Cluster analysis and its application in geochemistry -- A workshop at Goldschmidt 2020

Shuang Zhang (Carnegie Earth and Planets Laboratory)

Step 1

1.1 Download R **first** from https://cran.r-project.org/

Note: if you already have R installed on your computer, just ignore this step. However, if you are not able to run the code of this workshop, your R might be very old. Consider updating R. This is a link showing you how to do the update (https://www.linkedin.com/pulse/3-methods-update-r-rstudio-windows-mac-woratana-ngarmtrakulchol).

Choose your system, download and install the latest R version (Linux needs to select the right distribution). My R version is 3.6.1. Therefore, if you choose to download and install the latest version, it will be newer than my version but it should work (let us know if it doesn't work). The packages you install afterwards should be compatible with your own R version.

For windows, click base and download the latest version as follows:

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

R-4.0.1 for Windows (32/64 bit)

Download R 4.0.1 for Windows (84 megabytes, 32/64 bit)
Installation and other instructions
New features in this version

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server. You will need a version of md5sum for windows: both graphical and command line versions are available.

Frequently asked questions

- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

1.2 Download RStudio from https://rstudio.com/products/rstudio/download/#download

Still, choose the right system. The download button will most likely be shown on the website when you click this click, as this website can automatically detect your system and make the recommendation. If not, then just select the right version, download and install.

RStudio Desktop 1.3.959 - Release Notes

1. Install R. RStudio requires R 3.0.1+.

2. Download RStudio Desktop. Recommended for your system:

DOWNLOAD RSTUDIO FOR MAC
1.3.959 | 148.57MB

Requires macOS 10.13+ (64-bit)

All Installers

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an older version of RStudio.

OS	Download	Size	SHA-256
Windows 10/8/7	RStudio-1.3.959.exe	171.41 MB	3d493ae5
macOS 10.13+	RStudio-1.3.959.dmg	148.57 MB	7c5b695d
Ubuntu 16	rstudio-1.3.959-amd64.deb	124.57 MB	c2931495
Ubuntu 18/Debian 10	rstudio-1.3.959-amd64.deb	126.11 MB	411ab500
Fedora 19/Red Hat 7	rstudio-1.3.959-x86_64.rpm	146.24 MB	a144e4e6

Open RStudio and make sure everything is working so far. By default, RStudio will automatically detect the R installed in your system and use that R version.

Step 2: Download the ZIP file (more straightforward way) that contains the code, data and slides from https://github.com/wadesnoopy/Goldschmidt-2020-cluster

Cluster analysis and its application in geochemistry -- A workshop at Goldschmidt 2020   Edit

Manage topics

| -○- 10 commits | ⑂ 1 branch | ⊕ 0 packages | ⬙ 0 releases | ⅍ 1 contributor |

Branch: master ▾   New pull request   Create new file   Upload files   Find file   Clone or download ▾

wadesnoopy Update README.md

Clone with HTTPS ⑦   Use SSH

Use Git or checkout with SVN using the web URL.

📁 R_code	Add files via upload	https://github.com/wadesnoopy/Goldsch 📋
📄 README.md	Update README.md	
📄 cluster_instruction_SZ.pdf	instructions to the cluster analysis	Open in Desktop Download ZIP

📖 README.md   ✎

Goldschmidt-2020-cluster

Code, data, and slides for the 2020 Goldschmidt workshop -- Cluster analysis and its application in geochemistry.

Follow the "cluster_instruction_SZ.pdf" to configure the running environment.

Step 3: Put the ZIP file in your working folder and unzip it. The unzipped folder name will be "Goldschmidt-2020-cluster-master"

Step 4: Navigate into the folder of "Goldschmidt-2020-cluster-master", then enter the folder "R_code". In this folder, you will see four files.

- "**pyrite_samples.csv**" is the data

- "**Goldschmidt-2020-cluster.R**" is the code

- "**Goldschmidt-2020-cluster.Rmd**" is a markdown file that generates the static "Goldschmidt-2020-cluster.html" file.

