

main.py

1  
2  
3  
4  
5  
6  
7  
8  
9

# CODE WITH REPL.IT

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

FROM BEGINNER TO EXPERT  
THROUGH GUIDED TUTORIALS



# **Programming walkthroughs: Coding with Python and Repl.it**

Gareth Dwyer

© 2020 Gareth Dwyer

# Contents

<b>Welcome to Code with Repl.it . . . . .</b>	<b>1</b>
Part 1: The basics of Repl.it . . . . .	1
Part 2: advanced Repl.it use . . . . .	2
Part 3: building your own projects . . . . .	3
<b>Understanding the Repl.it IDE: a practical guide to building your first project with Repl.it . . . . .</b>	<b>5</b>
Introduction: creating an account and starting a project . . . . .	5
Adding more files to your software project . . . . .	9
Sharing your application with others . . . . .	12
Sharing write-access: Multiplayer . . . . .	13
Make it your own . . . . .	15
Where next? . . . . .	15
<b>Working with Files using Repl.it . . . . .</b>	<b>16</b>
Working with files using Python . . . . .	16
Creating files using Python . . . . .	17
Building a weather logging system using Python and Repl.it . . . . .	18
Exporting our weather data files . . . . .	23
Make it your own . . . . .	25
Where next? . . . . .	25
<b>Managing dependencies using Repl.it . . . . .</b>	<b>26</b>
Understanding Repl.it's magic import tool and the universal package manager . . . . .	26
Installing packages through the GUI . . . . .	27
Building an NLP project using spaCy . . . . .	29
Make it your own . . . . .	36
Where next? . . . . .	36
<b>Data science with Repl.it: Plots and graphs . . . . .</b>	<b>37</b>
Installing Matplotlib and creating a basic line plot . . . . .	37
Making a scatter plot of US cities by state . . . . .	38
More advanced plotting with seaborn and pandas . . . . .	40
Saving plots to PNG files . . . . .	42
Make it your own . . . . .	43
Where next? . . . . .	44

## CONTENTS

<b>Multiplayer: Pair programming with Repl.it . . . . .</b>	<b>45</b>
Extending our data science article using pair programming: Getting help . . . . .	45
Make it your own . . . . .	52
Where next? . . . . .	52
<b>Repl.it and GitHub: Using and contributing to open-source projects . . . . .</b>	<b>54</b>
Importing a project from GitHub and running it on Repl.it . . . . .	54
Looking at the version control panel in Repl.it . . . . .	57
Forking the project to your own GitHub account . . . . .	59
Make it your own . . . . .	62
Where next? . . . . .	63
<b>Building a game with PyGame and Repl.it . . . . .</b>	<b>64</b>
Creating a PyGame repl . . . . .	64
Displaying the sprite using PyGame . . . . .	66
Making our tennis ball move with each frame . . . . .	68
Processing events: Detecting mouse clicks . . . . .	69
Making the ball bounce off the edges and move randomly . . . . .	70
Adding more balls . . . . .	72
Make it your own . . . . .	76
Where next? . . . . .	76
<b>Staying safe: Keeping your passwords and other secrets secure . . . . .</b>	<b>77</b>
Understanding .env files . . . . .	77
Refactoring our weather project to keep our API key secure . . . . .	78
Testing that the file is not copied into others' forks . . . . .	79
Using environment variables in our script . . . . .	80
Time travelling to find secrets . . . . .	81
Rotating credentials . . . . .	82
Make it your own . . . . .	83
Where next? . . . . .	83
<b>An introduction to pytest and doing test-driven development with Repl.it . . . . .</b>	<b>84</b>
Creating a project structure for pytest . . . . .	84
Defining examples for the name split function . . . . .	85
Writing the test cases for our names function . . . . .	86
Fixing our split_name function . . . . .	90
Using our function . . . . .	93
Make it your own . . . . .	93
Where next . . . . .	94
<b>Productivity hacks . . . . .</b>	<b>95</b>
Using the command palette . . . . .	95
Duplicating entire lines . . . . .	96

