

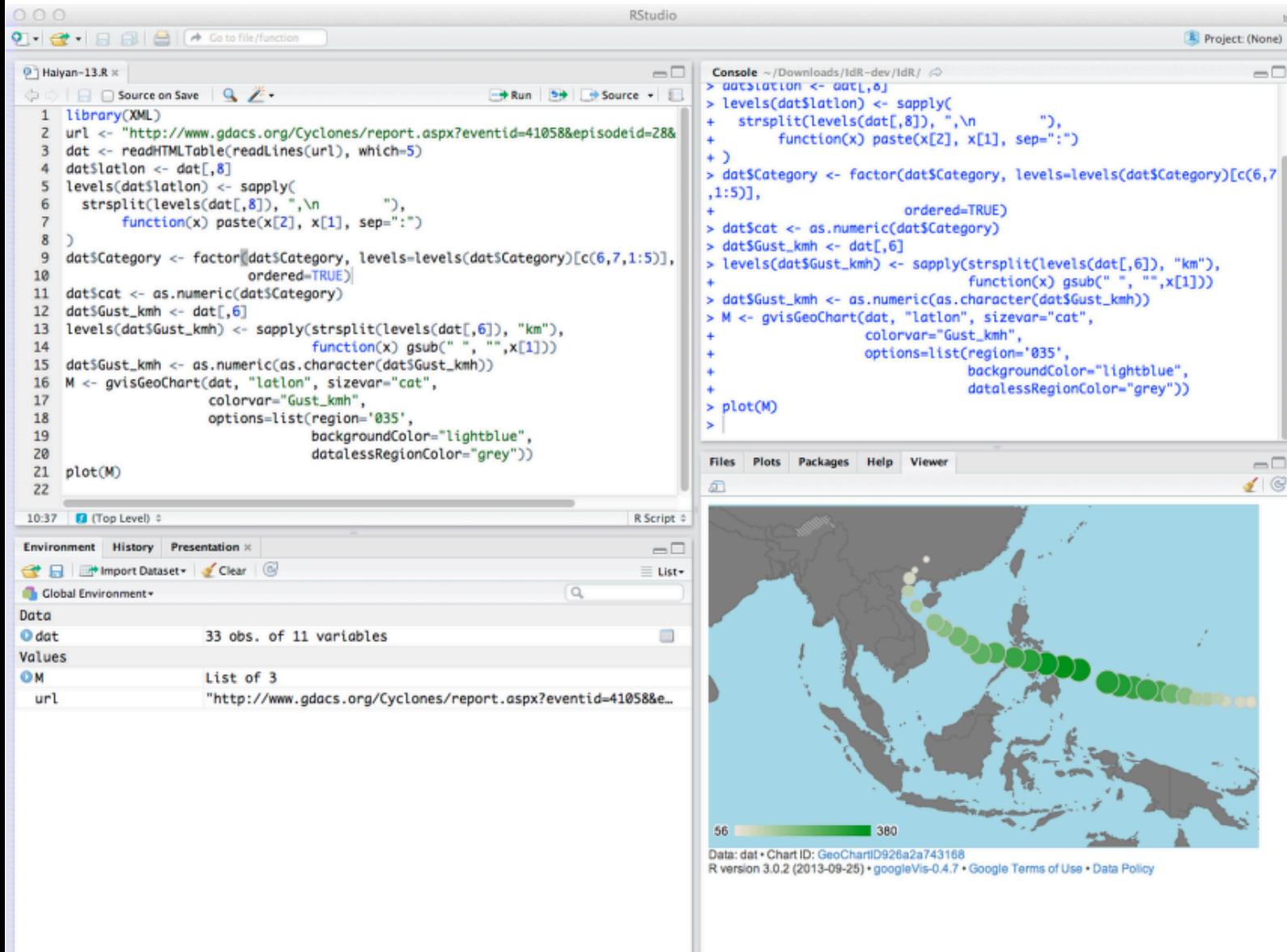
# Tableau, R, & Rstudio

Carolina A. de Lima Salge  
Assistant Professor  
Terry College of Business  
University of Georgia

*Business Intelligence*  
Spring 2021



Terry College of Business  
UNIVERSITY OF GEORGIA



# Tableau

A *paid* software for interactive data visualization

- Easy to learn
- Widely adopted
- Powerful visuals



# Why Tableau?



EDUCATION » Home Colleges Grad Schools Online Colleges Global Universities K-12 SkillBuilder



College Compass » Get instant access to full rankings and complete school data.

Education

## Learn Tableau to Jump-Start Your Data and Visualization Projects: A U.S. News Guide

Learning Tableau could boost your career prospects.

*U.S. News & World Report Education takes an unbiased approach to our recommendations. When you use our links to buy products, we may earn a commission but that in no way affects our editorial independence.*



By Erik Sherman, Contributor Nov. 6, 2020, at 12:36 p.m.



### On This Page

[What Is Tableau?](#)

[Why Tableau?](#)

[Tableau Software and Tableau Products](#)

[Resources for Learning Tableau](#)

[How Long Does It Take to Learn Tableau?](#)

