

# Introduction to the R Statistical Environment

Sarah Hunter

*sghunter@umich.edu*

ICPSR Day 1

# Overview

- 1 Introductions
- 2 Getting to know R
- 3 Downloading the Programs
- 4 Taking a Tour
- 5 Getting Started in R
- 6 Workflow in R
- 7 Working with Packages

## Introductions

- Getting to know R
- Downloading the Programs
- Taking a Tour
- Getting Started in R
- Workflow in R
- Working with Packages

# An Introduction to Me



## Introductions

Getting to know R

Downloading the Programs

Taking a Tour

Getting Started in R

Workflow in R

Working with Packages

# How to Get in Touch

**Office:** Newberry 223

**Office Hours:** 3:30-5 M-F (Zoom Link on Canvas)

**Email:** [sghunter@umich.edu](mailto:sghunter@umich.edu)

# What is this course and Why should I take it?

## Course Description

This is a class designed to introduce you to the basics of the **R** statistical Environment. This is a free, open-source statistical program used by many fields in academia and in industry. **R** is flexible and has a wide range of uses.

# What is this course and Why should I take it?

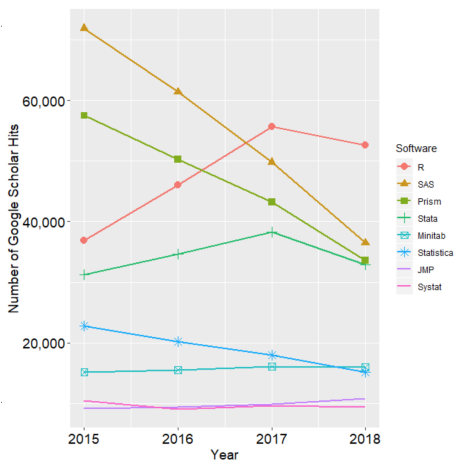


Figure 26 The number of Google Scholar hits for Google Scholar that list the following software

# What to expect from this course

By the end of this course, you will...

- Understand the workflow in R using R Studio
- Know the basic structure and logic of the R programming language
- Perform data management with R
- Calculate and interpret descriptive statistics in R
- Use data visualization tools such as base plot and ggplot
- Be familiar with the tidyverse suite of tools for data management and data visualization
- Perform simple linear regression in R
- Understand simple programming in R

# Class Structure

**Lecture** Most days will be a lecture explaining the logic behind R coding and the basics of how to use the program.

**Labs** Some days will have in class labs, or a set of tasks to perform in R



## Topics Covered in this Class

- Data Manipulation
- Descriptive Statistics
- Data Visualization
- Debugging
- Linear Models and Extensions
- Model Presentation
- Simple Programming
- Special Topics if time permits

## Books and Materials

- Recommended Texts
  - Harris, Jenine K. *Statistics with R: Solving Problems Using Real-World Data*. 2021. Los Angeles: Sage.
  - Wickham, Hadley. and Garrett Golemund. 2017. *R for Data Science: Visualize, Model, Transform, Tidy, and Import Data*. O'Reilly Media, Inc. [FREE at <https://r4ds.had.co.nz/index.html>]
  - Fox, John, and Sanford Weisberg. 2018. *An R Companion to Applied Regression, 3rd Edition*. SAGE.
- Software
  - R
  - R Studio

# How to Get the Most from this course

- 1 Come to class
- 2 Listen/read
- 3 **Practice**

# Course Outline

Day	Topics
1	Introduction and Workflow
2	Loading and cleaning Data in R
3	Using the tidyverse for advanced data manipulation
4	Descriptive Statistics
5	Data Visualization with ggplot
6	Debugging your code
7	Common Models
8	Presenting results
9	Loops and functions