## CONTENTS

Deleting entire lines . . . . .	97
Inserting blank lines . . . . .	97
Indenting and dedenting lines . . . . .	98
Adding cursors . . . . .	98
Navigating to specific pieces of code . . . . .	99
Vim and Emacs key bindings . . . . .	100
Make it your own . . . . .	100
Where next? . . . . .	101
<b>Beginner web scraping with Python and Repl.it . . . . .</b>	<b>102</b>
Overview and requirements . . . . .	102
Webpages: beauty and the beast . . . . .	103
Downloading a web page with Python . . . . .	107
Using BeautifulSoup to extract all URLs . . . . .	109
Fetching all of the articles from the homepage . . . . .	113
Where next? . . . . .	116
<b>Building news word clouds using Python and Repl.it . . . . .</b>	<b>118</b>
Overview and requirements . . . . .	118
Web scraping . . . . .	119
Taking a look at RSS Feeds . . . . .	119
Setting up our online environment (Repl.it) . . . . .	121
Pulling data from our feed and extracting URLs . . . . .	121
Setting up a web application with Flask . . . . .	122
Downloading articles and extracting the text . . . . .	124
Returning HTML instead of plain text to the user . . . . .	126
Generating word clouds from text in Python . . . . .	129
Adding some finishing touches . . . . .	131
Where next? . . . . .	133
<b>Building a Discord Bot with Python and Repl.it . . . . .</b>	<b>135</b>
Overview and requirements . . . . .	135
Creating a Repl and installing our Discord dependencies . . . . .	139
Setting up authorization for our bot . . . . .	139
Keeping our bot alive . . . . .	145
Forking and extending our basic bot . . . . .	147
<b>Building a Discord bot with Node.js and Repl.it . . . . .</b>	<b>149</b>
Overview and requirements . . . . .	149
Creating a Repl and installing our Discord dependencies . . . . .	153
Setting up authorization for our bot . . . . .	154
Keeping our bot alive . . . . .	159
Forking and extending our basic bot . . . . .	161

## CONTENTS

<b>Creating and hosting a basic web application with Django and Repl.it . . . . .</b>	<b>162</b>
Setting up . . . . .	163
Changing our static content . . . . .	164
Calling IPIFY from JavaScript . . . . .	165
Adding a new route and view, and passing data . . . . .	166
Calling a Django route using Ajax and jQuery . . . . .	169
Using ip-api.com for geolocation . . . . .	171
Getting weather data from OpenWeatherMap . . . . .	173
<b>Building a CRM app with NodeJS, Repl.it, and MongoDB . . . . .</b>	<b>178</b>
Overview and requirements . . . . .	178
Creating a Repl and connecting to our Database . . . . .	180
Making a user interface to insert customer data . . . . .	182
Updating and deleting database entries . . . . .	187
Putting it all together . . . . .	193
<b>Introduction to Machine Learning with Python and Repl.it . . . . .</b>	<b>194</b>
Overview and requirements . . . . .	194
Setting up . . . . .	194
Creating some mock data . . . . .	195
Understanding vectorization . . . . .	196
Understanding classification . . . . .	199
Building a manual classifier . . . . .	202
<b>Quicksort tutorial: Python implementation with line by line explanation . . . . .</b>	<b>206</b>
Overview and requirements . . . . .	206
The Partition algorithm . . . . .	207
The Quicksort function . . . . .	211
Testing our algorithm . . . . .	211
<b>Closing note . . . . .</b>	<b>214</b>

# Welcome to Code with Repl.it

In a series of tutorials, you'll go from beginner to expert in coding with Repl.it. While these lessons are designed to be taken in order, they each make sense on their own too, so feel free to jump in wherever looks the most interesting to you.

Part 1 covers the basics of using Repl.it: how to create projects, work with files, use third party dependencies, do plotting and graphing, and use multiplayer to code as part of a team.

