

Chapter 0

Introduction to R

This guidebook will serve as an introduction to R and RStudio. We begin by showing how to download and install the necessary programs. The guide is written so that it follows the organization of the text. The following online tutorials/readings are very useful if you want to read additional R guides:

- <https://www.rstudio.com/resources/cheatsheets/>
- <https://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>
- <https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf>

Installing R

To download and install R to your computer, go to the following R CRAN website:
<https://cran.r-project.org/mirrors.html>

This webpage includes a list of R mirrors. See Figure 1. Each R mirror listed is a separate download for R. These are called mirrors, because each link is a replica of the rest. The reason there are so many mirrors is to allow faster download speeds and to ensure that network traffic is well maintained. All of these downloads are exactly the same, so scroll down the page. Click on the mirror that is in closest proximity to your location to ensure faster download speeds. If there is confusion about which mirror to choose, it does not matter. The only negative aspect of choosing the wrong mirror is slower download speeds.

Figure 1

| | |
|---|---|
| https://cran.cnr.berkeley.edu/ | University of California, Berkeley, CA |
| http://cran.cnr.berkeley.edu/ | University of California, Berkeley, CA |
| http://cran.stat.ucla.edu/ | University of California, Los Angeles, CA |
| https://mirror.las.iastate.edu/CRAN/ | Iowa State University, Ames, IA |
| http://mirror.las.iastate.edu/CRAN/ | Iowa State University, Ames, IA |
| https://ftp.ussg.iu.edu/CRAN/ | Indiana University |
| http://ftp.ussg.iu.edu/CRAN/ | Indiana University |
| https://rweb.crmda.ku.edu/cran/ | University of Kansas, Lawrence, KS |
| http://rweb.crmda.ku.edu/cran/ | University of Kansas, Lawrence, KS |
| https://cran.mtu.edu/ | Michigan Technological University, Houghton, MI |
| http://cran.wustl.edu/ | Michigan Technological University, Houghton, MI |
| http://cran.wustl.edu/ | Washington University, St. Louis, MO |
| http://archive.linux.duke.edu/cran/ | Duke University, Durham, NC |

After clicking on the appropriate R mirror, a webpage like Figure 2 will appear. If you have a Windows computer, click “Download R for Windows.” If you have a Mac computer, click “Download R for (Mac) OS X.”

Figure 2

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](#)
- [Download R for \(Mac\) OS X](#)
- [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.

From there, Windows computer users will see Figure 3.

Figure 3



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

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 \geq 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. |

Click on the base subdirectory. This allows a download of all of the base packages for R. Mac computer users will see Figure 4.

Figure 4



[FAQ](#)
[errors](#)
[what's new?](#)
[ask Views](#)
[arch](#)

[about R](#)
[Homepage](#)
[the R Journal](#)

[software](#)
[Sources](#)
[Binaries](#)
[packages](#)
[here](#)

[documentation](#)
[manuals](#)
[FAQs](#)
[contributed](#)

R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) [here](#). Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

As of 2016/03/01 package binaries for R versions older than 2.12.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting accordingly.

R 3.5.1 "Feather Spray" released on 2018/07/05

Important: since R 3.4.0 release we are now providing binaries for OS X 10.11 (El Capitan) and higher using non-Apple toolkit to provide support for OpenMP and C++17 standard features. To compile packages you may have to download tools from the [tools](#) directory and read the corresponding note below.

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type

```
md5 R-3.5.1.pkg  
in the Terminal application to print the MD5 checksum for the R-3.5.1.pkg image. On Mac OS X 10.7 and later you can also validate the signature using  
pkgutil --check-signature R-3.5.1.pkg
```

Lastest release:

[R-3.5.1.pkg](#)

```
MD5-hash: 58ea2f65f0d0242d07ef1e521e17e78  
SHA1-  
hash: 76c01b5f62a6896d5f9e4511e25d17276d149621  
(ca. 74MB)
```

R 3.5.1 binary for OS X 10.11 (El Capitan) and higher, signed package.

Contains R 3.5.1 framework, R.app GUI 1.70 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 5.2. 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 since it is no longer part of OS X. Always re-install XQuartz when upgrading your macOS to a new major version.

Important: this release uses Clang 6.0.0 and GNU Fortran 6.1, neither of which is supplied by Apple. If you wish to compile R packages from sources, you will need to download and install those tools - see the [tools](#) directory.