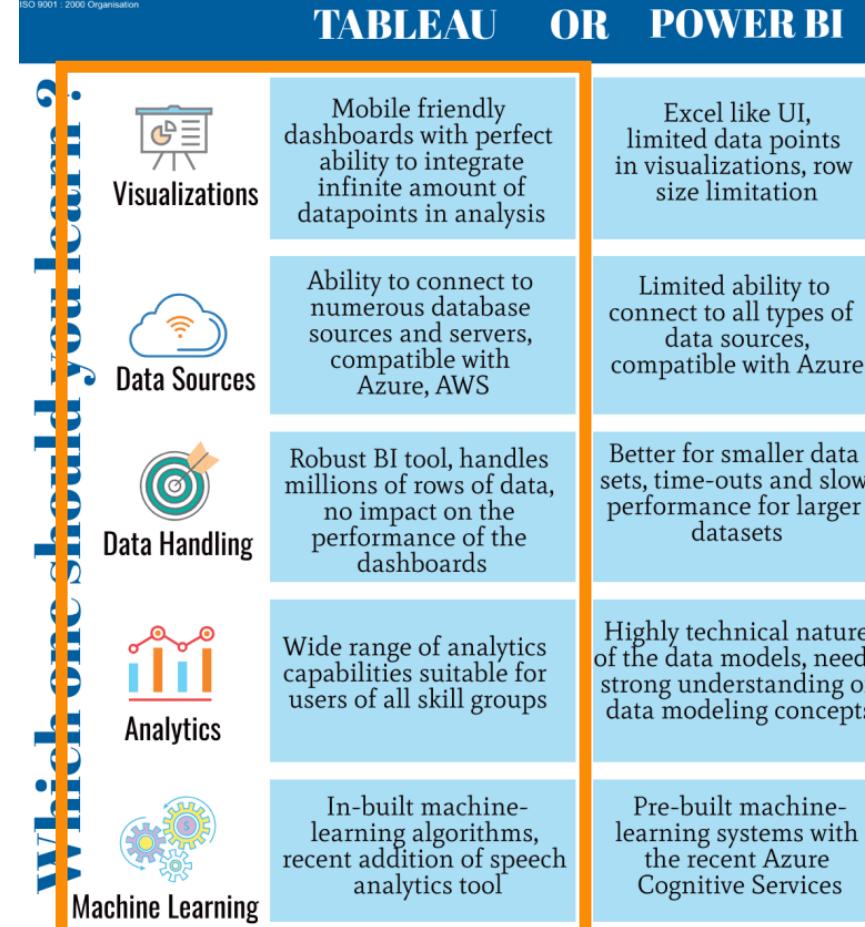
**IN A WORLD OF** data-driven business, learning Tableau could give you a leg up on getting jobs and advancing your career. Whether in marketing, operations, strategic planning, product development, HR or other aspects of business, data helps companies make smarter decisions.

Pulling insights from data and presenting them in ways people understand can be a challenge. Tableau gives you the power to do both with an integrated platform that has been used for nearly 20 years. Use this guide to find resources and gain skills to begin using Tableau.



(COURTESY OF TABLEAU)

### What Is Tableau?



### #Pro Tip

Fresher to BI tools? Go for Tableau with a comparatively easier learning curve, user friendly seamless navigation options, while the same in Power BI? A little uncharted, bulky interface but less expensive. However, in the end it's all about your business needs

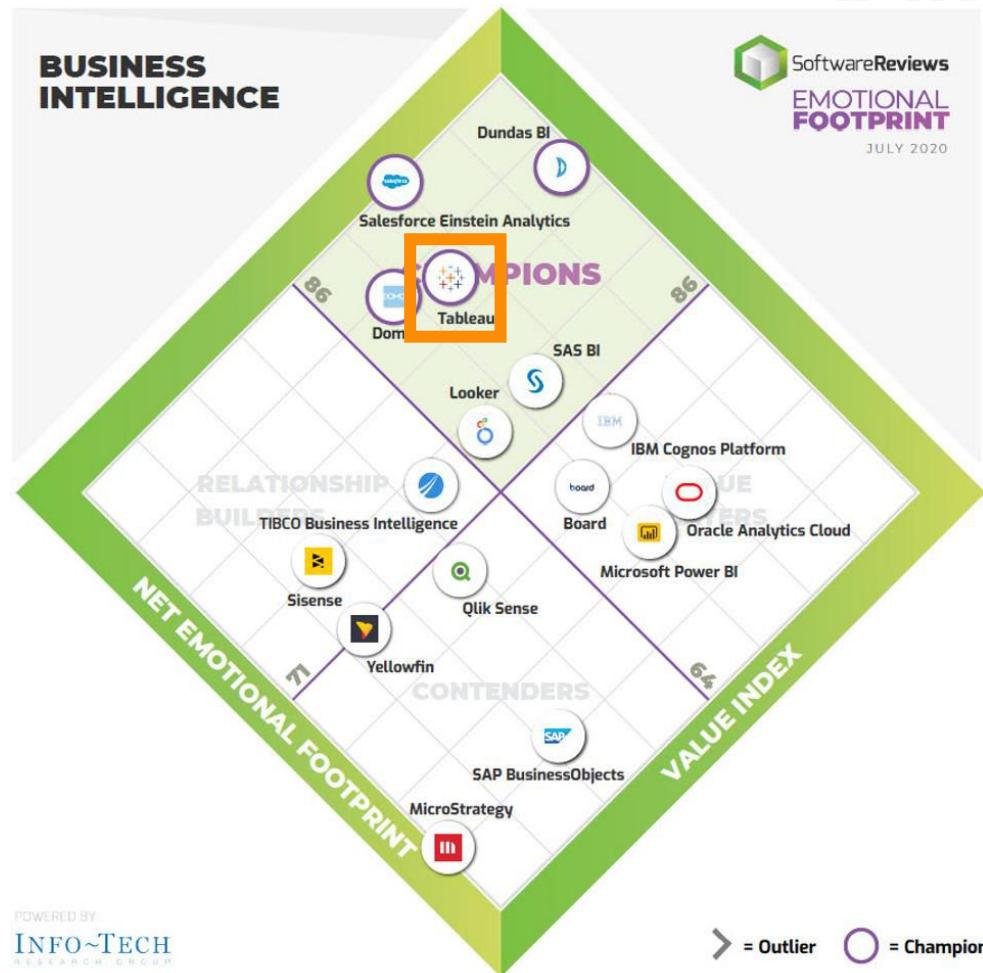


Terry College of Business  
UNIVERSITY OF GEORGIA

<https://ivypf.medium.com/which-one-should-you-learn-tableau-or-power-bi-6e43cef441b>

