



Getting Started with R

David Keyes // R for the Rest of Us



Installation



Install R

The first thing you need to do is download the R software. Go to the [Comprehensive R Archive Network \(aka “CRAN”\) website](#) and download the software for your operating system (Windows, Mac, or Linux).



Working Directly in R



RStudio

R: Engine



RStudio: Dashboard



Courtesy [Modern Dive](#)



RStudio

If you use RStudio, you'll have a graphical user interface, the ability to see all of your stored information, and much more.

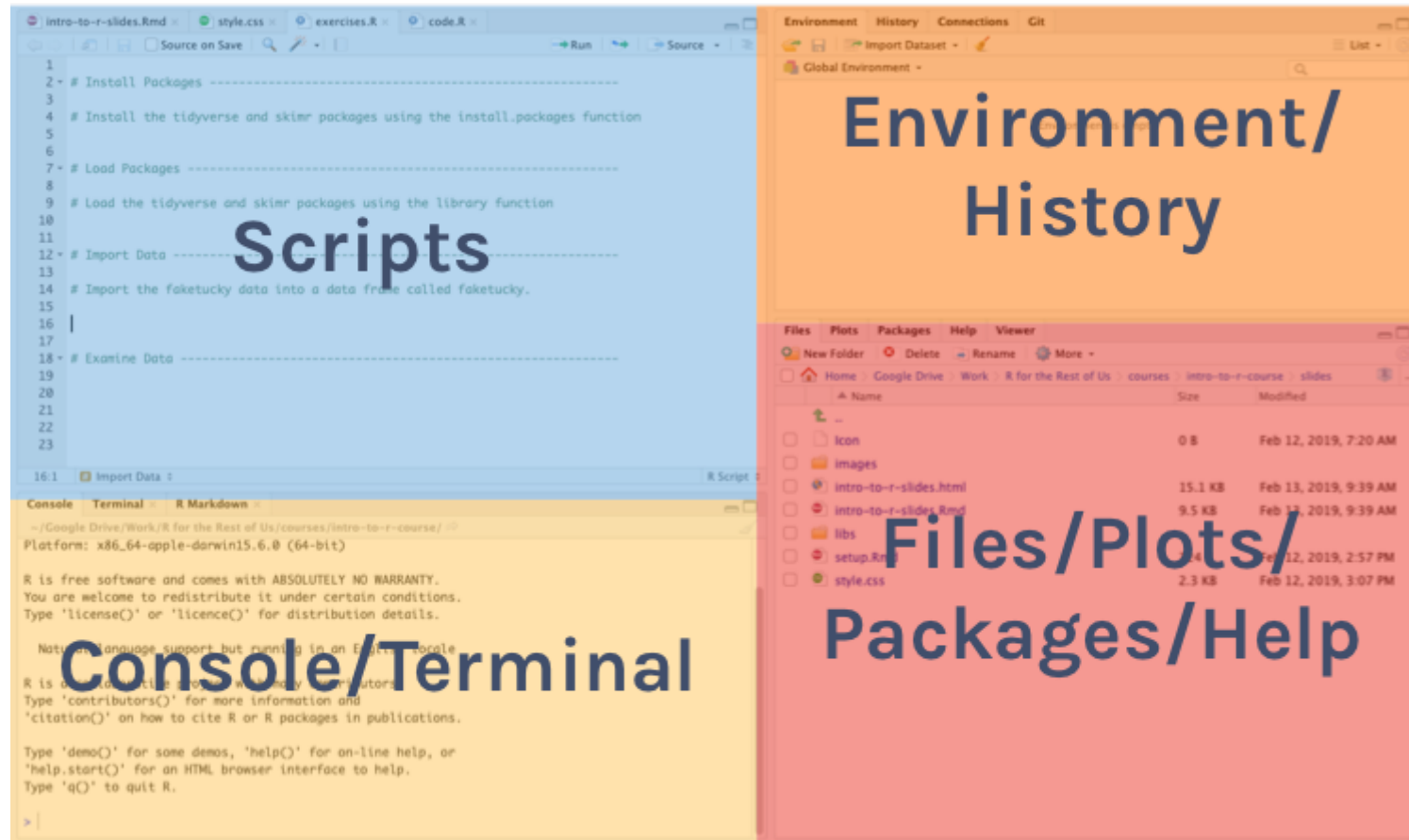


Download RStudio

Download RStudio at the [RStudio website](#). Ignore the various versions listed there. All you need is the latest version of RStudio Desktop.