Click on the latest release package. Because R constantly updates and improves, new updates are released frequently. In Figure 4, the most recent update is "R-3.5.1 pkg." Your most recent update will likely be different! After clicking on the most recent update "R-3.X.X pkg," open the downloaded file and complete the prompted download instructions. And that is it. R is now officially downloaded on your computer! If issues arise, please refer to one of the guides linked at the beginning of this document.

Installing RStudio

The next step will be installing RStudio. RStudio is an open source license that is designed to increase productivity with R. It includes a code editor, as well as debugging and visualization tools. RStudio is not necessary to code in R, but it is highly recommended. The interface is much simpler than R and will make life easier. To download RStudio, go to the following website: <https://www.rstudio.com/products/RStudio/>

Click on the Download RStudio Desktop Open Source License and Figure 5 will appear.

Figure 5

Installers for Supported Platforms

| Installers | Size | Date | MD5 |
|--|---------|------------|----------------------------------|
| RStudio 1.1.456 - Windows Vista/7/8/10 | 85.8 MB | 2018-07-19 | 24ca3fe0dad8187aabd4bfb9dc2b5ad |
| RStudio 1.1.456 - Mac OS X 10.6+ (64-bit) | 74.5 MB | 2018-07-19 | 4fc4f4f70845b142bf96dc1a5b1dc556 |
| RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit) | 89.3 MB | 2018-07-19 | 3493f9d5839e3a3d697f40b7bb1ce961 |
| RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit) | 97.4 MB | 2018-07-19 | 863ae806120358fa0146e4d14cd75be4 |
| RStudio 1.1.456 - Ubuntu 16.04+/Debian 9+ (64-bit) | 64.9 MB | 2018-07-19 | d96e63548c2add890bac633bdb883f32 |
| RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit) | 88.1 MB | 2018-07-19 | 1df56c7cd80e2634f8a9fdd11ca1fb2d |
| RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit) | 90.6 MB | 2018-07-19 | 5e77094a88fdbdddb0d35708752462 |

Windows computer users should click on the first download (Windows vista/7/8/10). Mac computer users should click on the second download (Mac OS C 10.6+ (64-bit)). After clicking on the download links, follow the prompted download instructions.

RStudio Cloud Rather than downloading the desktop version of RStudio, you might consider the RStudio Cloud at rstudio.cloud.

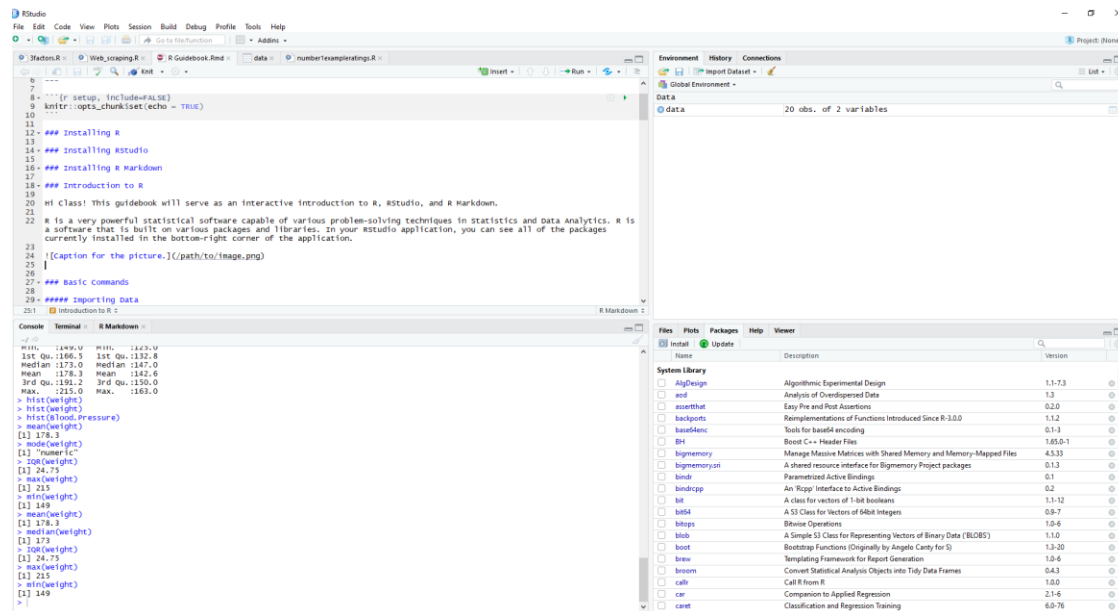
Introduction to R and RStudio

R is a very powerful statistical language capable of various problem-solving techniques in Probability, Statistics, and Data Analytics. For most Data Scientists, R is one of the two or three most commonly used programming languages.

