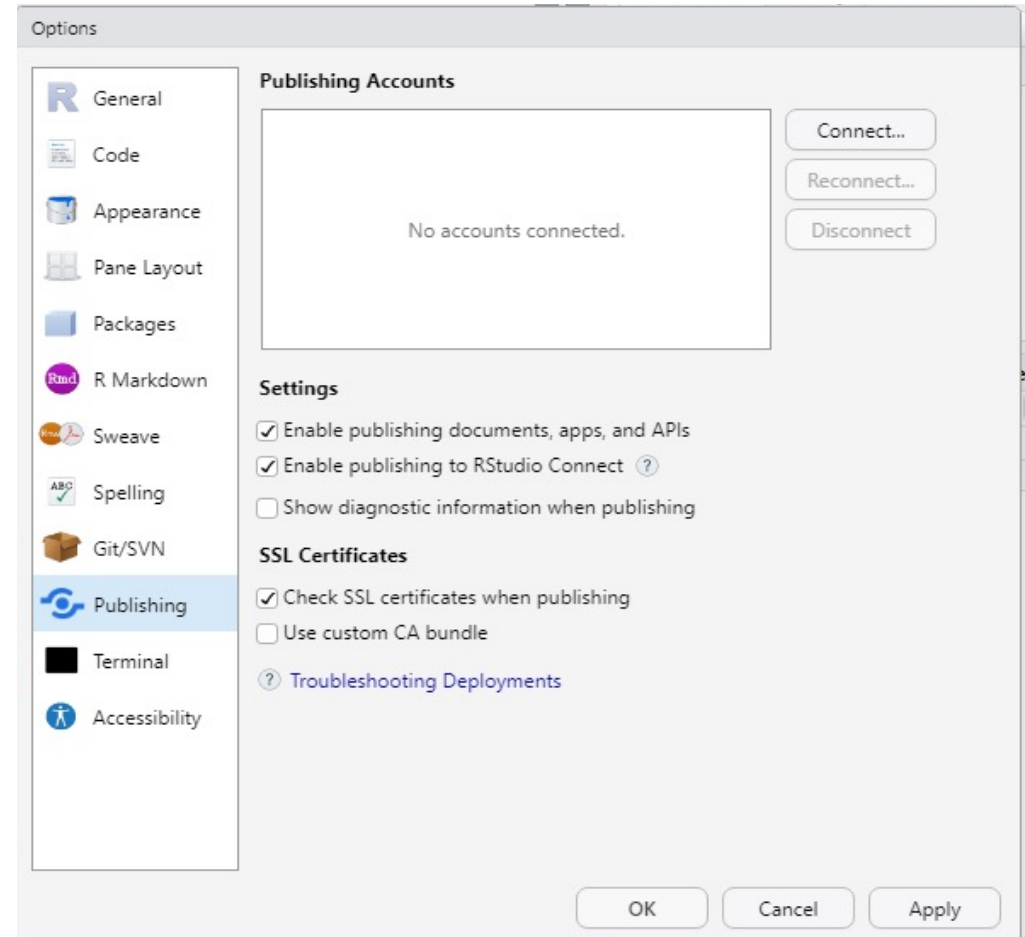
- Use **renv::init()** to initialize renv with a new or existing project.
- This will set up your project with a private library, and also make sure to install all of the packages you’re using into that library.
- The packages used in your project will be recorded into a *lockfile*, called `renv.lock`

Alternatively, we can use “packrat” too (define while creating project): <https://rstudio.github.io/packrat/rstudio.html>

Working with Project:

- Publishing with R studio:
- Markdown/R notebook files using:
- **Rpubs**
- R Studio connect

More here: <https://support.rstudio.com/hc/en-us/articles/228270928-Push-button-publishing-to-RStudio-Connect>



More here: <https://docs.rstudio.com/connect/user/publishing/>

Question/queries?

- Next classes
- R markdown “nitty gritty”
- R notebook

Thank you!

@shitalbhandary