<https://www.usnews.com/education/learn-tableau-guide>

# Why Tableau?



## BusinessMatters

DING BUSINESS MAGAZINE



Advice Finance Legal Opinion In Business Technology Get Funded Profiles

News: Electric vehicles close to 'tipping point' of mass adoption



MarkLogic DHS  
MarkLogic Data Hub Service

### Why Tableau?

Data experts need tools that are powerful, reliable, user-friendly, and flexible along with many other qualities. Tableau has all the features along with some additional qualities to crave for. It allows experts to clean, process, analyze, and visualize the raw data to actionable insights helping the data expert teams to make timely decisions. Its flexibility allows it to connect to multiple data sources of various kinds simultaneously. This is important as you may have data from various sources in different formats to analyze and prepare consolidated reports. It is highly reliable and provides consistent, real-time output for larger data. It has all sorts of charts, graphs, and tables to choose from to prepare an aesthetic report or dashboard. Charts are very colorful and attractive which help to find hidden patterns, trends in data, and makes it easier to communicate finding to all stakeholders. It allows you to create static or dynamic reports of exceptional qualities.

You need basic data analysis concepts and mathematical knowledge to [learn Tableau](#). It is very user-friendly and you would need to put minimal effort to master it. One of the great features is to create colorful dashboards that present your data with ease on different platforms such as web or mobile. You would only need to spend very little time to create an interactive dashboard. Due to all the qualities of Tableau, various companies are using it, some of the prominent names include Honeywell, LinkedIn, Lufthansa, and Lenovo.

<https://bmmagazine.co.uk/business/tableau-career-options-why-you-should-get-into-tableau/>

<https://www.forbes.com/sites/louiscolombus/2020/08/03/which-bi-systems-are-the-most-popular-with-users/?sh=2f5c68b07f18>



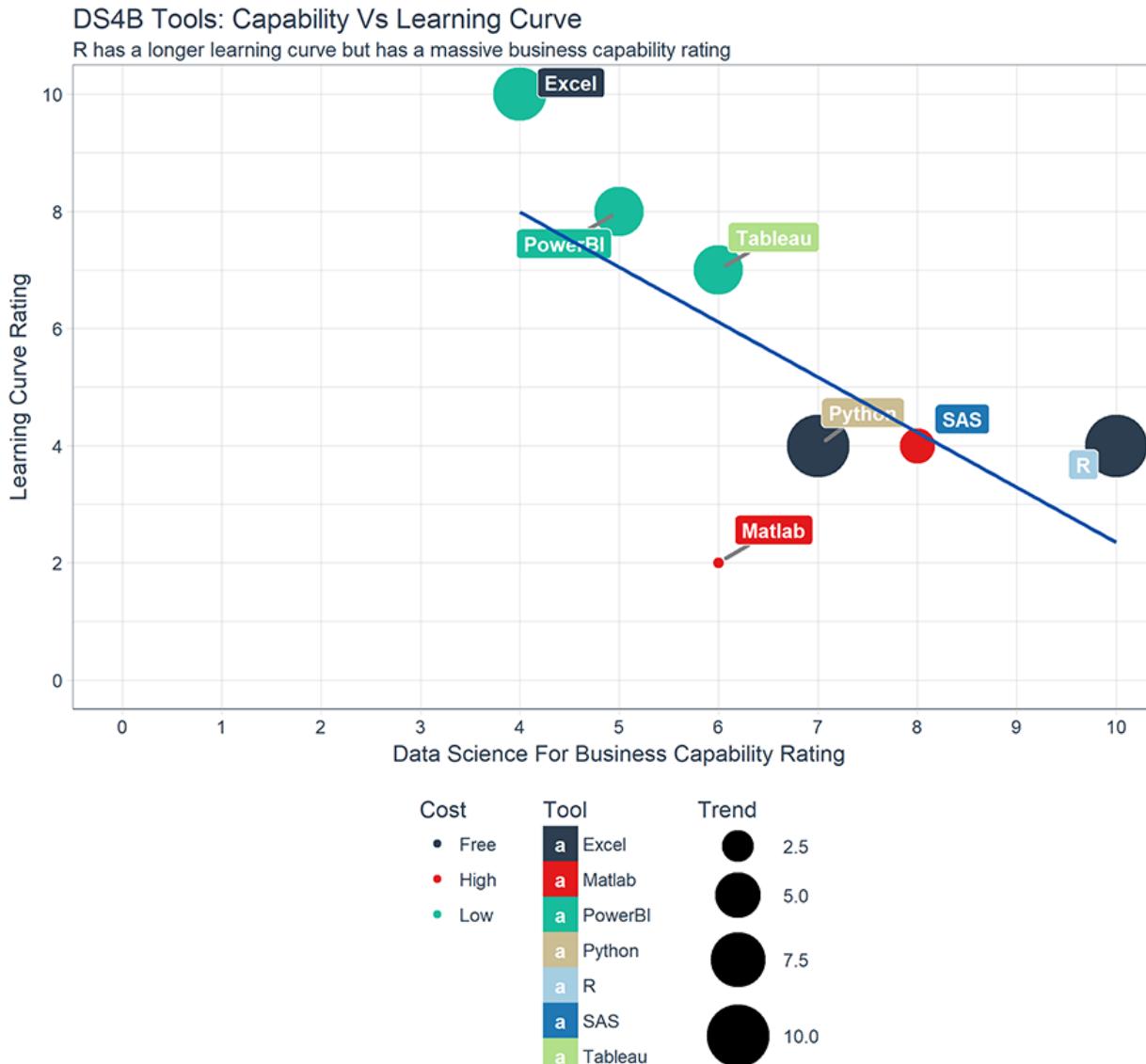
# R

A *free* software environment for statistical computing and graphics

- Object-oriented
- Runs on a wide variety of platforms
- Highly extensible (through packages)