R is a programming language built on various packages and libraries, ranging from simpler visualization/ graphing packages to more complex machine learning libraries. The R community is constantly developing new libraries to be used for data analysis.

RStudio was created to be a simple interface to perform statistical analysis using R. Figure 6 is a screenshot of the RStudio interface. The top-left corner is where the R commands are written (called R script). The user can type R commands in this area and save it without having to run code line by line. The bottom-left corner is known as the console. The user can write R commands in this area to try it line by line instead of writing an entire R program in the script window. The top-right corner is the data window. RStudio will display all imported datasets in this corner. In Figure 6, note the term “data” in the data window. This is a dataset that was imported into RStudio. As seen in Figure 6, this dataset has 20 observations of 2 variables. In order to view this dataset, the user simply clicks on the word “data” in the data environment area. A benefit of working in RStudio is the ability to review data sets. Finally, in the bottom-right corner, the user can view and search for different libraries and packages. This is also where plots/visualizations are displayed. All of these capabilities play a role in the user-friendly nature of RStudio.

Figure 6



R is valuable, because it is one of the few statistical softwares that can handle large amounts of data and process it quickly. Data science is a rapidly growing professional field, and data science is needed in nearly every industry. Annual demand for data scientists is expected to grow by nearly 364,000 listings, 28% growth a year.

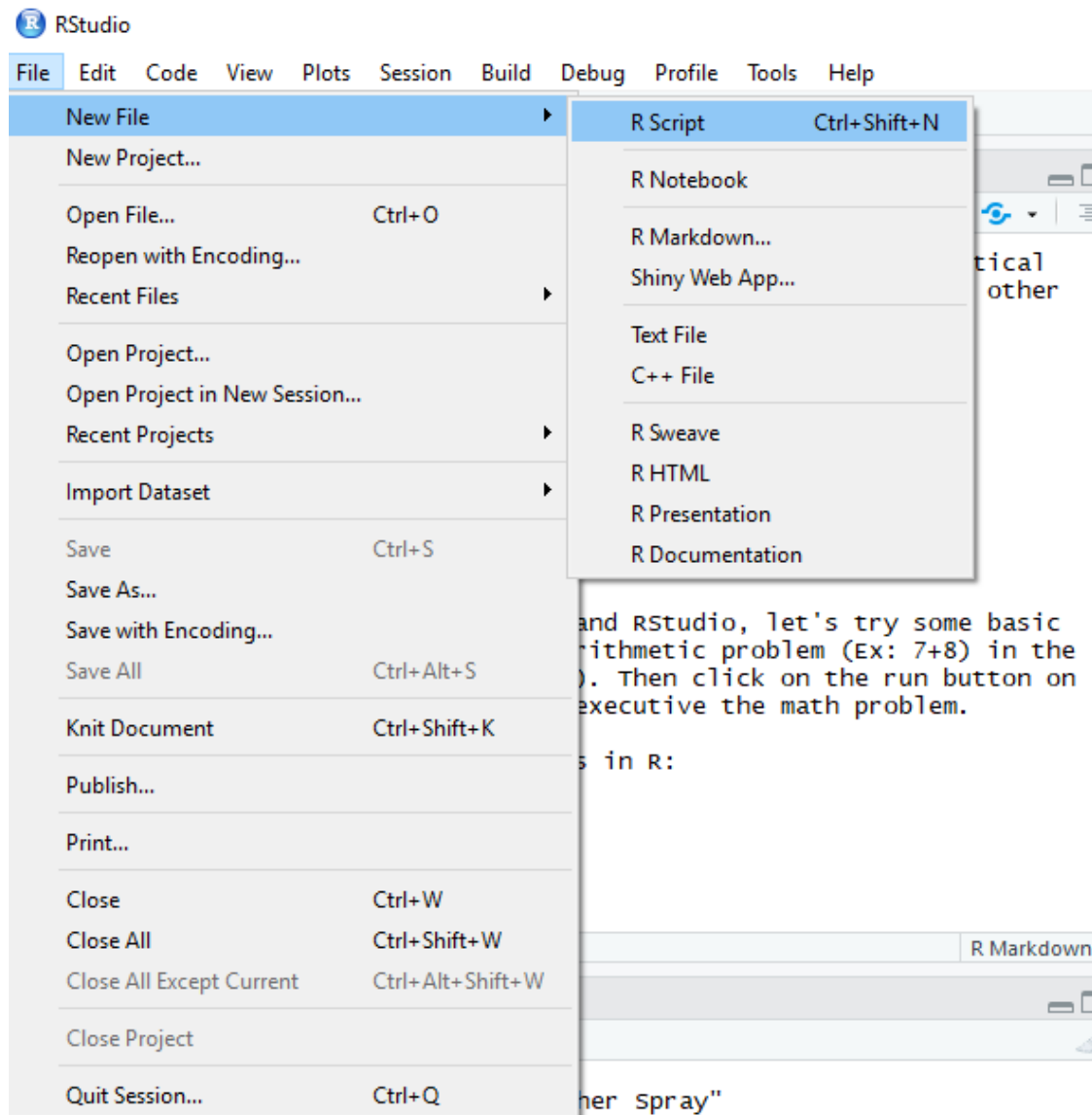
R can be used for a number of analyses, including mathematical calculations, statistical analysis, data visualization, or other advanced data science techniques.