Part 2 covers more advanced Repl.it use. You'll see how to pull projects from GitHub and collaborate on open source software, build a game, and keep your code secure. You'll build a full web application using Test Driven Development (TDD), and find out how to be an elite hacker by using the shortcuts offered by Repl.it

Once you've completed Part 1 and Part 2, you'll be able to build nearly any project that you want, and deploy it for the world to use. If you're stuck for ideas, you can go through the examples given in Part 3, which consists of practical tutorials to build everything from web scrapers to chat bots.

Note that this set of lessons does not focus on teaching you to code, though we will explain some key concepts along the way. If you don't already know how to code, it's best to take this course in conjunction with a more traditional course. If you're not sure what to do next, jump right in and see if you can keep up. We're beginner friendly.

## Part 1: The basics of Repl.it

In this section of the course, you'll learn the basic of Repl.it. But that doesn't mean you won't build some fun stuff along the way.

### Tutorial 1: Introduction to Repl.it and using the IDE

Learn the basics of the Repl.it IDE. Why use an online IDE and what are all those different panes? Build a simple program to solve your maths homework.

### Tutorial 2: Working with files using Repl.it

Computers were initially created to read and write files, and although we've come a long way files remain central to everything we do. Learn how to create them, read from them, write to them, and import and export them in bulk.

## Tutorial 3: Managing dependencies with Repl.it

No one is an island, and if you build software you'll build it on top of existing modules that others have written. Here we show you how to work with other people's code in a variety of ways: in many cases all you need to do is [import antigravity and fly away](#)<sup>1</sup>.

## Tutorial 4: Data science: plotting and graphing

Data is only useful if it can be easily understood. Plots, charts, and graphs are the easiest way to know what's happening in the world around you. And did you know that data science is [the sexiest job of the 21st century](#)<sup>2</sup>. Follow along to plot every city in the USA and find out if richer people live longer.

## Tutorial 5: Pair programming and using multiplayer

Did we mention that no one is an island? Coders don't have to work alone. You can invite your friends to code along with you, a technique used by beginners and experts alike. Learn how to code collaboratively, as if you were using a Google Doc.

# Part 2: advanced Repl.it use

Once you know the basics, it's time to build larger and more complicated projects and keep them secure.

## Tutorial 6: Running projects from GitHub

Most open source software lives on GitHub and it's easy to take advantage of all of this free software by pulling code from GitHub to Repl.it and running it with one click. Some software needs to be configured in specific ways so you'll also learn how to modify what happens when you press that big green "run" button.

## Tutorial 7: Building a game with PyGame

Do you want to develop games? Of course, you can do that with Repl.it too. We'll build a 2D juggling game using PyGame in this lesson and you'll learn more about graphics programming at the same time: sprites, physics, and more.

## Tutorial 8: Can you keep a secret? What about from time travellers?

Have you been hacked? It's only a matter of time if you haven't. Learn how to keep your secrets safe, even when coding in public spaces. Pro tip: if you accidentally paste a password into your code and then remove it, others might still find it in your history, so you'll learn how to navigate that too.

---

<sup>1</sup><https://xkcd.com/353/>

<sup>2</sup><https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

## Tutorial 9: Introduction to TDD using PyTest

By this stage you'll have made a few mistakes. Learn the TDD way and how to write code that tests your other code to catch frustrating errors before they can hurt anyone. Repetitive jobs is what computers are best at after all.

## Tutorial 10: Become an elite hacker with productivity hacks

Have you seen the Matrix? Learn to be the Neo of coding by getting more than one cursor, using keyboard shortcuts, and all of the other productivity features that Repl.it offers. You'll be soon producing more code in less time.

# Part 3: building your own projects

Once you've gone through everything, you might think "but what should I build?". It's a common problem and we've got you covered. Choose your favourite projects from a list (or turn on the coffee machine, order some pizza, and get through them all). Once you've gone through the step-by-step guides you can easily modify or extend the projects to make them your own.