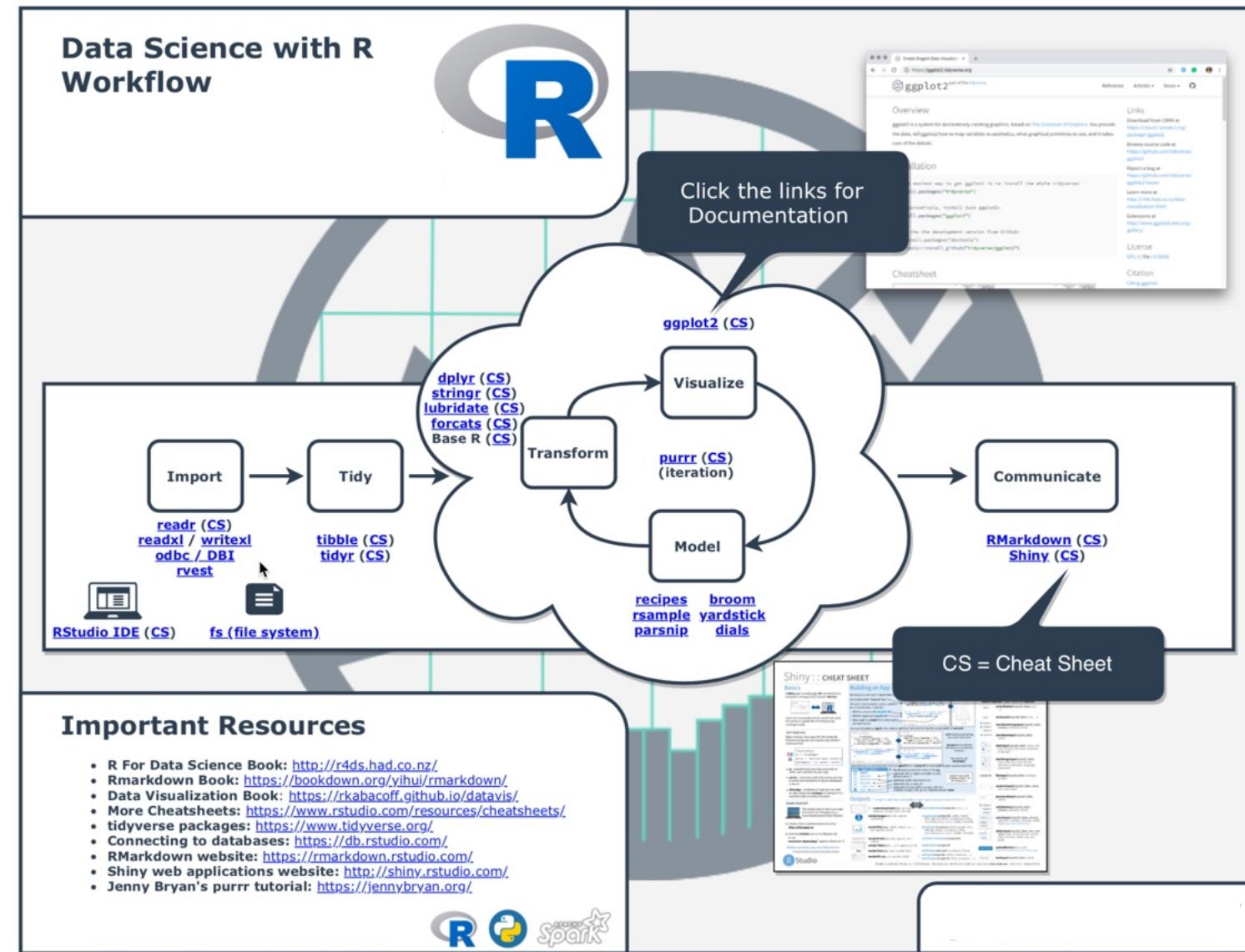
# Why R?



Why R? Tools like Excel, Tableau, PowerBI are easier to learn, but have lower Business Capability. Tools like Python, SAS, and Matlab have high Data Science Capability, but lack the visualization and interactive application tools needed for business. R has the best data science, visualization, and interactive tools plus it's free!



# Why R?



# Why R?

- Are you a computer scientist or software engineer? If yes, learn Python.
- Are you an analytics professional or mechanical/industrial/chemical engineer looking to get into data science? If yes, learn R.

Think about what you are trying to do:

- Are you trying to build a self-driving car? If yes, learn Python.
- Are you trying to communicate business analytics throughout your organization? If yes, learn R.



## Popularity on Stack Overflow, Python vs. R



Source: DZone

# A Note Here

*R and Python > R or Python*



**“R is currently head-and-shoulders above Python for data analysis, but I remain convinced that Python can catch up, easily and quickly.”**

JAN GALKOWSKI, COMPUTATIONAL ENGINEER



Terry College of Business  
UNIVERSITY OF GEORGIA

<https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis>

# Why R?

Learning R used to be a major challenge. Base R was a complex and inconsistent programming language. Structure and formality was not the top priority as in other programming languages. This all changed with the “**tidyverse**”, a set of packages and tools that have a consistently structured programming interface.