Tour of RStudio





Packages



Packages

Packages add functionality that is not present in base R.

They're where much of the power of R is found.

R: A new phone	R Packages: Apps you can download
	 

Courtesy [Modern Dive](#)



Packages We'll Use



tidyverse

The [tidyverse](#) is a collection of packages.

We'll use [readr](#) to import data.



Packages We'll Use

skimr

[skimr](#) provides easy summary statistics.





Install Packages

The syntax to install packages is as follows.

```
install.packages("tidyverse")  
install.packages("skimr")
```

The package name must be in quotes.

Packages should be installed **once per computer** (i.e. once you've installed a package, you don't need to do it again on the same computer).



Load Packages

To load packages, use the following syntax:

```
library(tidyverse)  
library(skimr)
```

Package names don't need to be quoted here (though they can be).

Packages should be loaded **once per session** (i.e. every time you start working in R, you need to load any packages you want to use).



Import Data



Import Data

Let's read data from a CSV file.

```
faketucky <- read_csv("data/faketucky.csv")
```

We now have a data frame/tibble called `faketucky` that we can work with in R.



Where Does our Data Live?

Data we have imported is available in the environment/history pane.

The screenshot displays the RStudio interface with four main panes:

- Scripts:** The top-left pane shows an R script with comments and code for installing packages, loading the tidyverse and skimr packages, importing data from faketucky, and examining the data.
- Environment/History:** The top-right pane, highlighted with an orange background, shows the Global Environment and a list of objects.
- Files/Plots/Packages/Help:** The bottom-right pane shows a file explorer view of the current project directory, listing files like icon, images, intro-to-r-slides.html, intro-to-r-slides.Rmd, libs, setup.Rmd, and style.css.
- Console/Terminal:** The bottom-left pane shows the R console output, including the R platform information (x86_64-apple-darwin15.6.0 (64-bit)) and the R license text.



Examine Our Data





Examine Our Data

There are many ways to look at our data. We'll talk about a few.



faketucky

If you type the name of your data frame (i.e. `faketucky`), R will output the following:

```
faketucky
```

```
## # A tibble: 57,855 x 9
##   student_id first_high_scho... school_district gender race_ethnicity
##   <dbl> <chr> <chr> <chr> <chr>
## 1     1622 Jackson Jackson Male Multiple/Nati...
## 2     1877 Jackson Jackson Male White
## 3     1941 Jackson Jackson Male White
## 4     3442 Jackson Jackson Female White
## 5     4623 Jackson Jackson Male White
## 6     4913 Jackson Jackson Male White
## 7     5754 Jackson Jackson Male White
## 8     6293 Jackson Jackson Female White
## 9     7010 Jackson Jackson Male White
## 10    8343 Jackson Jackson Male White
## # ... with 57,845 more rows, and 4 more variables: percent_absent <dbl>,
## #   gpa <dbl>, act_reading_score <dbl>, act_math_score <dbl>
```



View

`View` (note capital V) opens the RStudio viewer (or click on a data frame in the environment pane).

```
View(faketucky)
```



skimr

The skimr package provides more detailed information about our data frame. It is also broken up by the type of variable.

```
skim(faketucky)
```

```
## Skim summary statistics
```

```
##   n obs: 57855
```

```
##   n variables: 9
```

```
##
```

```
## — Variable type:character —
```

	variable	missing	complete	n	min	max	empty	n_unique
--	----------	---------	----------	---	-----	-----	-------	----------

##	first_high_school_attended	0	57855	57855	4	14	0	393
----	----------------------------	---	-------	-------	---	----	---	-----

##	gender	14	57841	57855	4	6	0	2
----	--------	----	-------	-------	---	---	---	---

##	race_ethnicity	794	57061	57855	5	24	0	5
----	----------------	-----	-------	-------	---	----	---	---

##	school_district	0	57855	57855	4	13	0	171
----	-----------------	---	-------	-------	---	----	---	-----

```
##
```

```
## — Variable type:numeric —
```

	variable	missing	complete	n	mean	sd	p0	p25
--	----------	---------	----------	---	------	----	----	-----

##	act_math_score	14101	43754	57855	18.99	4.65	1	16
----	----------------	-------	-------	-------	-------	------	---	----