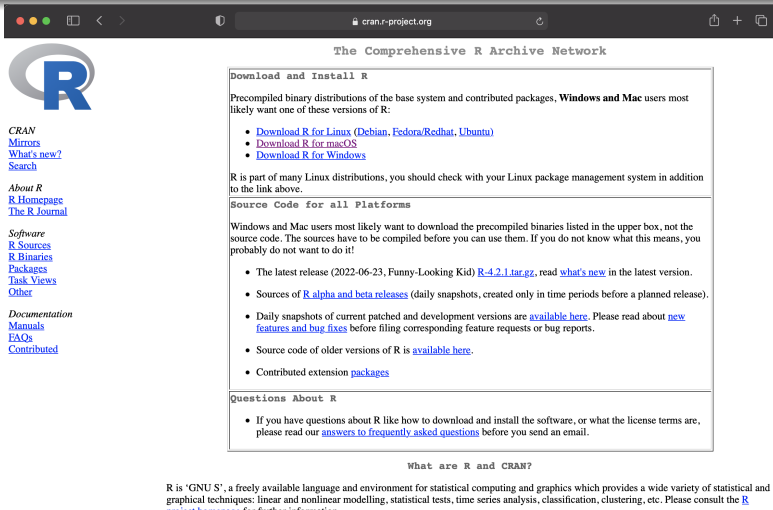
## So, R... What is so special about it?

- Free and open-source
- User-created packages
- Flexible and wide range of uses
- Better data visualization
- Highly valued skill in academia and industry

# The not-so-great parts of R

- Steep learning curve
- 10 ways to do any one task
- Highly sensitive
- No “point-and-click”

# Downloading Basic R



The screenshot shows the CRAN website with the following content:

**The Comprehensive R Archive Network**

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2022-06-23, Funny-Looking Kid) [R-4.2.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

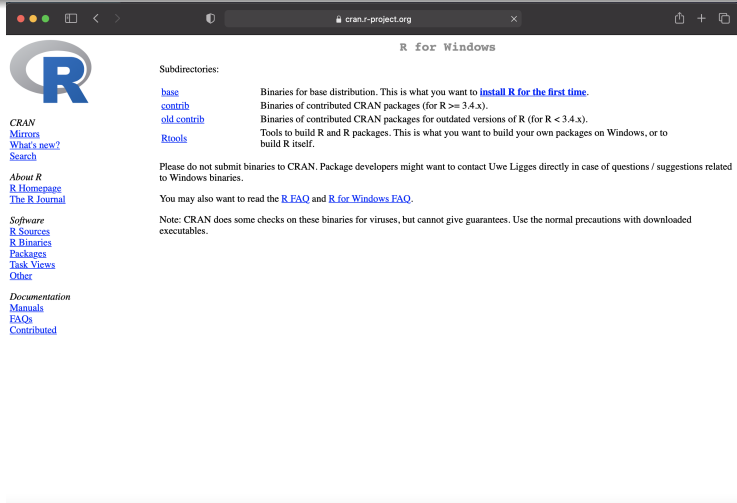
### Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

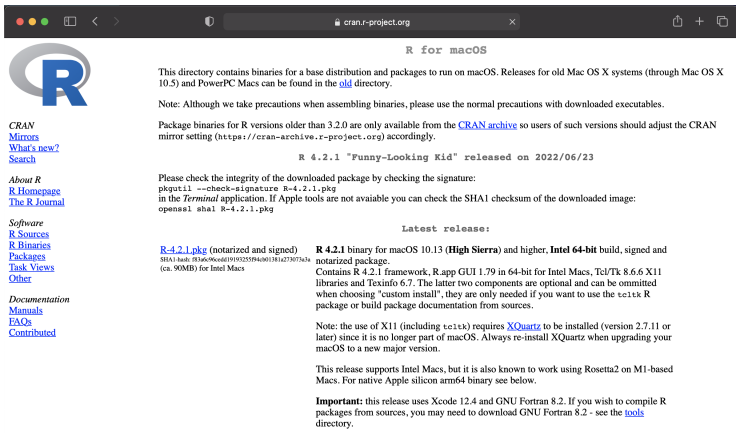
# R for Windows



The screenshot shows a web browser window with the address bar set to `cran.r-project.org`. The page title is "R for Windows". On the left side, there is a navigation menu with the following links: **CRAN**, [Mirrors](#), [What's new?](#), [Search](#), **About R**, [R Homepage](#), [The R Journal](#), **Software**, [R Sources](#), [R Binaries](#), [Packages](#), [Task Views](#), [Other](#), **Documentation**, [Manuals](#), [FAQs](#), and [Contributed](#). The main content area is titled "Subdirectories:" and lists four categories: [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 >= 3.4.x).), [old contrib](#) (Binaries of contributed CRAN packages for outdated versions of R (for R < 3.4.x).), and [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.). Below this, a paragraph states: "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." Another paragraph says: "You may also want to read the [R FAQ](#) and [R for Windows FAQ](#)." A final note reads: "Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables." The browser's address bar and navigation icons are visible at the top, and the operating system's taskbar is partially visible at the bottom right.