When tools such as `dplyr` and `ggplot2` came to fruition, it made the learning curve much easier by providing a consistent and structured approach to working with data. As **Hadley Wickham** and many others continued to evolve R, the `tidyverse` came to be, which includes a series of commonly used packages for data manipulation, visualization, iteration, modeling, and communication. The end result is that R is now much easier to learn - [Learn R From A Master Data Scientist's Code](#).

```
> library(tidyverse)
— Attaching packages —                               tidyverse 1.2.1 —
✓ ggplot2 2.2.1.9000    ✓ purrr   0.2.4
✓ tibble  1.3.4        ✓ dplyr   0.7.4
✓ tidyr   0.7.2        ✓ stringr 1.2.0
✓ readr   1.1.1        ✓ forcats 0.2.0
— Conflicts —                                     tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()   masks stats::lag()
> |
```



# Why R?

## R has brains

R implements cutting-edge algorithms including:

- **H2O** (`h2o`) - High-end machine learning package
- **Keras/TensorFlow** (`keras`, `tensorflow`) - Go-to deep learning packages
- **xgboost** - Top Kaggle algorithm
- **Modeltime** - Time Series forecasting
- And many more!

## R has muscle

R has powerful tools for:

- Vectorized Operations - R uses vectorized operations to make math computations lightning fast right out of the box
- Loops (`purrr`)
- Parallelizing operations (`parallel`, `future`)
- Speeding up code using C++ (`Rcpp`)
- Connecting to other languages (`rJava`, `reticulate`)
- Working With Databases - **Connecting to databases** (`dbplyr`, `odbc`, `bigrquery`)
- Handling Big Data - **Connecting to Apache Spark** (`sparklyr`)
- And many more!

## R has heart

We already talked about the infrastructure, the **tidyverse**, that enables the ecosystem of applications to be built using a consistent approach. It's this infrastructure that brings life into your data analysis. The **tidyverse** enables:

- Data manipulation (`dplyr`, `tidyr`)
- Working with data types (`stringr` for strings, `lubridate` for date/datetime, `forcats` for categorical/factors)
- Visualization (`ggplot2`)
- Programming (`purrr`, `tidyeval`)
- Communication (`Rmarkdown`, `shiny`)



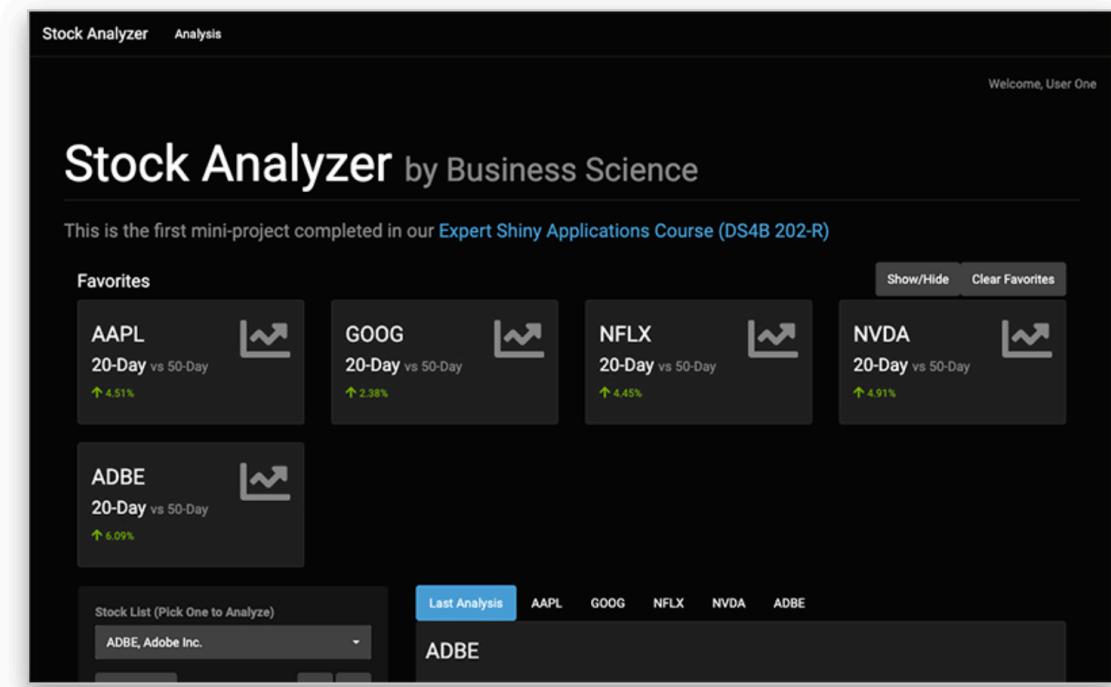
# Why R?