- "**Goldschmidt-2020-cluster.html**" is the static HTML file

Open "Goldschmidt-2020-cluster.html" to see the code output.
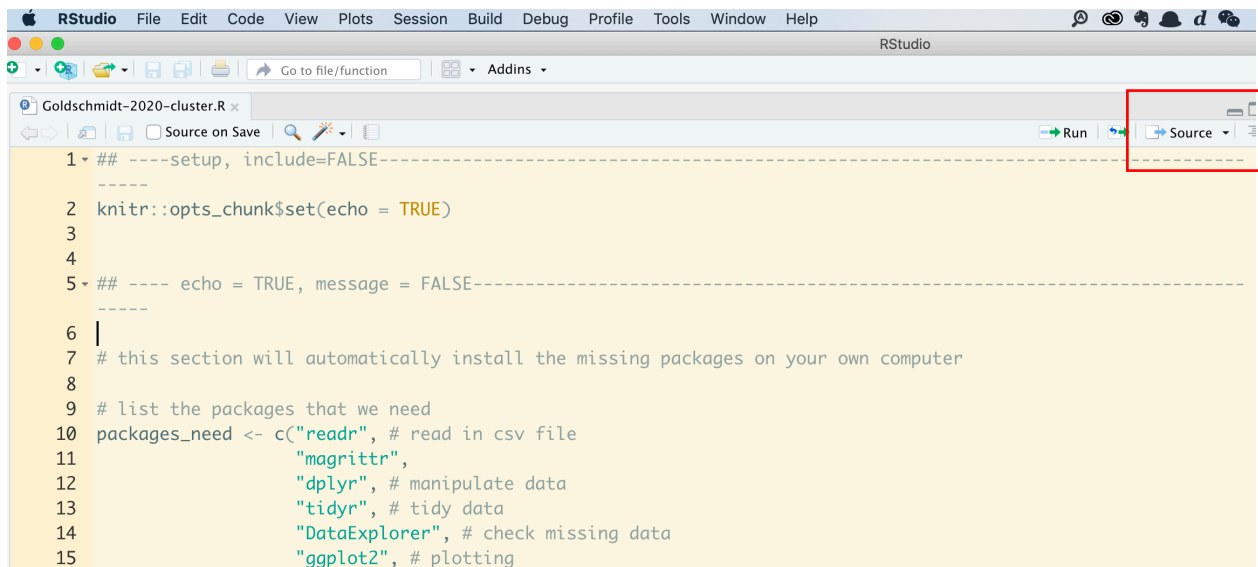
Step 5: Run the code.

"**Goldschmidt-2020-cluster.R**" is the code that we will run. Open it in RStudio. Two ways to run the code.

1) You can run the whole code by first setting the path of working folder to your current "R_code" folder and then clicking the "Source" button. The packages that are needed in this exercise should be automatically installed if they are not on your computer.

Note that for windows, you also need to use the forward slash "/" instead of the backslash "\" when setting up the path.

```
## ---- echo = TRUE, message = FALSE----------------------------------------------------------
-----
# Navigate to your the R_code folder downloaded from github or from this workshop

# change "/Users/shzhang/Documents/Research/Meeting/2020-Goldschmidt/R_code" to your own local path of the
R_code folder downloaded from github or from this workshop

setwd("/Users/shzhang/Documents/Research/Meeting/2020-Goldschmidt/R_code")
```

FYI, The whole code can be finished within 2 minutes on my macbook pro (2019) with 2.3 GHz Intel Core i9. I would guess in general, it can be finished within ~8 minutes using a normal laptop. But if you run the code chunk by chunk, time will not be a problem as each chunk runs relatively fast.