# R for Macs



The screenshot shows a web browser window displaying the CRAN R for macOS page. The browser's address bar shows 'cran.r-project.org'. The page title is 'R for macOS'. On the left side, there is a navigation menu with links for 'CRAN', 'Mirrors', 'What's new?', 'Search', 'About R', 'R Homepage', 'The R Journal', 'Software', 'R Sources', 'R Binaries', 'Packages', 'Task Views', 'Other', 'Documentation', 'Manuals', 'FAQs', and 'Contributed'. The main content area contains the following text:

**R for macOS**

This directory contains binaries for a base distribution and packages to run on macOS. Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

Package binaries for R versions older than 3.2.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting (<https://cran-archive.r-project.org>) accordingly.

**R 4.2.1 "Funny-Looking Kid" released on 2022/06/23**

Please check the integrity of the downloaded package by checking the signature:

```
pkgtutil --check-signature R-4.2.1.pkg
```

in the *Terminal* application. If Apple tools are not available you can check the SHA1 checksum of the downloaded image:

```
openssl sha1 R-4.2.1.pkg
```

**Latest release:**

[R-4.2.1.pkg](#) (notarized and signed)  
SHA1 hash: 934696c0d19193255f84001381a273075a34  
(ca. 90MB) for Intel Macs

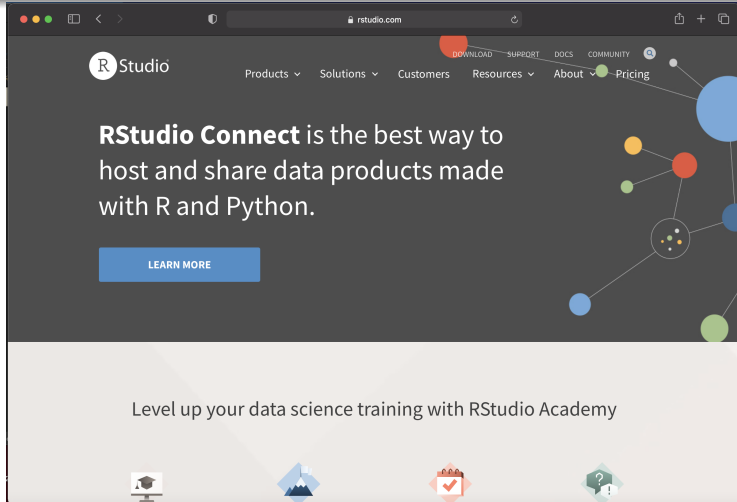
**R 4.2.1** binary for macOS 10.13 (**High Sierra**) and higher, **Intel 64-bit** build, signed and notarized package. Contains R 4.2.1 framework, R app GUI 1.79 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 6.7. The latter two components are optional and can be omitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) to be installed (version 2.7.11 or later) since it is no longer part of macOS. Always re-install XQuartz when upgrading your macOS to a new major version.

This release supports Intel Macs, but it is also known to work using Rosetta2 on M1-based Macs. For native Apple silicon arm64 binary see below.

**Important:** this release uses Xcode 12.4 and GNU Fortran 8.2. If you wish to compile R packages from sources, you may need to download GNU Fortran 8.2 - see the [tools](#) directory.