Basic R Operations

Start a New Project

Now that you have installed RStudio, let's start a new project! Open RStudio on your computer or go to rstudio.cloud and select "New Project". Using the top menu, click on File -> New File -> R Script as shown in Figure 7. Save the untitled workbook in the desired folder on your computer and you are ready to get started.

Figure 7



The other way to open a new workbook is to click on the green plus sign right below “File”. Then click on R Script and you have opened a new workbook just like the previous method.

Arithmetic

Now that you know how to open a workbook in RStudio, let’s try some basic commands in R. Type a simple arithmetic expression in the R text editor (top-left corner). Then click on the run button on the toolbar. R will do the arithmetic.

Example: Type the following problems in R:

- $7 + 8$
- $999 - 876$

- $7*4 - 3$

$8/3+9$



The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar. The main editor window, titled 'Untitled2*', contains a script with four lines of code:

```
1 7+8
2 999-876
3 7*4-3
4 8/3+9
```

Red text in the center of the editor reads: "Type the expressions here. Highlight all and click Run." The bottom console window shows the output of these expressions:

```
> 7+8
[1] 15
> 999-876
[1] 123
> 7*4-3
[1] 25
> 8/3+9
[1] 11.66667
>
```

Red text in the center of the console reads: "The results appear down here."

Clearly, R can handle simple math and knows Order of Operations. What about more advanced calculations?

More Advanced Calculations

R knows the value of pi and how to do more advanced calculations like squareroot, log, and exponentials. Note that the `log()` command in R is the natural logarithm and the `exp()` command is the number e raised to a power. In RStudio, type the following directly in the console (bottom-left):

```

> pi
[1] 3.141593
> log(8)
[1] 2.079442
> sqrt(9)
[1] 3
> exp(3)
[1] 20.08554
> 6^2
[1] 36
> |

```

Variables

If you want, you can also create variables in R. For example, the command `x <- 3` will store the value 3 in the variable `x`. Variable names do not need to be a single character. For example, `Weight <- 130` will store the value 130 in the variable `Weight`. See below. Notice that R will not assume multiplication in the expression $4x - 12$.

```

> x <- 3
> x
[1] 3
> Weight <- 130
> Weight
[1] 130
> 4x - 12
Error: unexpected symbol in "4x"
> 4*x-12
[1] 0
> |

```

Importing Data

Data should always be imported into R. Most data sets are available as csv (comma separated values) files that are used with most spreadsheets (such as Excel or GoogleSheets). Data is going to either be stored on the Internet or on your computer. For data on the Internet, all you need is the URL that identifies the data. The data for the Sullivan texts are housed at GitHub. The URL for most data sets is

https://sullystats.github.io/Statistics6e/Data/Chapterx/x_s_p.csv

where

- x is the chapter number
- s is the section number (r is use for the chapter review problems and ct is used for the chapter test problems).
- p is the problem number

