Setup instructions

Contents

You need to install and test R and RStudio before the workshop. The installation process is simple on all major platforms, and mostly consists of downloading and running the installers appropriate to your platform. Please contact the instructors if you run into difficulties so we can help you show up with a computer properly set up for the workshop.

R and RStudio

- Install R, a free software environment for statistical computing and graphics.
  - It is *highly recommended* to install a precompiled binary distribution for your operating system – use the links up at the top of the page linked to above!
  - If you currently have R installed on your laptop, please make sure it is version 3.4.0 or later. *Please update if it is not!*
- Install RStudio, a powerful user interface for R.

Testing testing

1. Do whatever is appropriate for your OS to launch RStudio. You should get a window similar to the screenshot below.
   - If you don't see a window that is divided into distinct areas labelled "Console", "Environment", etc., you are probably running the user interface that comes bundled with R. Check that RStudio

      is present in your applications and start it instead.
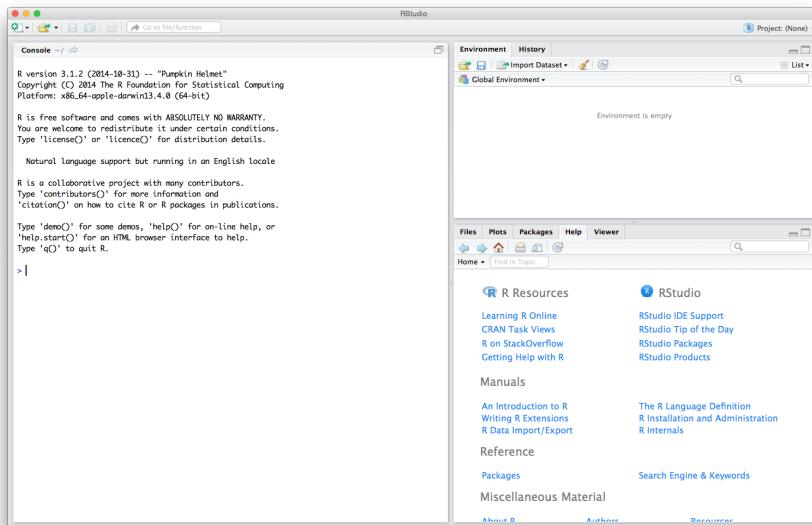
2. Put your cursor in the left pane labelled "Console", which is where you interact with the live R process. Create a simple object with code like `x <- 2 * 4` (followed by enter or return). In the "Environment" pane, to the right of the Console, you should see an entry for "x" with appropriate value (in this case, 8).

3. Then inspect the `x` object by typing `x` in the Console followed by enter or return. You should see the value 8 printed to the screen.

4. Finally, check that plotting works. In the console, type `plot(cars)`. If you see a scatterplot appear in the "Plots" pane to the lower right, you are good to go.

Add-on packages

R is an extensible system and many people share useful code they have developed as a *package* via the Comprehensive R Archive Network CRAN, Bioconductor, or personal GitHub repositories.

From CRAN

To install a package from CRAN in RStudio:

1. Go to the "Tools" menu and choose "Install Packages" or click "Install" in the Packages area in the lower right pane of RStudio.
2. The following window appears
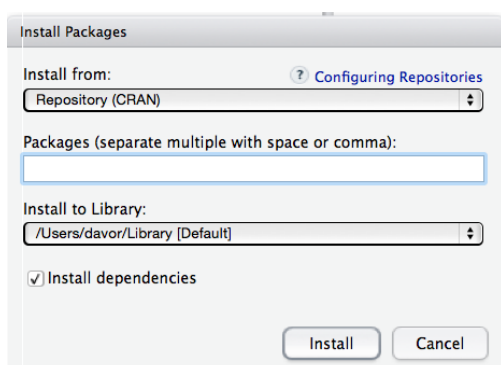
3. Enter the name of the desired package
   - The checkmark by "Install dependencies" should be **on**.
4. Click on the "Install" button

Please install the following packages and their dependencies for use in this workshop:

- tidyverse
  - *Please note that if you have **R v3.3 or older**, you may not be able to install `tidyverse`. In this case, you need to separately install each package within the tidyverse. This includes:* `readr`, `tibble`, `dplyr`, `tidyr`, `stringr`, `ggplot2`, `purr`, `forcats`
- ape
- seqinr
- vegan
- betapart
- abind
- Matrix
- cowplot
- BiocManager
- devtools

