15.1.1 Introduction to R

Jeremy is ecstatic that he has been given the opportunity to lead the data analytics team. He's confident that his 10 years working with the manufacturing and research team has provided him with sufficient expertise in the subject matter. However, he is much less confident about his statistics background and his programming ability. He took some stats and programming courses in college, but it has been a long time since he had to think about either subject.

But Jeremy has never been one to walk away from a challenge. He knows that if he fully commits to learning stats and R programming, he'll feel comfortable in no time! With his previous experience in programming, learning R seems to be a good starting point to prepare for his new role.

R is a programming language that has a variety of uses in data science. R has solidified itself in academia and industry as one of the go-to programming languages for statistical modelling and hypothesis testing. In recent years, R developers have extended R's capabilities to generate machine learning algorithms and other advanced models to ensure that R can be used in every stage of data analytics.

Benefits of R

R is a versatile and extensible programming language with many benefits. One of the benefits of using an interpreted programming language such as R (or Python) is that the analysis scripts are written in plaintext. The versions of plaintext files are easy to control using tools such as Git, which means that **R analysis scripts** (or **RScripts**) are highly reproducible and easy to share with peers and collaborators.

Another benefit to using R is that the R programming language was specifically designed for data analysis. This means that the process of loading in a dataset, visualizing the data, and performing statistical tests is straightforward and easy to interpret. In fact, many of the statistical tests in Python have been directly ported from R due to how well they were implemented. In addition to the native statistical functions, there are many other useful data transformation and modelling libraries, such as the tidyverse package, that simplify the process of ETL and visualizations.

Drawbacks of R

Still, R is not perfect. The biggest drawback is its licensing. R and most of R's libraries are licensed as General Public License, version 2 (GPL 2). This means that if you program or model anything using R, GPL forces your application, program, or script to be open source.

In many personal and academic uses, this is not a problem because you're either (a) not trying to monetize your program or (b) going to publish your analysis and findings. However, if you are working for a company with intellectual property, or proprietary data and programs, this can be an issue. Therefore, many companies use R for internal analysis and regulatory testing, but use Python for any application or script that contains proprietary information.

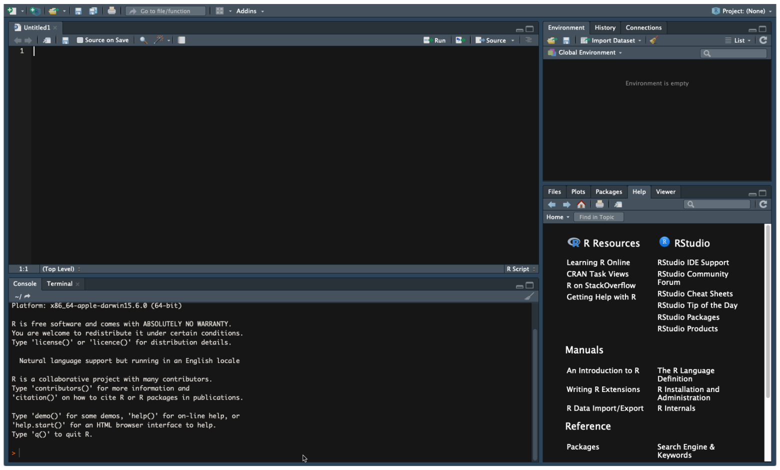
Despite the licensing drawback, R is still a highly valuable programming language for data analysis and is used by data professionals at all levels across many fields.

RStudio Integrated Development Environment

Just as Jupyter Notebooks are an integrated development environment (IDE) used to help design and test Python scripts, RStudio is an IDE used to help design and test RScripts. RStudio provides users a graphical user interface (GUI) with multiple dynamic windows and perpetual access to their RScript and R console.

Similar to Jupyter Notebooks, RStudio enables users to test their analysis scripts line by line while allowing users to view different environment variables and outputs. This means that for each line of code written and executed, users can verify the results and troubleshoot any problems quickly and easily.

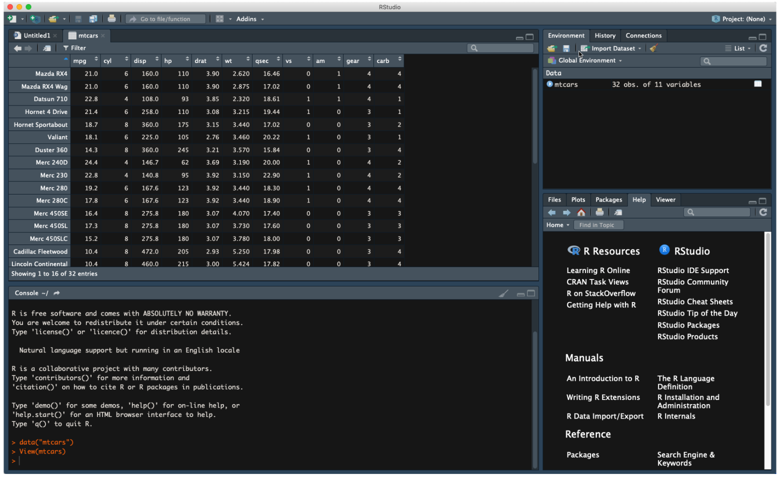
The following image shows what an empty session in RStudio will look like:



Environment is empty

```
Platform: x86_64-apple-darwin15.6.0 (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.

> 
```

And here's what an active RStudio looks like. Note the DataFrame in the top left pane and the mtcars object in the top right pane:



	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4

Showing 1 to 16 of 32 entries

Data
mtcars 32 obs. of 11 variables

```
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.

> data("mtcars")
> View(mtcars)
> 
```

Now that you know what R and RStudio do, let's install them on your machine.