For example, to retrieve the data set for Problem 31 in Section 4.1, use the URL

https://sullystats.github.io/Statistics6e/Data/Chapter4/4_1_31.csv

To retrieve the data set for Problem 10 in the Chapter 2 Review, use the URL

https://sullystats.github.io/Statistics6e/Data/Chapter2/2_r_10.csv

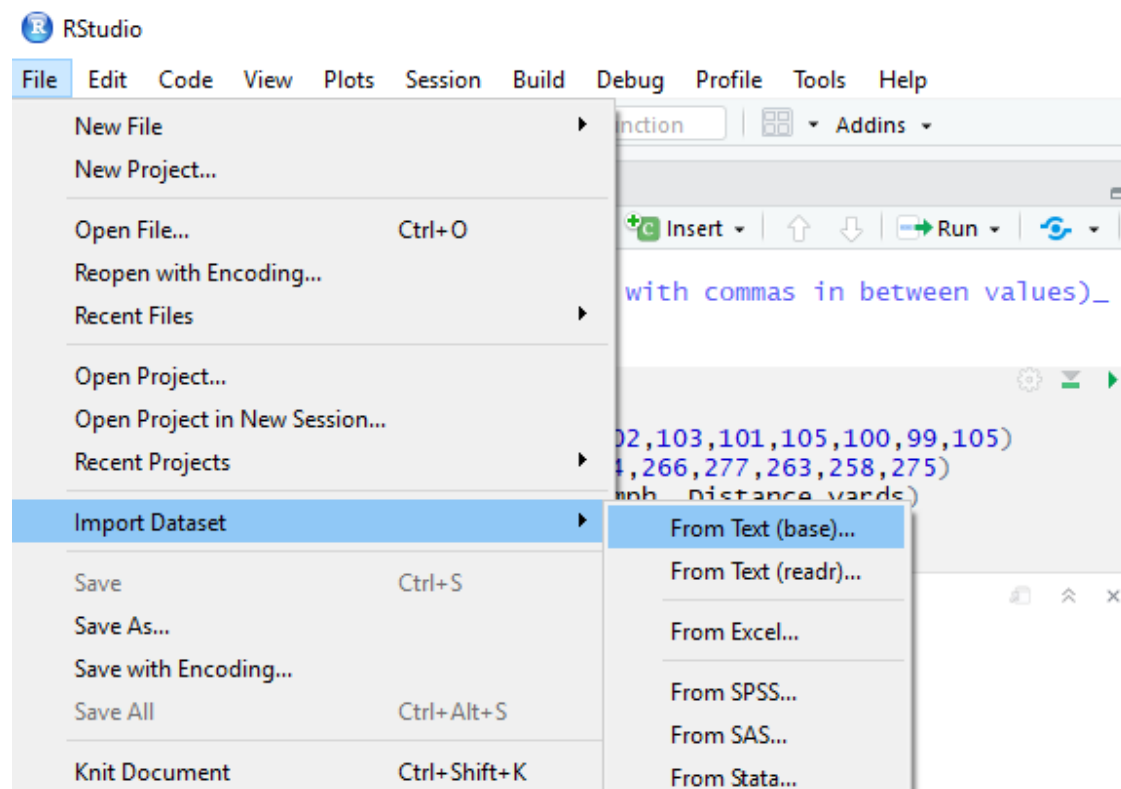
To retrieve data online from the text, go to www.pearsonhighered.com/sullivanstats. A new window will open. The first bullet item displayed is Data Sets. Click on the TXT.zip link. You will be prompted to save the zip file to your computer's hard-drive. Create a folder for this class and save the zip file. Once the file downloads (it is a large file so this may take a few minutes), click on the file in your system tray. An "unzipped" folder will appear in the same location as the zip file. Feel free to rename the folder (something like "StatsData"). Within this folder, there is a separate folder that contains data files for each chapter of the text. Complete the same process for the CSV.zip link.

Importing TXT Files

Now that you have all the data downloaded, you are ready to import a dataset.

For .txt files, from the toolbar at the top of RStudio, click File -> Import Dataset -> From Text(base)... See Figure 8.