From Bioconductor

To install a package from Bioconductor in RStudio:

1. Make sure you have already installed the `BiocManager` package that contains the Bioconductor install function.
2. Then run the following in the R console to install

- phyloseq
- ggtree

```
BiocManager::install("phyloseq")
BiocManager::install("ggtree")
```

From GitHub

To install a package from GitHub in RStudio:

1. Make sure you have already installed the `devtools` package that contains the GitHub install function.
2. Then run the following in the R console to install

- phylofactor

```
devtools::install_github("reptalex/phylofactor")
```

FastTree

Outside of R/RStudio, we will also be using FastTree to build phylogenetic trees. This is a command line program.

1. Install FastTree by following the instructions for your operating system.

2. Open your terminal program.

   - Mac and Linux users can use the pre-installed terminals available with your operating systems.

- Windows users need to install a new terminal since the Windows one is not Linux-based. There are many options but we recommend GitBash as it also provides Git, which is used in other related workshops.

3. Test FastTree in your terminal by running the following, replacing the `PathToSoftware` with wherever you've installed the program on your machine.

```
[PathToSoftware]/FastTree -help
```

You should see the help page like so.

```
/Users/kim/Applications/FastTree/FastTree -help
```

```
## FastTree 2.1.10 SSE3:
##   FastTree protein_alignment > tree
##   FastTree < protein_alignment > tree
##   FastTree -out tree protein_alignment
##   FastTree -nt nucleotide_alignment > tree
##   FastTree -nt -gtr < nucleotide_alignment > tree
##   FastTree < nucleotide_alignment > tree
## FastTree accepts alignments in fasta or phylip interleaved formats
##
## Common options (must be before the alignment file):
##   -quiet to suppress reporting information
##   -nopr to suppress progress indicator
##   -log logfile -- save intermediate trees, settings, and model details
##   -fastest -- speed up the neighbor joining phase & reduce memory usage
##        (recommended for >50,000 sequences)
##   -n <number> to analyze multiple alignments (phylip format only)
##        (use for global bootstrap, with seqboot and CompareToBootstrap.pl)
##   -nosupport to not compute support values
##   -intree newick_file to set the starting tree(s)
##   -intree1 newick_file to use this starting tree for all the alignments
##        (for faster global bootstrap on huge alignments)
##   -pseudo to use pseudocounts (recommended for highly gapped sequences)
##   -gtr -- generalized time-reversible model (nucleotide alignments only)
##   -lg -- Le-Gascuel 2008 model (amino acid alignments only)
##   -wag -- Whelan-And-Goldman 2001 model (amino acid alignments only)
##   -quote -- allow spaces and other restricted characters (but not ' ) in
##          sequence names and quote names in the output tree (fasta input only;
##          FastTree will not be able to read these trees back in)
##   -noml to turn off maximum-likelihood
##   -nome to turn off minimum-evolution NNIs and SPRs
##        (recommended if running additional ML NNIs with -intree)
##   -nome -mllen with -intree to optimize branch lengths for a fixed topology
##   -cat # to specify the number of rate categories of sites (default 20)
##      or -nocat to use constant rates
##   -gamma -- after optimizing the tree under the CAT approximation,
##      rescale the lengths to optimize the Gamma20 likelihood
##   -constraints constraintAlignment to constrain the topology search
##        constraintAlignment should have 1s or 0s to indicates splits
##   -expert -- see more options
## For more information, see http://www.microbesonline.org/fasttree/
```

Further resources

The above is sufficient preparation for this workshop but here are some links if you are interested in reading a bit further.

- How to Use RStudio:
    - https://support.rstudio.com/hc/en-us/sections/200107586-Using-RStudio
- RStudio Public Discussion & Troubleshooting Guide:
    - https://support.rstudio.com/hc/en-us/sections/203994097-RStudio-IDE
- How to Install R:
    - http://cran.r-project.org/doc/manuals/R-admin.html
    - http://cran.stat.sfu.ca/doc/FAQ/R-FAQ.html#How-can-R-be-installed_003f
- R FAQ:
    - http://cran.r-project.org/doc/FAQ/R-FAQ.html
- More about add-on packages in the R Installation and Administration Manual
    - http://cran.r-project.org/doc/manuals/R-admin.html#Add_002don-packages