2) You can run the code line by line or chunk by chunk from top to bottom. Select the code chunk, then click the "Code" button on the top panel, and click "Run Selected Line(s)". Tip: to save some time, you can remember the shortcut of "Run Selected Line(s)".
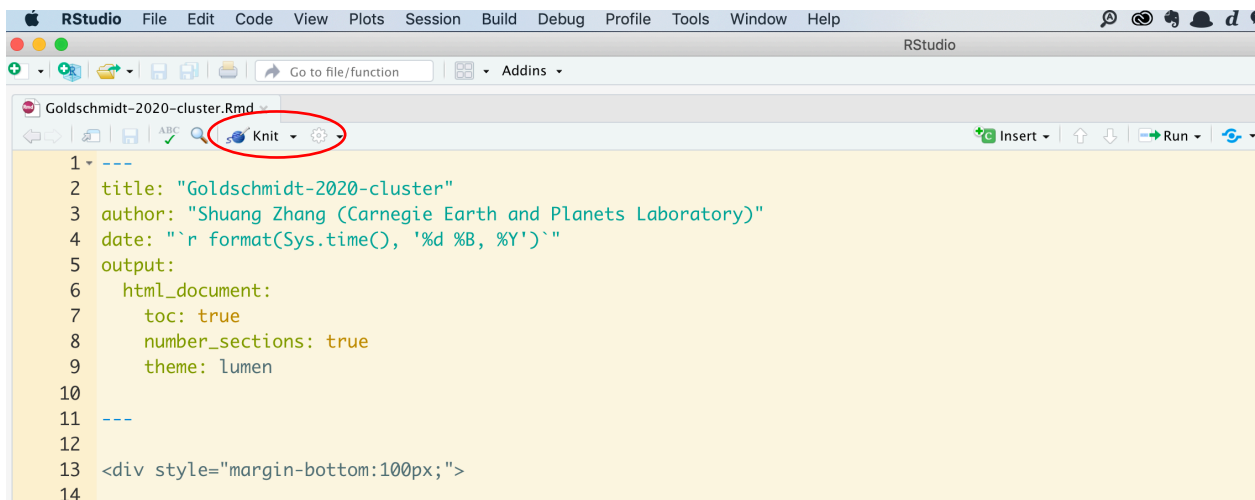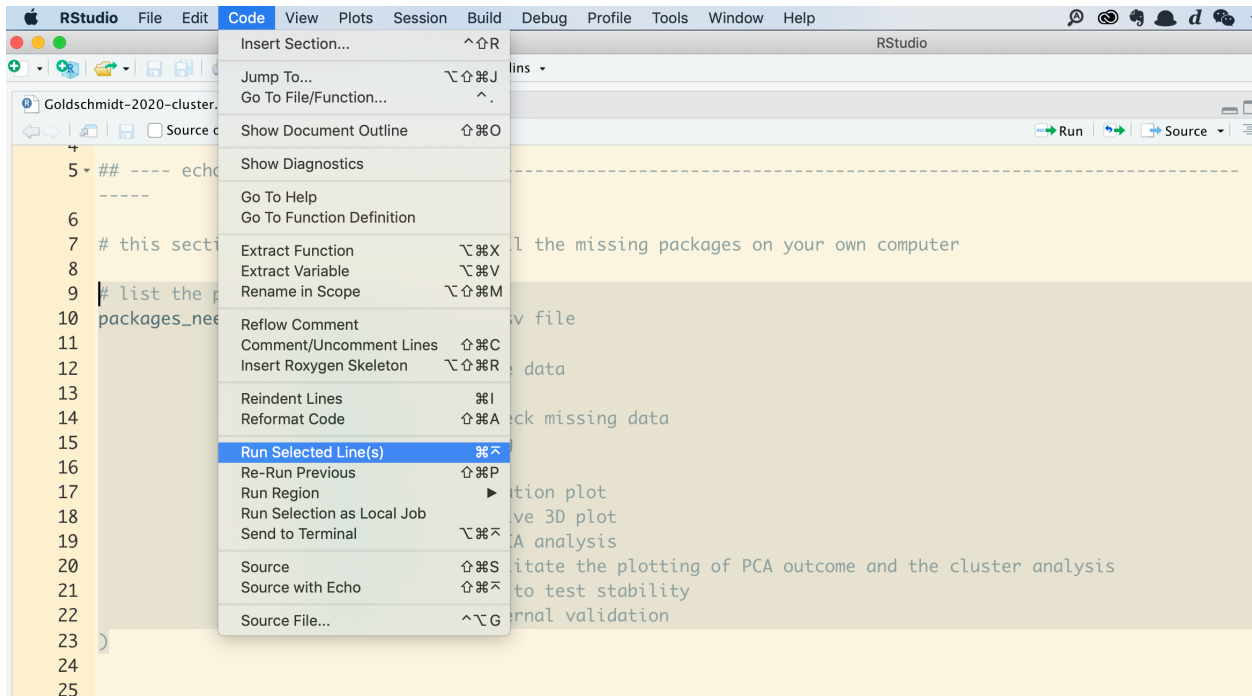
Step 6: Bonus step and optional (you can continue if you like)

I "Knit" the code and the code output to the "**Goldschmidt-2020-cluster.html**" file, which is a good way to share the code output among people. This file is generated by the "**Goldschmidt-2020-cluster.Rmd**" file, which is a markdown file and has the same code content with "**Goldschmidt-2020-cluster.R**".

If you want to generate the "**Goldschmidt-2020-cluster.html**" yourself, you need to open the "**Goldschmidt-2020-cluster.Rmd**" file in RStudio, and then click "Knit". Note that RStudio will notify you that you need to install "Knit" first if "Knit" is not on your computer. Just install it. And then click "Knit".

FYI, The whole code can be finished within 2 minutes on my macbook pro (2019) with 2.3 GHz Intel Core i9. I would guess in general, it can be finished within ~8 minutes using a normal laptop.