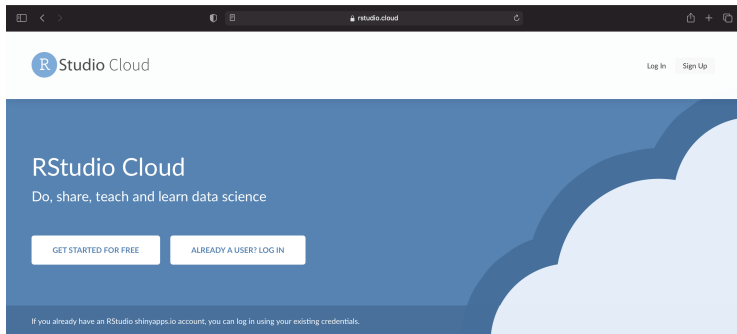
# R Studio



The screenshot shows the RStudio website homepage. The browser address bar displays 'rstudio.com'. The navigation menu includes 'DOWNLOAD', 'SUPPORT', 'DOCS', 'COMMUNITY', 'Products', 'Solutions', 'Customers', 'Resources', 'About', and 'Pricing'. The main headline reads: 'RStudio Connect is the best way to host and share data products made with R and Python.' Below this is a blue 'LEARN MORE' button. At the bottom, there is a section titled 'Level up your data science training with RStudio Academy' with four icons representing different educational or training paths.

Introductions  
Getting to know R  
Downloading the Programs  
Taking a Tour  
Getting Started in R  
Workflow in R  
Working with Packages

# R Studio Cloud



The screenshot shows the RStudio Cloud website. At the top left is the RStudio logo and the text "Studio Cloud". At the top right are "Log In" and "Sign Up" links. The main content area has a blue background with a white cloud graphic on the right. It features the heading "RStudio Cloud" and the tagline "Do, share, teach and learn data science". Below this are two white buttons: "GET STARTED FOR FREE" and "ALREADY A USER? LOG IN". At the bottom of the main area, a small line of text reads: "If you already have an RStudio shinyapps.io account, you can log in using your existing credentials."

## Data science without the hardware hassles

RStudio Cloud is a lightweight, cloud-based solution that allows anyone to do, share, teach and learn data science online.

- Analyze your data using the RStudio IDE, directly from your browser.

[\\$ AVAILABLE PRICING PLANS](#)

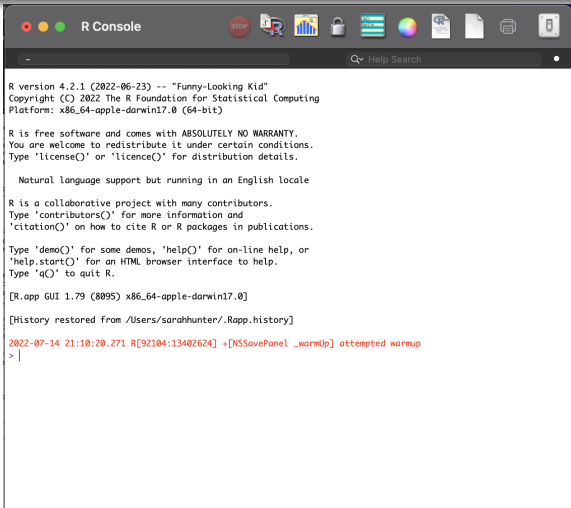
[📖 RSTUDIO CLOUD GUIDE](#)

[🌐 RSTUDIO.COM](#)

# Downloading the Programs

Take Time Now to Download the Programs

# Basic R



```
R Console
R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

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.

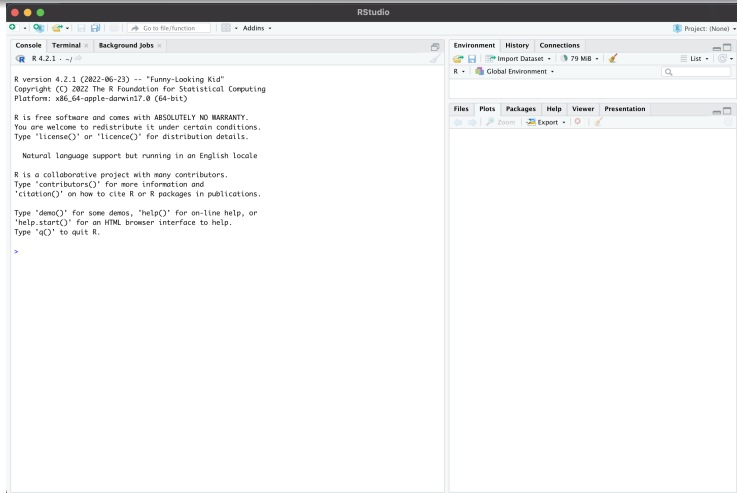
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.79 (8095) x86_64-apple-darwin17.0]

[History restored from /Users/sarahhunter/.Rapp.history]

2022-07-14 21:10:20.271 R[92104:13402624] +[NSSavePanel _warmUp] attempted warmup
> |
```