## Rmarkdown

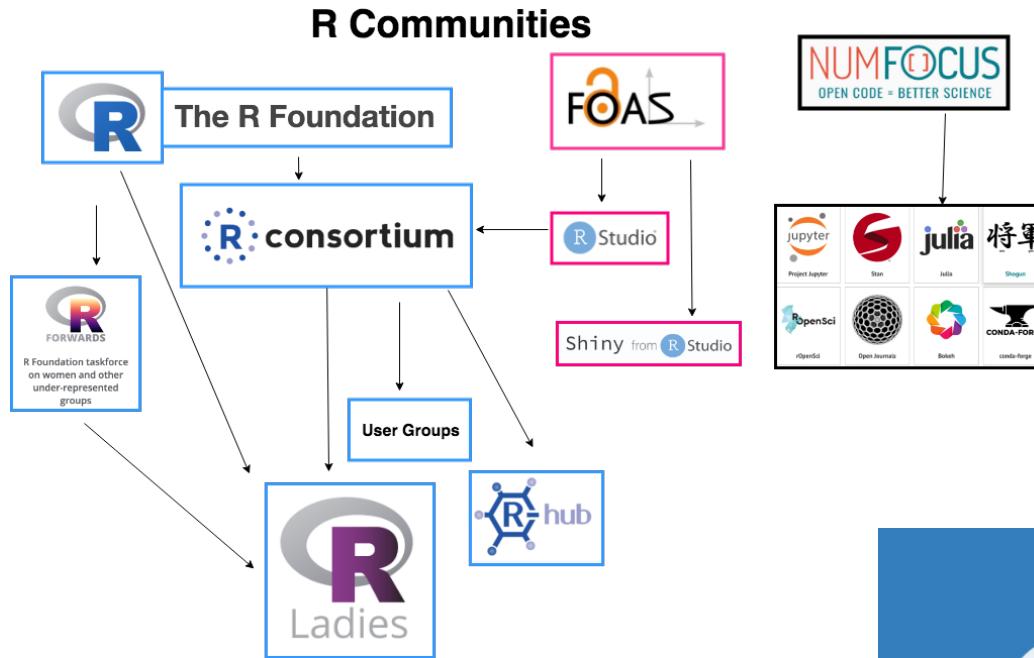
Rmarkdown is a framework for creating reproducible reports that has since been extended to building blogs, presentations, websites, books, journals, and more. It's the technology that's behind this blog, and it allows us to include the code with the text so that anyone can follow the analysis and see the output right with the explanation. What's really cool is that the technology has evolved so much. Here are a few examples of its capability:

- [rmarkdown](#) for generating HTML, Word and PDF reports
- [rmarkdown](#) for generating presentations
- [flexdashboard](#) for creating web apps via the user-friendly Rmarkdown format.
- [blogdown](#) for building blogs and websites
- [bookdown](#) for creating online books
- [Interactive documents](#)
- [Parameterized reports](#) for generating custom reports (e.g. reports for a specific geographic segment, department, or segment of time)

## Shiny



# Why R?



## R User Community

From: The R Ecosystem  
[bit.ly/R-ecosystem](http://bit.ly/R-ecosystem)



Local R User Groups (55)  
<http://blog.revolutionanalytics.com/local-r-groups.html>



100 repositories  
<https://github.com/languages/R/created>

crantastic!

inside-R

R users (2M+)

Quora

stackoverflow  
6000 Questions

Revolutions  
News about R statistics and the world of open source from the staff of Revolution Analytics  
<http://blog.revolutionanalytics.com>

R-bloggers  
R news and tutorials contributed by (231) R bloggers

twitter  
#rstats  
<http://bit.ly/r-twitter>

REVOLUTION ANALYTICS



# RStudio

Scripts

The screenshot shows the RStudio interface. In the top-left, there's a code editor window titled "RStudio.R" containing R code. In the top-right, the "Environment" tab of the global environment pane shows two objects: "CoffeeChain" (4248 obs. of 19 variables) and "df" (20 obs. of 2 variables). Below the code editor is the "Console" tab, which displays the R session history. To the right of the console is a "Plots" tab showing a horizontal bar chart titled "reorder(State, -Count)". The x-axis is labeled "Count" and ranges from 0 to 300. The y-axis lists states from Massachusetts at the top to California at the bottom. The bars show the count for each state, with California having the highest count (around 280) and Massachusetts the lowest (around 150).

```
library(tidyverse)
library(readxl)
CoffeeChain <- read_excel("Dropbox/BAN6003_2018 3/Datasets/CoffeeChain.xlsx")

library(sqldf)
df <- sqldf("select State, count(*) as `Count` from CoffeeChain group by State order by Count desc;")

ggplot(df, aes(reorder(State, -Count), Count)) + geom_bar(stat="identity") + coord_flip()
```

```
> library(sqldf)
Loading required package: gsubfn
Loading required package: proto
xcrun: error: invalid active developer path (/Library/Developer/CommandLineTools), missing xcrun at: /Library/Developer/CommandLineTools/usr/bin/xcrun
Loading required package: RSQLite
Warning messages:
1: package 'sqldf' was built under R version 4.0.2
2: package 'gsubfn' was built under R version 4.0.2
3: package 'proto' was built under R version 4.0.2
4: In system2("/usr/bin/otool", c("-l", shQuote(OSO)), stdout = TRUE) :
  running command '/usr/bin/otool' -L '/Library/Frameworks/R.framework/Resources/library/tcltk/lib //tcltk.so' had status 1
5: package 'RSQLite' was built under R version 4.0.2
> df <- sqldf("select State, count(*) as `Count` from CoffeeChain group by State order by Count desc;")
> ggplot(df, aes(reorder(State, -Count), Count)) + geom_bar(stat="identity") + coord_flip()
>
```

Datasets

Results

Files, plots, packages, & help



# Why RStudio?



```
R version 4.0.1 (2020-06-06) -- "See Things Now"  
Copyright (C) 2020 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 1.72 (7845) x86_64-apple-darwin17.0]
```

```
[Workspace restored from /Users/carolinaalvesdelimasalge/.RData]  
[History restored from /Users/carolinaalvesdelimasalge/.Rapp.history]
```

```
>
```



# Packages

Collections of functions and data sets developed by the community.  
They increase the power of R by improving existing base R  
functionalities, or by adding new ones

- Over 4,000 packages
- Stored in repositories ([CRAN](#) is the official)
- Most people *use* packages, few *write* packages



# Packages

```
# install ONCE on your computer  
# can also use Rstudio to install
```

```
install.packages("packageName")
```

```
# require EVERY TIME before using a package in a session  
# loads the package to memory
```

```
require(packageName)  
library(packageName)
```



# Tidyverse

A collection of R packages designed for data science

- All packages share an understanding design philosophy, grammar, and data structures