### **Beginner web scraping with Python and Repl.it**

Learn more about what web scraping is, how websites are built, and how to automatically scrape data from websites.

### **Building news word clouds using Python and Repl.it**

Extending the beginner's web scraping tutorial, you'll build a more advanced scraper that extracts the plain text from news articles, stripping away the 'boilerplate' content, such as text in sidebars.

### **Building a Discord Bot with Python and Repl.it**

Build an echo bot using the Discord API. Your bot will always respond with exactly what you send it, but you can customize it afterwards to do something more useful.

### **Building a Discord bot with Node.js and Repl.it**

A NodeJS version of the Discord bot tutorial above. Even if you prefer Python, it's good to go through this one too to get experience with other languages.

### **Creating and hosting a basic web application with Django and Repl.it**

Build a django web application and host it with Repl.it. You'll use geolocation a weather API to show the user their local weather forecast.

## **Building a CRM app with NodeJS, Repl.it, and MongoDB**

Another web application, but using NodeJS instead of Django. This is a different application where you'll build a basic app to manage customer information.

## **Introduction to Machine Learning with Python and repl.it**

Build a machine-learning based text classifier. We skip the maths but show how you can use machine learning libraries to implement useful solutions without in-depth theoretical knowledge.

## **Quicksort tutorial: Python implementation with line by line explanation**

Whether you're applying for jobs or just like algorithms, it's useful to understand how sorting works. In real projects, most of the time you'll just call `.sort()`, but here you'll build a sorter from scratch and understand how it works.

# **Understanding the Repl.it IDE: a practical guide to building your first project with Repl.it**

Software developers can get pretty attached to their Integrated Development Environments (IDEs) and if you look for advice on which one to use, you'll find no end of people advocating strongly for one over another: VS Code, Sublime Text, IntelliJ, Atom, Vim, Emacs, and no shortage of others.

In the end, an IDE is just a glorified text editor. It lets you type text into files and save those files, functionality that has been present in nearly all computers since those controlled by punch cards.

In this lesson, you'll learn how to use the Repl.it IDE. It has some features you won't find in many other IDEs, namely:

- It's fully online. You can use it from any computer that can connect to the internet and run a web browser, including a phone or tablet.
- It'll fully manage your environment for building and running code: you won't need to mess around with making sure you have the right version of Python or the correct NodeJS libraries.
- You can deploy any code you build to the public in one click: no messing around with servers, or copying code around.

In the first part of this guide, we'll cover the basics and also show you how multiplayer works so that you can code alone or with friends.

## **Introduction: creating an account and starting a project**

Although you don't need an account to use Repl.it (you can just navigate to [repl.it<sup>3</sup>](https://repl.it/) and press the "start coding" button), let's set one up in order to have access to all of the features.

Visit <https://repl.it/signup<sup>4</sup>> and follow the prompts to create a user account, either by entering a username and password or by logging in with Google, GitHub, or Facebook.

Once you're done, hit the `+ new repl` button in the top right. In the example below, we choose to create a new Python project. Repl.it will automatically choose a random name for your project, or

---

<sup>3</sup><https://repl.it/>

<sup>4</sup><https://repl.it/signup>

you can pick one yourself. Note that by default your repl will be public to anyone on the internet; this is great for sharing and collaboration, but we'll have to be careful to not include passwords or other sensitive information in any of our projects.