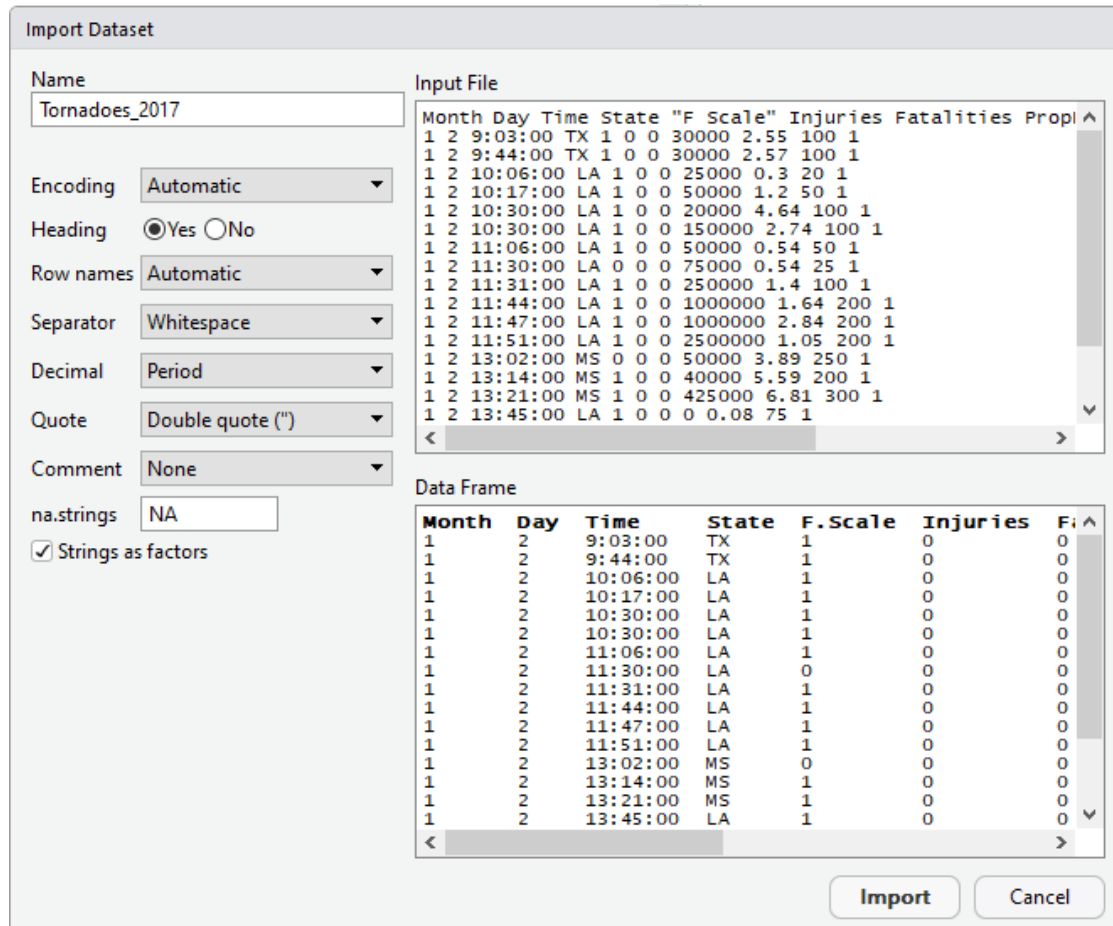
Figure 8



A popup will appear. Browse to the desired file and Click Open.

Then another popup will appear (See Figure 9). Click Import in the popup box and your txt file will appear both as a tab in you text editor area (top-left corner) and as a dataset stored in RStudio's global environment (top-right corner).

