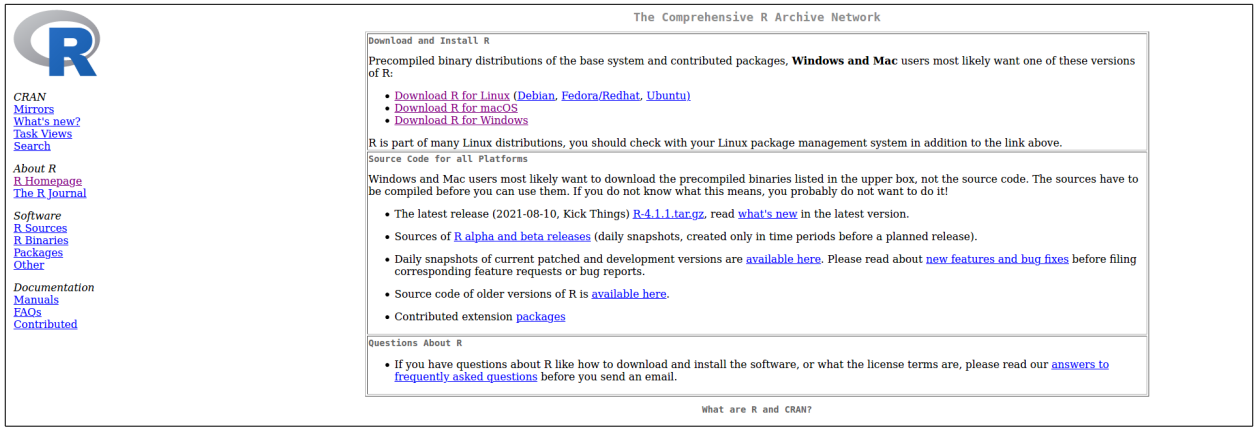
Lipidomics Data Analysis Workshop

Date/time: September 23, 2021, 3:00 pm – 4:30 pm
Where: SRB 120

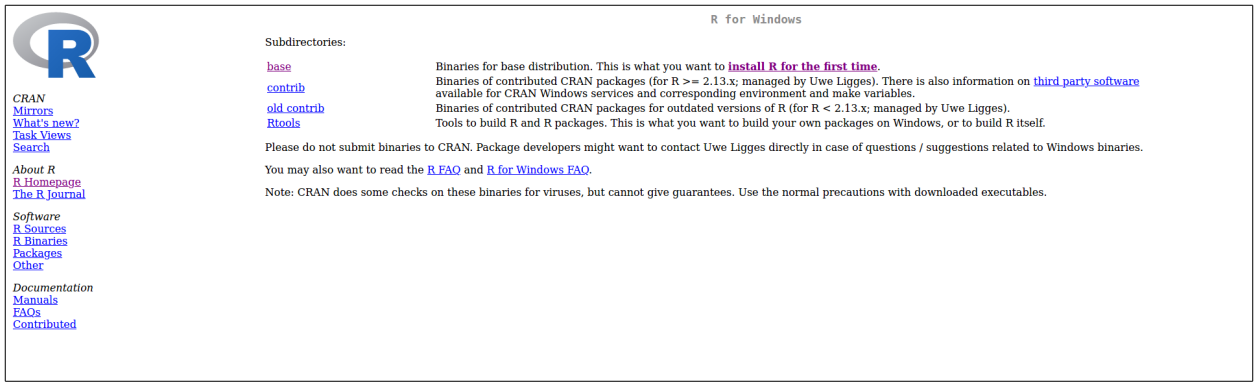
I. The data analysis makes use of an R script. To run this R script on your own computer you will need R and RStudio installed. If you would like to analyze the sample data during the workshop, then the following software will need to be installed:

A. Download R from https://rweb.crmda.ku.edu/cran/

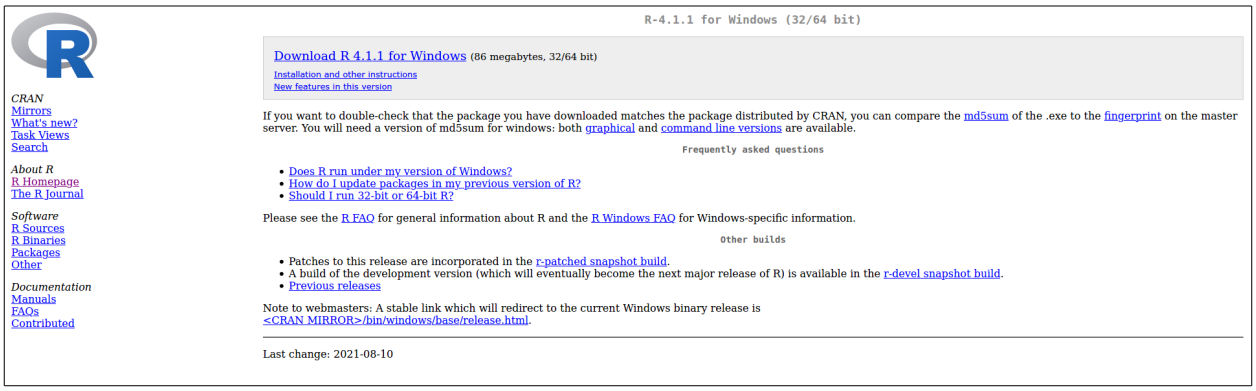
i. Choose your operating system



ii. Choose ‘base’



iii. Click on ‘Download R 4.1.1 for Windows’ and follow install instructions



CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

R-4.1.1 for Windows (32/64 bit)

Download R 4.1.1 for Windows (86 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?

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

Other builds

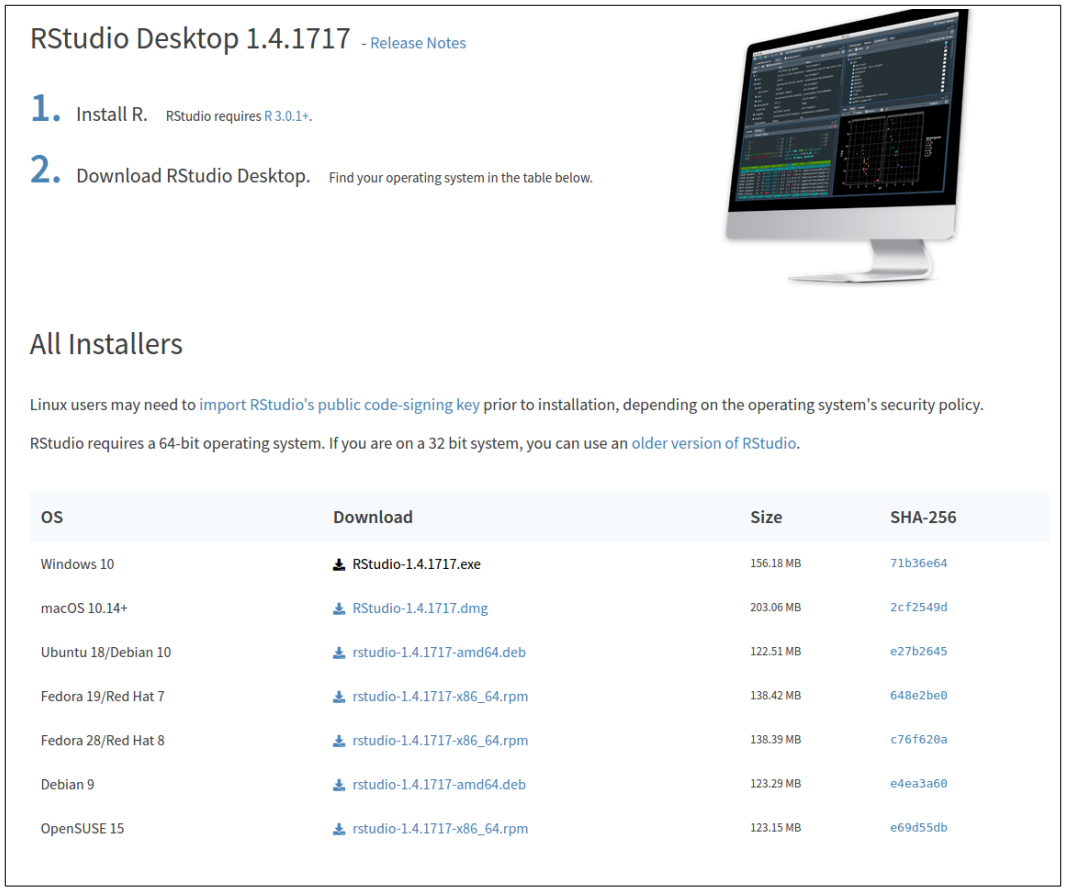
- Patches to this release are incorporated in the r-patched snapshot build.
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
- Previous releases

Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>/bin/windows/base/release.html.

Last change: 2021-08-10

B. Download and Install RStudio from

https://www.rstudio.com/products/rstudio/download/



RStudio Desktop 1.4.1717 - Release Notes

1. Install R. RStudio requires R 3.0.1+.

2. Download RStudio Desktop. Find your operating system in the table below.

All Installers

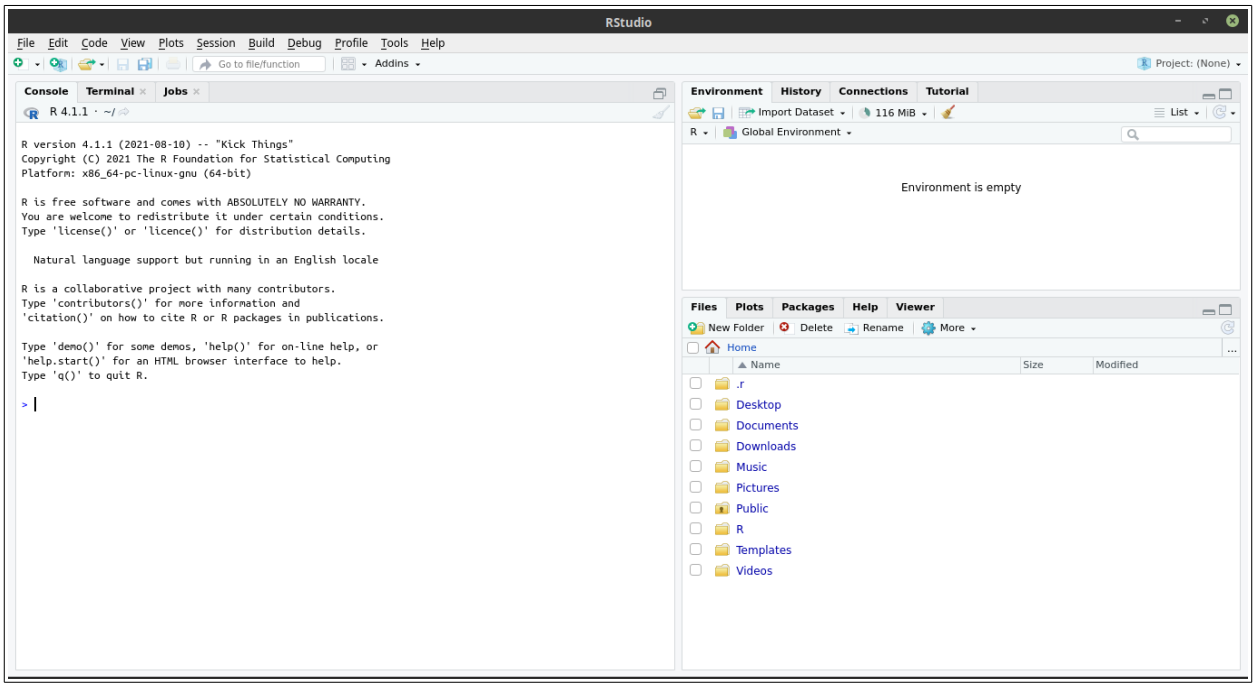
Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