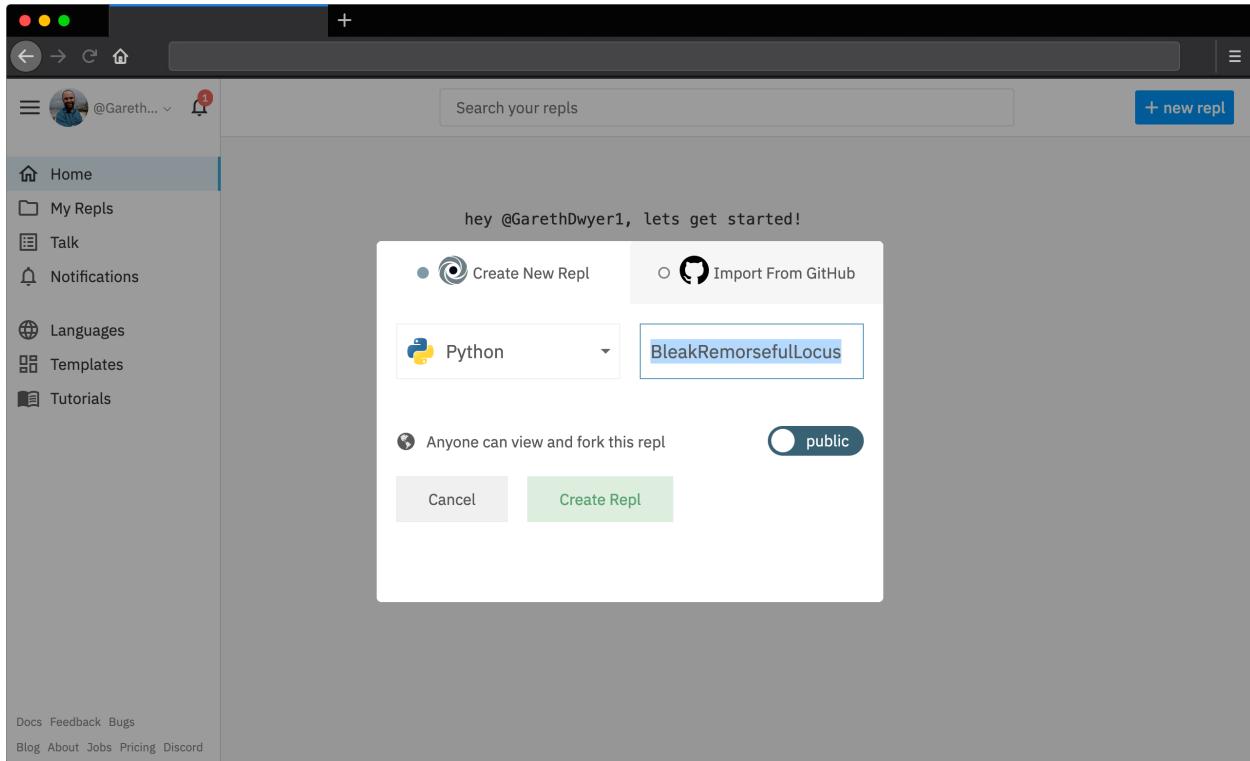


Image 1: *Creating a new Python project*

You'll also notice an "Import from GitHub" option. Repl.it allows you to import existing software projects directly from GitHub, but we'll create our own for now. Once your project is created, you'll be taken to a new view with several panes. Let's take a look at what these are.

## Understanding the Repl.it panes

You'll soon see how configurable Repl.it is and how most things can be moved around to suit your fancy. However, by default, you'll get the following layout.