Figure 9

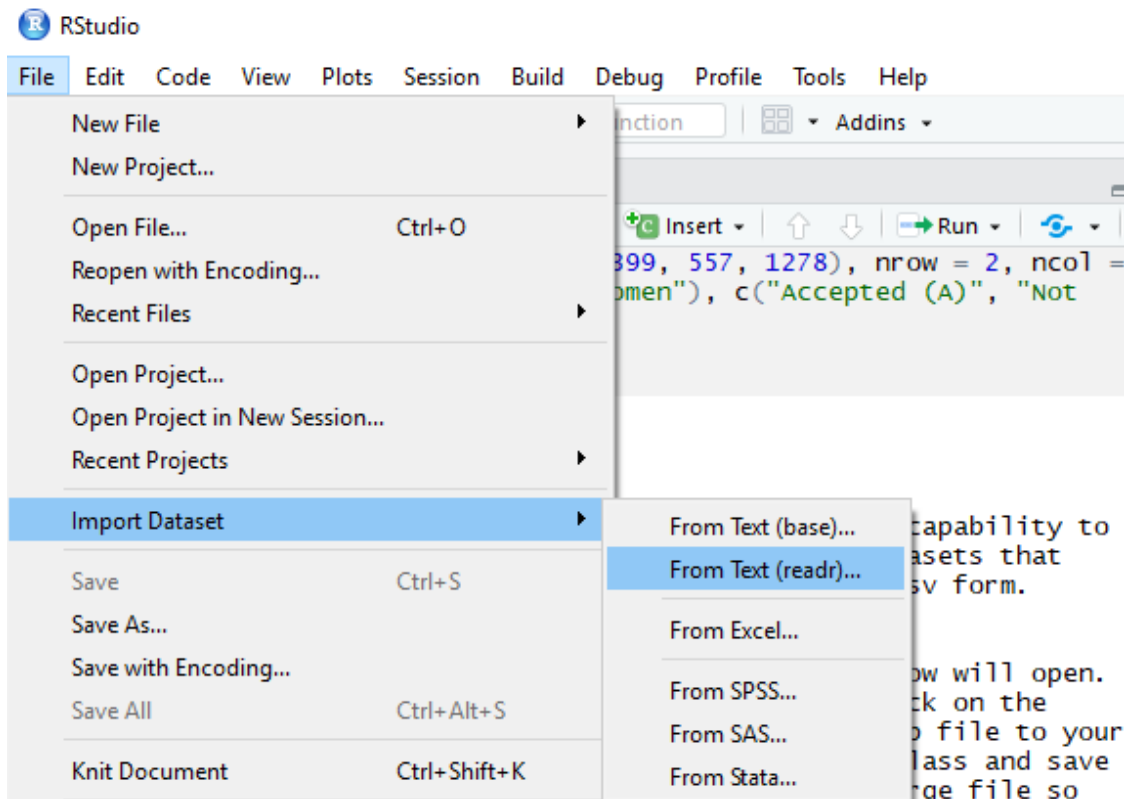


Importing CSV Files

Importing csv files follows a similar but slightly different process from txt files.

In the top toolbar, click File -> Import Dataset -> From Text(readr).... See Figure 10.

Figure 10



A popup will appear. Click Browse and find the desired dataset you want to import. Then click Import. The imported file will appear as a tab in the text editor area of RStudio and as a dataset in the Global Environment.

Creating New Data Files

Do not bother with entering data into R. Rather, use a spreadsheet such as Excel or GoogleSheets to create a csv file. Save it to GitHub or your computer, and then load the data set.

Final Note All of the data files we are working with are technically referred to as *data frames*. If you every want to see the contents of a data frame, use the following command:

