

# Resources

Below are miscellaneous resources (mostly for R) that you may find useful.

## Getting Started with R and RStudio

RStudio 2021.09.0+351 “Ghost Orchid” Release (077589bcad3467ae79f318afe8641a1899a51606, 2021-09-20) for Windows Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.8 Chrome/69.0.3497.128 Safari/537.36

You should do the following as soon as possible to be sure you are ready to use R. 1. Download and install R from [www.r-project.org](http://www.r-project.org) for your operating system. If you already have R installed, be sure you have the latest version installed — i.e., R version 4.1.2 (2021-11-01) nicknamed “Bird Hippie.” You can see the version when starting R/RStudio, or by typing the command `version$version.string` in the console. It should show `R version 4.1.2 (2021-11-01)`. For this course you must use this version of R (or later if a newer version is released during the semester).

0. Download and install the free open source license desktop version RStudio from [www.rstudio.com](http://www.rstudio.com). If you already have RStudio installed, be sure you have the latest version installed (version 2021.9.0.351 or later if a new version is released this semester).
1. Run RStudio to verify that everything is working. Note that R runs “under” RStudio (the latter is what is sometimes called an “integrated development environment” or IDE). Tinker around with RStudio/R if you’ve not used it before. Explore some of the options in “Global Options...” under the “Tools” drop-down menu to suit your tastes. Many of these options you should not change if you don’t understand fully what you are doing, but many of the options under *Code*, *Appearance*, and *Pane Layout* are safe.
2. Try installing a package from the Comprehensive R Archive Network (CRAN) repository by typing the command `install.packages("devtools")` at the `>` prompt in the console window of RStudio to install the **devtools** package. This package is used to install my **trtools** package (see the next step). Also install the package **tidyverse** with `install.packages("tidyverse")`. This will install several packages that we will be using throughout the semester for data manipulation and plotting. You can also try updating any packages you already have installed or that came with R by typing the command `update.packages()` at the prompt in the console window. Note that `update.packages()` will only update packages that are on the CRAN repository. Updating a package from elsewhere (like GitHub) can be done by simply re-installing the package in the usual way. Installing or updating packages from CRAN can also be done via the “Tools” menu in RStudio.
3. Install my **trtools** package using by typing the command `devtools::install_github("trobinj/trtools")` at the prompt in the console window of RStudio, assuming you have already installed the package **devtools** as described in the previous step. Installing **trtools** require a different command because it is hosted on GitHub rather than on CRAN.

If you run into problems with any of the steps above let me know.

## R Packages

The following are a few interesting and sometimes useful R packages.

**anytime** is useful for converting a variety of variable types into dates/times. See the vignette for some examples.

**colorspace** provides tools to select and manipulate individual colors and color palettes for graphics and plots.

**colourpicker** is useful when trying to find the name or HEX value for a color for a plot. It adds an add-in to RStudio that helps identify different colors.

**cowplot** extends the capabilities of **ggplot** in several ways, but what I find particularly useful is the **plot\_grid** function for combining and aligning several plots into one. See the vignettes that come with the package for some examples.

**dplyr** is very helpful for manipulating data and computing basic descriptive statistics. It is part of the “tidyverse” of R packages.

**emmeans** is designed to estimate “marginal means” for linear and generalized linear models. It is a very useful package for making specific inferences based on a linear or generalized linear (mixed) model model. Its functionality is similar to the **contrast** function in the **trtools** package.

**forcats** includes some useful functions for working with factors. It is part of the “tidyverse” of R packages.

**gganimate** lets you produce animated plots with **ggplot** that can be exported as gifs.

**lubridate** is very useful when working with time or dates. It is part of the “tidyverse” of R packages.

**rmarkdown** lets you create documents that combine text, R code, and the results of running that code. Almost all of the documents I create for my classes, include this web page, are created using **rmarkdown**.

**Rcpp** facilitates the interfacing C++ code with R. It allows for more computationally efficient code, but it also extends C/C++ with useful classes and functions. This is highly recommended for C++ programmers (or anyone who would like to learn C++). There are also several packages that allow you to easily install and access various C++ libraries. Some examples are **RcppArmadillo** (my favorite) for using the Armadillo C++ linear algebra library with Rcpp, **RcppEigen** for using the Eigen numerical library, **RcppGSL** for using the GNU Scientific Library (GSL), and **RcppEnsmallen** for using the Ensmallen optimisation library. For an introduction to using the **Rcpp** package I would recommend starting with the book chapter in *Advanced R* on Rcpp, the github “book” *Rcpp for Everyone*, and the blog entry *Introduction to Rcpp*.

**tesseract** lets you use the tesseract optical character recognition (OCR) software from R. I have found this useful for reading data into R from a scanned document from a book or article, although it can be a little tricky to calibrate.

**tidyr** is used to rearrange data frames, particularly between “long form” (with one observation per row) and “wide form” with multiple observations per row. It is part of the “tidyverse” of R packages. It is a successor to the **reshape2** package which still exists but is no longer being updated except to keep it working.

**tidyverse** is actually a collection of packages for manipulating and plotting including **dplyr**, **forcats**, **lubridate**, **tidyr**, and **ggplot2**. You can install all of these packages at once by installing the **tidyverse** package. Also you can load all of these packages at once with **library(tidyverse)**.

**trtools** is a package I created originally for teaching, but now I and others also use it for research. It contains some data sets that I use in classes as well as several utility functions to facilitate certain kinds of tasks that are not (in my opinion) as easily done with other packages or functions. It is not available on CRAN so it cannot be installed using **install.packages**. To install it use **devtools::install\_github("trobinj/trtools")** (assuming you already have the **devtools** package installed). Note that this package is a work in progress.

There are several packages that add new color palettes for use with the **ggplot2** package. Several of based on themes such as the **wesanderson** package for color palettes used in Wes Anderson films, the **ghibli** package for palettes inspired by films produced by Studio Ghibli, and the **gameofthrones** package for palettes inspired by the Game of Thrones television series.

## Books

The following are some free books about using R that you might find useful.

*Advanced R* covers some of the more advanced aspects of the R programming language.

*Data Visualization for Social Science* is a very nice book on visualization using R with the **ggplot2** package and some supporting packages. And despite the title, this book would be useful for applications outside the social sciences.

*The Hitchhiker's Guide to Ggplot2* is a introduction to the **ggplot2** package with many fun examples.

*R Packages* is an introduction to creating your own R package. Making a R package is a useful way to organize your own R code and to disseminate that work to others.

*R for Data Science* is an introduction to R. It also features using some capabilities of the tidyverse packages (e.g., **dplyr**, **tidyr**, and **ggplot2**).

*R Programming for Data Science* is an introduction to R.