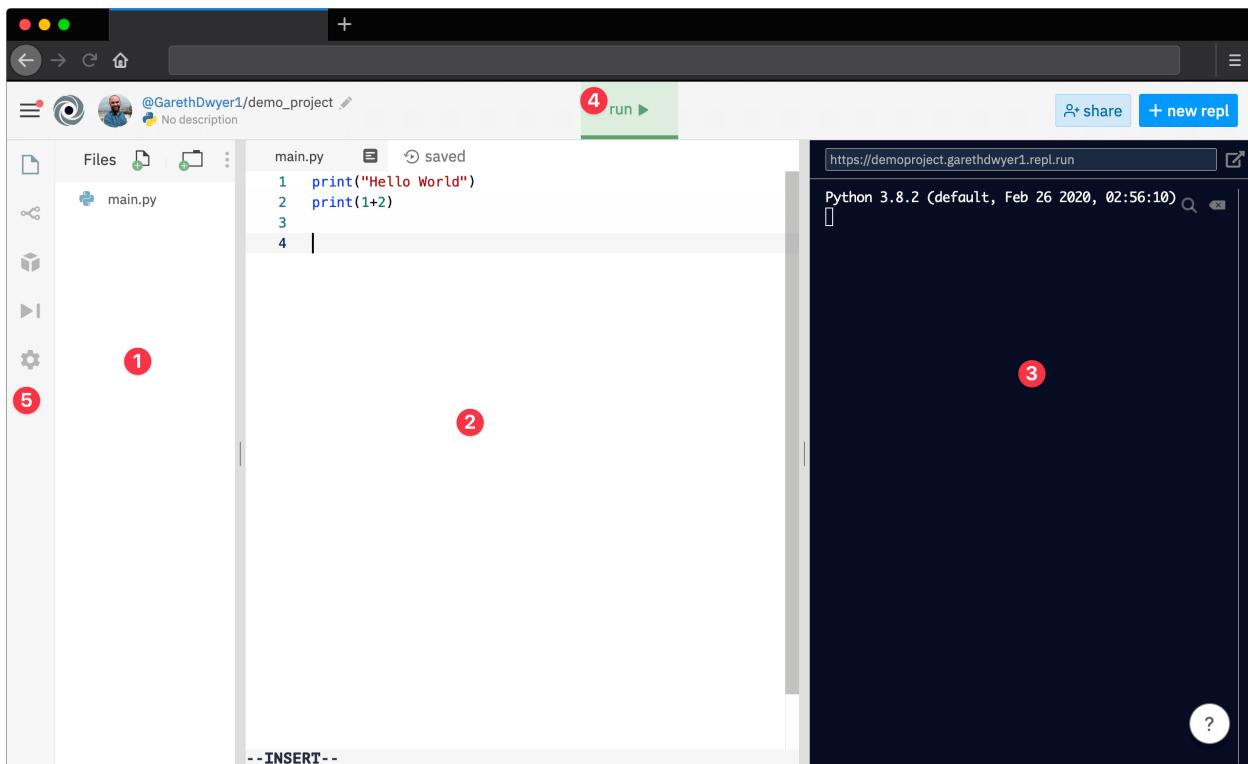


Image 2: The Repl.it panes

1. **Left pane: files and configuration.** This, by default, shows all the files that make up your project. Because we chose a Python project, Repl.it has gone ahead and created a `main.py` file.
2. **Middle pane: code editor.** You'll probably spend most of your time using this pane. It's a text editor where you can write code. In the screenshot, we've added two lines of Python code, which we'll run in a bit.
3. **Right pane: output sandbox.** This is where you'll see your code in action. All output that your program produces will appear in this pane, and it also acts as a quick sandbox to run small pieces of code, which we'll look at more later.
4. **Run button.** If you click the big green run button, your code will be executed and the output will appear on the right.
5. **Menu bar.** This lets you control what you see in the main left pane (pane 1). By default, you'll see the files that make up your project but you can use this bar to view other things here too by clicking on the various icons. We'll take a look at these options later.

Don't worry too much about all of the functionality offered right away. For now, we have a simple goal: write some code and run it.

## Running code from a file

Usually, you'll enter your code as text in a file, and run it from there. Let's do this now. Enter the following code in the middle pane (pane 2), and hit the run button.

```
1 print("Hello World")
2 print(1+2)
```

The screenshot shows the Repl.it IDE interface. On the left, the code editor pane displays a file named 'main.py' with the following content:

```
1 print("Hello World")
2 print(1+2)
3
```

A green 'run' button is visible above the code. To the right, the output pane shows the URL <https://demoproject.garethdwyer1.repl.run>. The output itself consists of two lines of text: "Hello World" and "3".