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



# Why Tidyverse?

- Consistency
- Comprehensive
- Community

```
> library(tidyverse)
— Attaching packages ————— tidyverse 1.2.1 —
✓ ggplot2 2.2.1.9000   ✓ purrr  0.2.4
✓ tibble  1.3.4        ✓ dplyr   0.7.4
✓ tidyr   0.7.2        ✓ stringr 1.2.0
✓ readr   1.1.1        ✓ forcats 0.2.0
— Conflicts ——————— tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()   masks stats::lag()
> |
```



# Why Tidyverse?

## Gather, Store, Access

- Read (readr, readxl)
- Scrape (rvest)
- Database (DBI)
- Web APIs (httr)
- XML (xml2)
- JSON (jsonlite)

## Analyze

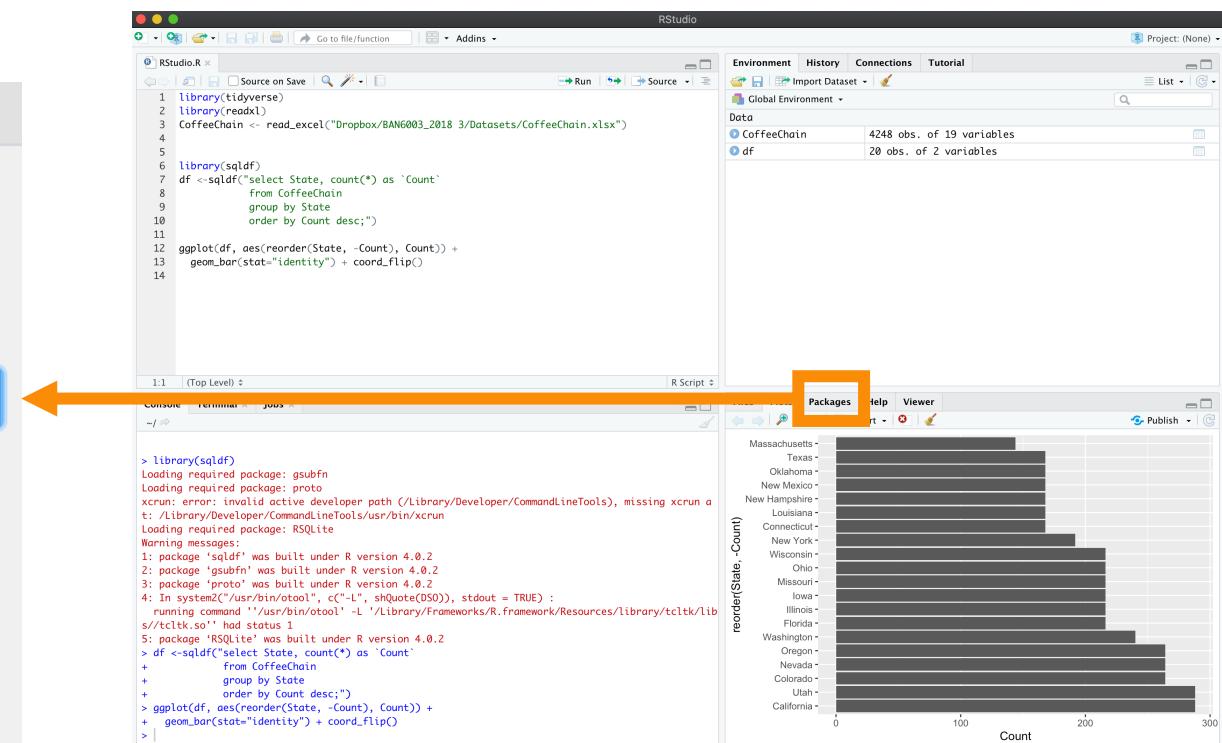
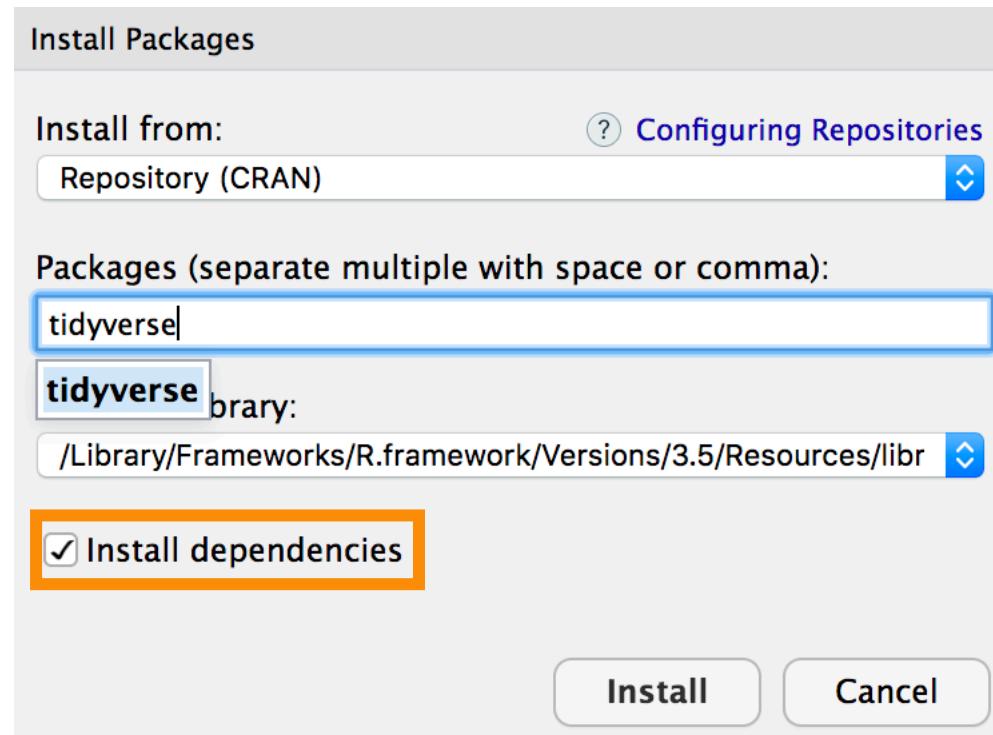
- Visualize (ggplot2)
- Transform (dplyr)
- Wrangle (tidyR)
- Model (modelr)