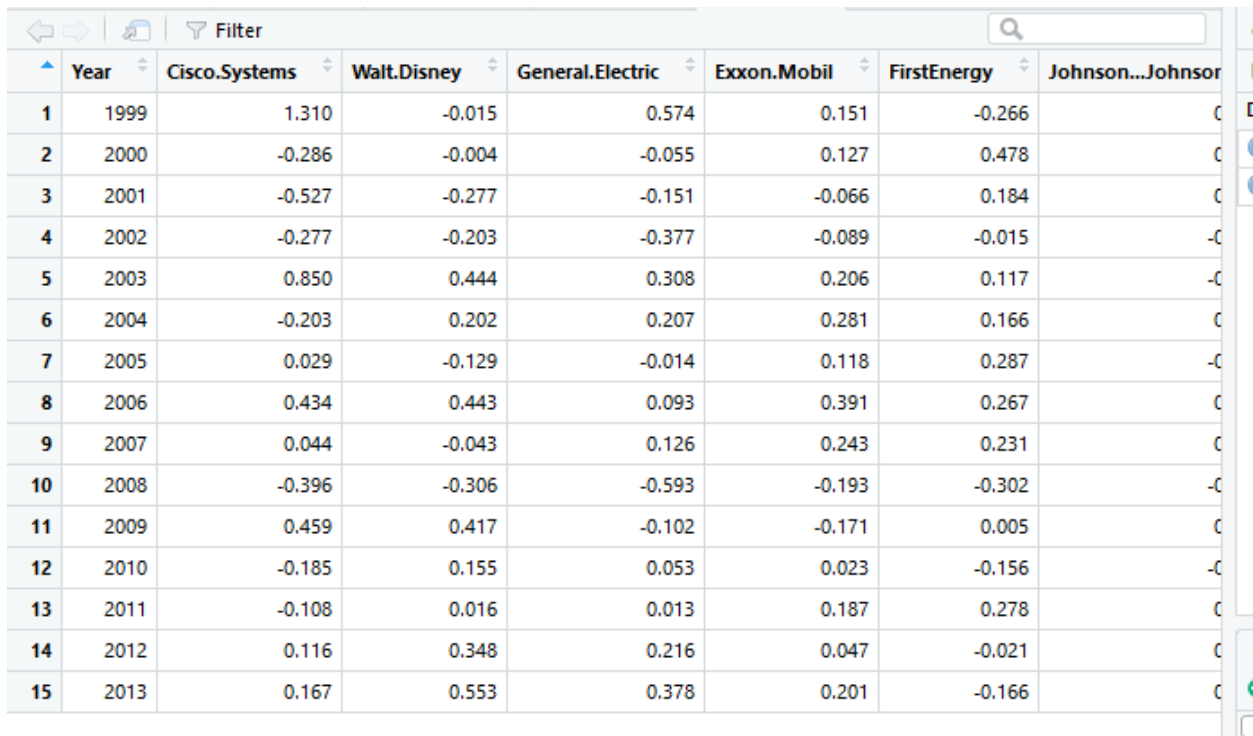
```
View(name)
```

For example, suppose we import the data frame “4_1_39.txt” and name it Stocks. We can view the contents of the data frame as follows:

```
> Stocks <- read.delim("C:/Users/Sullivan Texts Inc/iCloudDrive/Stats/Stats  
5e/sst5e_data_txt/ch04/4_1_39.txt")  
> View(Stocks)
```

The data frame is shown in Figure 11 on the next page.

Figure 11



| | Year | Cisco.Systems | Walt.Disney | General.Electric | Exxon.Mobil | FirstEnergy | Johnson...Johnson |
|----|------|---------------|-------------|------------------|-------------|-------------|-------------------|
| 1 | 1999 | 1.310 | -0.015 | 0.574 | 0.151 | -0.266 | |
| 2 | 2000 | -0.286 | -0.004 | -0.055 | 0.127 | 0.478 | |
| 3 | 2001 | -0.527 | -0.277 | -0.151 | -0.066 | 0.184 | |
| 4 | 2002 | -0.277 | -0.203 | -0.377 | -0.089 | -0.015 | |
| 5 | 2003 | 0.850 | 0.444 | 0.308 | 0.206 | 0.117 | |
| 6 | 2004 | -0.203 | 0.202 | 0.207 | 0.281 | 0.166 | |
| 7 | 2005 | 0.029 | -0.129 | -0.014 | 0.118 | 0.287 | |
| 8 | 2006 | 0.434 | 0.443 | 0.093 | 0.391 | 0.267 | |
| 9 | 2007 | 0.044 | -0.043 | 0.126 | 0.243 | 0.231 | |
| 10 | 2008 | -0.396 | -0.306 | -0.593 | -0.193 | -0.302 | |
| 11 | 2009 | 0.459 | 0.417 | -0.102 | -0.171 | 0.005 | |
| 12 | 2010 | -0.185 | 0.155 | 0.053 | 0.023 | -0.156 | |
| 13 | 2011 | -0.108 | 0.016 | 0.013 | 0.187 | 0.278 | |
| 14 | 2012 | 0.116 | 0.348 | 0.216 | 0.047 | -0.021 | |
| 15 | 2013 | 0.167 | 0.553 | 0.378 | 0.201 | -0.166 | |

Packages in R

The R language is comprised of a variety of packages that have been written by members of the R community. These core packages (called **base R**) represent a collection of *functions* that allow us to perform a variety of statistical procedures. In addition, there are additional packages available to install. One such package is *MOSAIC*, which was written by Nicholas J. Horton, Randall Pruim, and Daniel T. Kaplan.

To use any package in R (beyond those in base R), it must be installed. This only needs to be done once (like installing any software on your computer). To install *MOSAIC* use the following command:

```
> install.packages("mosaic")
```

Please be patient when installing because R must communicate and upload to your computer. Once the package is installed, you will need to load it into your session. This must be done for each R session. Think of this as opening your software on your computer. The following command will load *MOSAIC*.

```
> library(mosaic)
```

Now, we are ready to work with R.