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	RStudio-1.4.1717.exe	156.18 MB	71b36e64
macOS 10.14+	RStudio-1.4.1717.dmg	203.06 MB	2cf2549d
Ubuntu 18/Debian 10	rstudio-1.4.1717-amd64.deb	122.51 MB	e27b2645
Fedora 19/Red Hat 7	rstudio-1.4.1717-x86_64.rpm	138.42 MB	648e2be0
Fedora 28/Red Hat 8	rstudio-1.4.1717-x86_64.rpm	138.39 MB	c76f620a
Debian 9	rstudio-1.4.1717-amd64.deb	123.29 MB	e4ea3a60
OpenSUSE 15	rstudio-1.4.1717-x86_64.rpm	123.15 MB	e69d55db

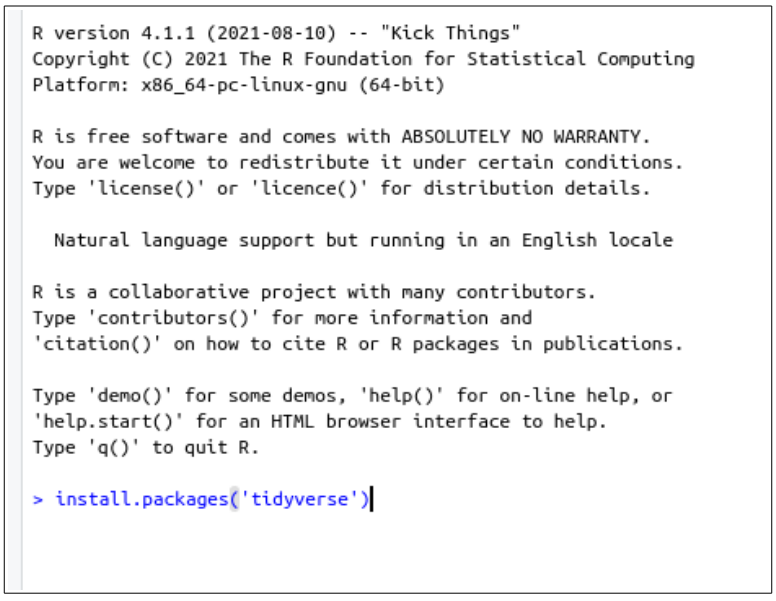
C. Install ‘tidyverse’ and ‘data.table’ R libraries

i. Open RStudio



ii. In the console panel type the following and then press Enter:

```
install.packages('tidyverse')
```



a) Note: you will see many packages install that will be used by R for data and statistical analysis. This will take several minutes.

iii. Next, in the console type the following and then press Enter:

```
install.packages('data.table')
```

iv. Test to make sure libraries have successfully installed by typing the following in the console and pressing Enter after each:

```
library(tidyverse)
```

```
library(data.table)
```

a) Note: If errors occur, you may need to determine if you have missing packages.

D. Download the zip file containing the R scripts and sample data we will be using from here. This is the data we will use to practice the data analysis workflow.

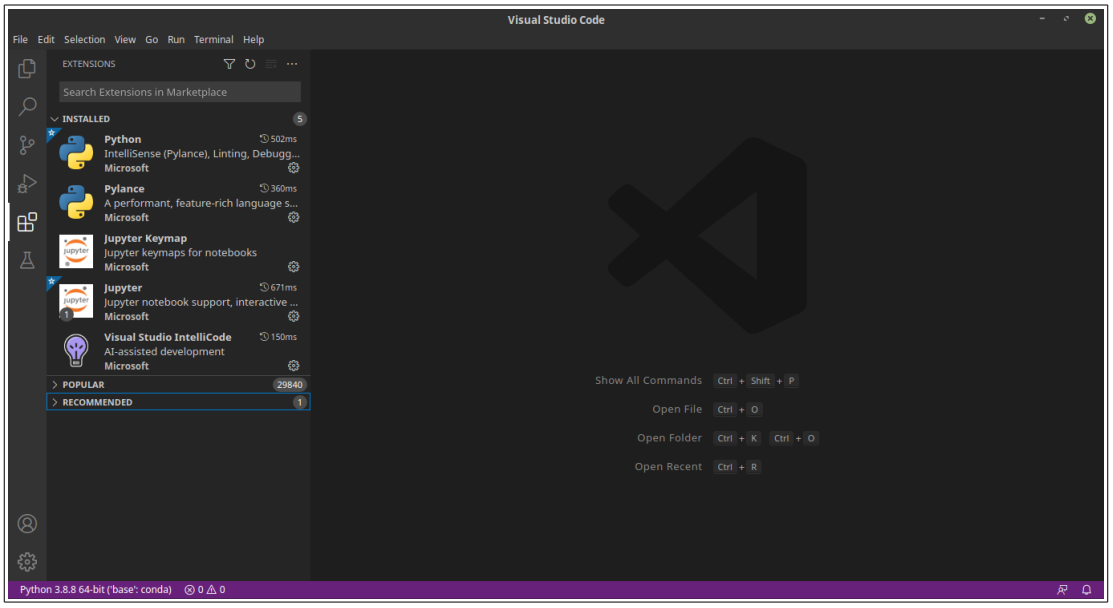
II. There are a few helpful Python scripts that were made and helpful for producing MRM tables or calculating lipid molecular masses and their adduct forms. This section is **OPTIONAL** and not required for the workshop, but if you'd like to make use of these Python scripts the following software is recommended:

A. Anconda is a useful suite of data analysis tools, many of which use Python3. There are two different flavors you can download, either the full suite called ‘Anaconda’ or the pared down version ‘Miniconda’. The former requires a few Gb of space and will take some time to install while the latter will be the minimal install and require additional steps afterwards. Either version can be installed from https://docs.conda.io/projects/conda/en/latest/user-guide/install/index.html. Follow the install instructions there.

   i. Note: If you chose to install Miniconda, be sure to also install the Python libraries for ‘Numpy’, ‘Pandas’, and ‘Matplotlib’. Others may be required.

B. Microsoft’s Visual Studio Code is an integrated development environment (IDE) that can be used for virtually any type of code. Here it is quite useful for Python3 and data analysis and can be downloaded from https://code.visualstudio.com/.

   i. Note: To make the most use of Python3 within Visual Studio Code, also be sure to install the internal apps: ‘Python’, ‘Pylance’, ‘Jupyter’, ‘Visual Studio Intellicode’. These apps will help autocomplete commands and allow for data analysis within the IDE.



III. Below are a list of useful resources and software for lipidomics, plant lipid metabolism, and data analysis. These are for your future reference.

A. R resources
   i. R: https://www.r-project.org/
      a) (additional libraries: tidyverse & data.table)
   ii. Rstudio: https://www.rstudio.com/products/rstudio/
   iii. R for Data Science (online ebook): https://r4ds.had.co.nz/index.html
   iv. Tidyverse library: https://www.tidyverse.org/
   v. Data.Table library: https://cran.r-project.org/web/packages/data.table/vignettes/datatable-intro.html
   vi. W3: https://www.w3schools.com/r/default.asp

B. Python Resources
   i. Anaconda: https://www.anaconda.com/
   ii. Visual Studio Code: https://code.visualstudio.com/
      a) (internal apps: Python, Pylance, Jupyter, Visual Studio Intellicode)
   iii. Python Data Science Handbook (online ebook): https://jakevdp.github.io/PythonDataScienceHandbook/
   iv. Numpy (included with Anaconda): https://numpy.org/
   v. Pandas (included with Anaconda): https://pandas.pydata.org/
   vi. Matplotlib (included with Anaconda): https://matplotlib.org/
   vii. W3: https://www.w3schools.com/python/default.asp

C. mMass (viewing spectra/chromatogram from .txt file): http://www.mmass.org/

D. MarvinSketch (chemical drawing): https://chemaxon.com/products/marvin

E. Metaboanalyst (statistical analysis): https://www.metaboanalyst.ca/

F. Lipid Metabolic Background Information
   i. AOCS Lipid Library: https://lipidlibrary.aocs.org/
   ii. ARALIP: The Arabidopsis Acyl-Lipid Metabolism Website: http://aralip.plantbiology.msu.edu/about
   iii. Acyl-Lipid Metabolism: https://doi.org/10.1199/tab.0133
   iv. LIPID MAPS: https://www.lipidmaps.org/