



# 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

The screenshot displays the RStudio interface with four main panes:

- Scripts**: Shows an R script with code for installing packages, loading tidyverse and skimr, importing faketucky data, and examining data.
- Environment/History**: Shows the Global Environment tab with the title "Environment/History".
- Files/Plots/Packages/Help**: Shows the Files tab with a file list:

Name	Size	Modified
Icon	0 B	Feb 12, 2019, 7:20 AM
images	15.1 KB	Feb 13, 2019, 9:39 AM
intro-to-r-slides.html	9.5 KB	Feb 12, 2019, 9:39 AM
intro-to-r-slides.Rmd		
libs		
setup.Rmd		Feb 12, 2019, 2:57 PM
style.css	2.3 KB	Feb 12, 2019, 3:07 PM

- Console/Terminal**: Shows the R startup message and a prompt for the user.



# 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 shows the RStudio interface with several panes open:

- Scripts**: An R script file containing code for installing packages, loading tidyverse and skimr, importing faketucky data, and examining data.
- Environment/History**: A large orange-paneled window titled "Environment/History". It lists the "Global Environment" with objects like "Icon", "images", "intro-to-r-slides.html", "intro-to-r-slides.Rmd", "libs", "setup.Rmd", and "style.css".
- Console/Terminal**: Shows the R startup message, license information, and a help message about the 'demo()' function.
- Files/Plots/Packages/Help**: A large orange-paneled window titled "Files/Plots/Packages/Help". It lists the "Global Environment" with the same objects as the History pane.



# 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
```



# Getting Help





# ?function

Use the ? to get help about anything you're confused about

```
?read_csv
```



# Tidyverse Website



## Tidyverse packages

### Installation and use

- Install all the packages in the tidyverse by running `install.packages("tidyverse")`.
- Run `library(tidyverse)` to load the core tidyverse and make it available in your current R session.

Learn more about the tidyverse package at <http://tidyverse.tidyverse.org>.



# Package Vignettes

## Using Skimr

*Elin Waring*

**2019-01-13**

`skimr` is designed to provide summary statistics about variables. It is opinionated in its defaults, but easy to modify.

In base R, the most similar functions are `summary()` for vectors and data frames and `fivenum()` for numeric vectors:

```
summary(iris)

##   Sepal.Length   Sepal.Width    Petal.Length    Petal.Width
##   Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##   1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##   Median :5.800   Median :3.000   Median :4.350   Median :1.300
```

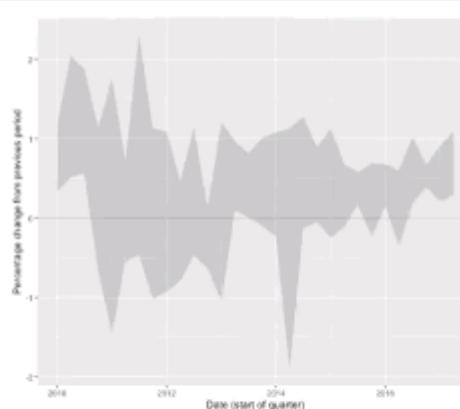


# Twitter

Sam Way and 12 others Retweeted

**Gina Reynolds @EvaMaeRey** · Feb 11  
my **#ggplot2** flipbook project is online! 😎😎😎 Incrementally walks through plotting code (**#MakeoverMonday**, soon **#TidyTuesday** plots). Using **#xaringan** with reveal function; thanks, **@statsgen** **@grrck**. **#rstats**.  
[evamaerey.github.io/ggplot\\_flipboo...](http://evamaerey.github.io/ggplot_flipboo...)

app %>% data %>%  
 select >= 300 & !is.na(.x) & !is.na(.y) &  
 mutate(pct\_change = percentage\_change\_from\_previous\_period) %>%  
 mutate(qtr = quarter(.x)) %>%  
 group\_by(qtr) %>% summarise = p, .by = "qtr") %>%  
 ggplot(aes(x = qtr, y = pct\_change)) +  
 geom\_line()



0:07 12.9K views

18 / 28

23 208 729

Show this thread



# R for Data Science Community

R4DS Online Learning Community

 [HOME](#) [GET INVOLVED](#) [ABOUT](#)

**R FOR DATA SCIENCE  
ONLINE LEARNING  
COMMUNITY**





# Google



E. Kale Edmiston PhD  
@EKaleEdmiston

Follow



A friend/colleague who is an excellent programmer offhandedly told me the other day that coding is 90% googling error messages & 10% writing code. Until this point, I thought that all the time I spent googling error messages meant I was bad at coding. What a perspective change!

8:12 AM - 4 Jan 2019

151 Retweets 1,069 Likes



27

151



1.1K