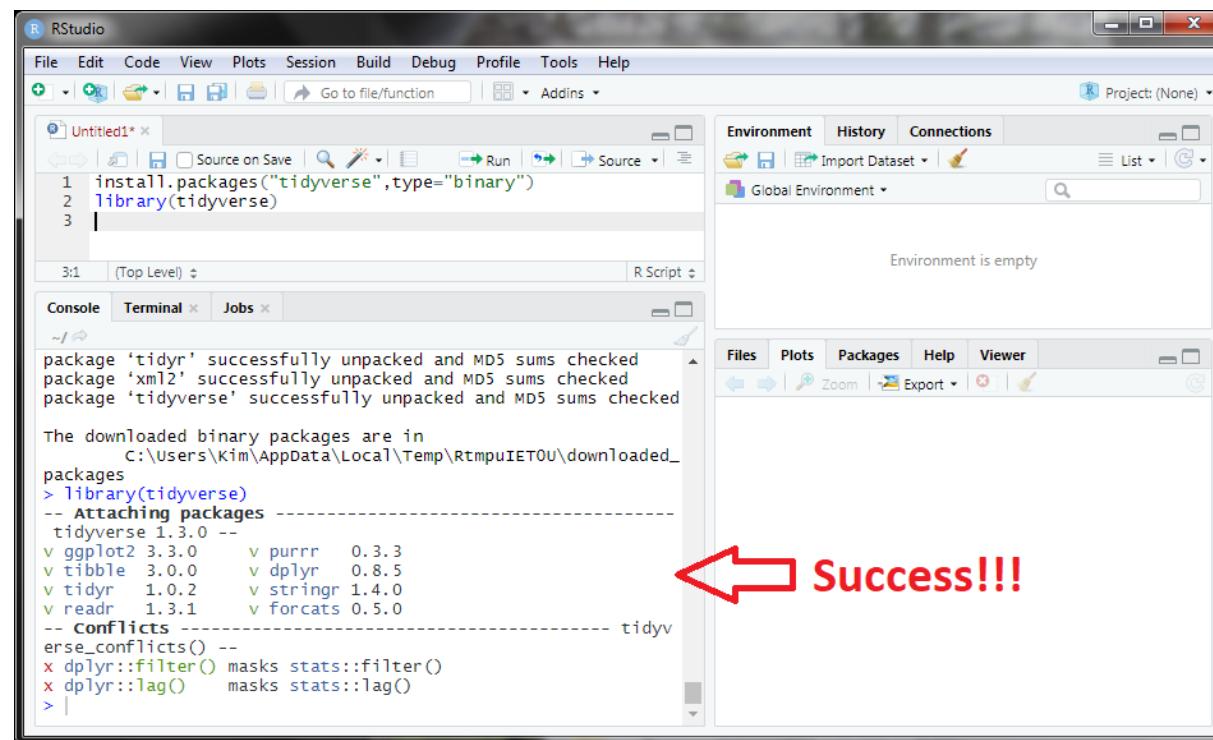
```
> library(tidyverse)
— Attaching packages ——————— tidyverse 1.2.1 —
✓ ggplot2 2.2.1.9000   ✓ purrr  0.2.4
✓ tibble  1.3.4        ✓ dplyr  0.7.4
✓ tidyR   0.7.2        ✓ stringr 1.2.0
✓ readr   1.1.1        ✓ forcats 0.2.0
— Conflicts ——————— tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()   masks stats::lag()
> |
```



# Install Tidyverse



# Install Tidyverse



A screenshot of the RStudio interface. The left pane shows an R script named 'Untitled1' with the following code:

```
install.packages("tidyverse", type="binary")
library(tidyverse)
```

The right pane shows the Environment tab of the Global Environment panel, which is currently empty.

Below the script pane, the Console tab displays the output of the package installation:

```
package 'tidyverse' successfully unpacked and MD5 sums checked
package 'xml2' successfully unpacked and MD5 sums checked
package 'tidyverse' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
  C:\Users\Kim\AppData\Local\Temp\RtmpUIET0U\downloaded_
packages
> library(tidyverse)
-- Attaching packages --
tidyverse 1.3.0 --
v ggplot2 3.3.0   v purrr  0.3.3
v tibble  3.0.0   v dplyr   0.8.5
v tidyverse 1.0.2  v stringr 1.4.0
v readr    1.3.1   v forcats 0.5.0
-- Conflicts --
x dplyr::filter() masks stats::filter()
x dplyr::lag()   masks stats::lag()
```

A red arrow points from the word "Success!!!" to the line "package 'tidyverse' successfully unpacked and MD5 sums checked".

```
> install.packages("tidyverse")
Installing package into 'C:/Users/laerd/Documents/R/win-library/3.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/tidyverse_1.2.1.zip'
Content type 'application/zip' length 92614 bytes (90 KB)
downloaded 90 KB
```

package 'tidyverse' successfully unpacked and MD5 sums checked

The downloaded binary packages are in  
C:\Users\laerd\AppData\Local\Temp\Rtmpqi5NcT\downloaded\_packages



# *Thank You!*



Terry College of Business  
UNIVERSITY OF GEORGIA