# R Studio



# R's Basic Functions

R is, at its heart, a fancy calculator.

# Objects in R

See R.

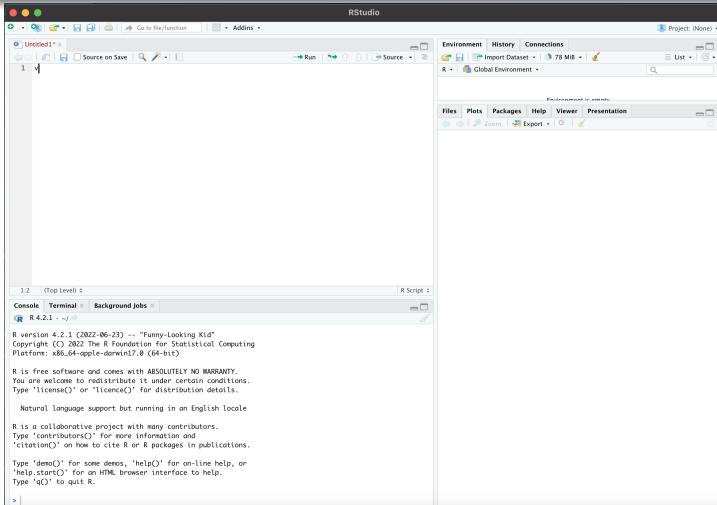


# The Script File

Save. Your. Code.

- Saving code for future reference'
- Replication
- So you don't have to figure it out again

# The Script File



# R Packages

## Defined

“collection of functions, data, and documentation that extends the capabilities of base R (Wickman and Grolemond)”

Packages help you:

- Perform specific (user-created) functions
- Access certain datasets
- Create specialized plots/figures

# tidyr: An Example

Files Plots Packages Help Viewer Presentation

Programming with tidyr Find in Topic

## Programming with tidyr

### Introduction

Most tidy verbs use **tidy evaluation** to make interactive data exploration fast and fluid. Tidy evaluation is a special type of non-standard evaluation used throughout the tidyverse. Here's some typical tidy code:

```
library(tidyr)

iris %>%
  nest(data = !Species)
#> # A tibble: 3 x 2
#>   Species data
#>   <fct>   <list>
#> 1 setosa  <tibble [50 x 4]>
#> 2 versicolor <tibble [50 x 4]>
#> 3 virginica <tibble [50 x 4]>
```

Tidy evaluation is why we can use `!Species` to say "all the columns except `Species`", without having to quote the column name ("`Species`") or refer to the enclosing data frame (`iris$Species`).

Two basic forms of tidy evaluation are used in tidy:

- **Tidy selection:** `drop_na()`, `fill()`, `pivot_longer()`/`pivot_wider()`, `nest()`/`unnest()`, `separate()`/`extract()`, and `unite()` let you select variables based on position, name, or type (e.g. `1:3`, `starts_with("x")`, or `is.numeric`). Literally, you can use all the same techniques as with `dplyr::select()`.
- **Data masking:** `expand()`, `crossing()` and `nesting()` let you refer to use data variables as if they were variables in the environment (i.e. you write `my_variable` not `df$myvariable`).

We focus on tidy selection here, since it's the most common. You can learn more about data masking in the equivalent vignette in dplyr: <https://dplyr.tidyverse.org/dev/articles/programming.html>. For other considerations when writing tidy code in packages, please see vignette "[In-packages](#)".

# Using R Packages

To R!

# Until Next Time