Image 3: Your first program

Your script will run and the output it generates will appear on the right pane (pane 3). Our code output the phrase “Hello World” (it’s a long-standing tradition that when you learn something new the first thing you do is build a ‘hello world’ project), and then output the answer to the sum  $1 + 2$ . You probably won’t be able to turn this script into the next startup unicorn quite yet, but let’s keep going.

## Running code from Repl.it’s REPL

In computer programming, a REPL is a [read-eval-print loop](#)<sup>5</sup>, and a REPL interface is often the simplest way to run short computer programs (and where Repl.it got its name).

While in the previous example we saved our code to a file and then executed the file, it’s sometimes quicker to execute code directly.

You can type code in the right-hand pane (pane 3) and press the “Enter” key to run it. Take a look at the example below where we print “Hello World” again and do a different sum, without changing our code file.

The screenshot shows the Repl.it IDE interface. The code editor pane on the left shows the same 'main.py' file as before. In the output pane (pane 3) on the right, a user has typed the following commands:

```
Hello World
3
> print("Hello World")
Hello World
> 1 + 5 - 2
4
>
```

The output shows the results of the printed statements and the arithmetic expression.

Image 4: Running code from the REPL

<sup>5</sup>[https://en.wikipedia.org/wiki/Read%E2%80%93eval%E2%80%93print\\_loop](https://en.wikipedia.org/wiki/Read%E2%80%93eval%E2%80%93print_loop)

This is useful for prototyping or checking how things work, but for any serious program you write you'll want to be able to save it, and that means writing the code in a file like in our earlier example.

## Adding more files to your software project

If you build larger projects, you'll want to use more than a single file to stay organised, grouping related code in different files.

So far, we've been using the 'main.py' file that was automatically created for us when we started the project, but we're not limited to this file. Let's add a new file, write some code in that, and import it into the main file for use.

As an example, we'll write code to solve [quadratic equations](#)<sup>6</sup>. If you've done this before, you'll know it can be tedious without a computer to go through all of the steps.

---

<sup>6</sup><https://www.mathsisfun.com/algebra/quadratic-equation.html>

**Example: Solve  $5x^2 + 6x + 1 = 0$**

Coefficients are:  $a = 5, b = 6, c = 1$

$$\text{Quadratic Formula: } x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

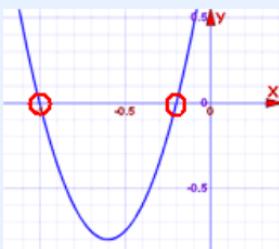
$$\text{Put in } a, b \text{ and } c: x = \frac{-6 \pm \sqrt{6^2 - 4 \times 5 \times 1}}{2 \times 5}$$

$$\text{Solve: } x = \frac{-6 \pm \sqrt{36 - 20}}{10}$$

$$x = \frac{-6 \pm \sqrt{16}}{10}$$

$$x = \frac{-6 \pm 4}{10}$$

$$x = -0.2 \text{ or } -1$$



**Answer:  $x = -0.2 \text{ or } x = -1$**

And we see them on this graph.

$$\begin{aligned} \text{Check } -0.2: \quad & 5 \times (-0.2)^2 + 6 \times (-0.2) + 1 \\ &= 5 \times (0.04) + 6 \times (-0.2) + 1 \\ &= 0.2 - 1.2 + 1 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Check } -1: \quad & 5 \times (-1)^2 + 6 \times (-1) + 1 \\ &= 5 \times (1) + 6 \times (-1) + 1 \\ &= 5 - 6 + 1 \\ &= 0 \end{aligned}$$

Image 5: A quadratic equation example

Let's make Python do the repetitive steps for us by creating a program called "solver". This could eventually have a lot of different solvers, but for now we'll just write one: `solve_quadratic`.

Add a new file to your project by clicking on the `new file` button, as shown below. Call the file `solver.py`. You now have two files in your project: `main.py` and `solver.py`. You can switch between your files by clicking on them.

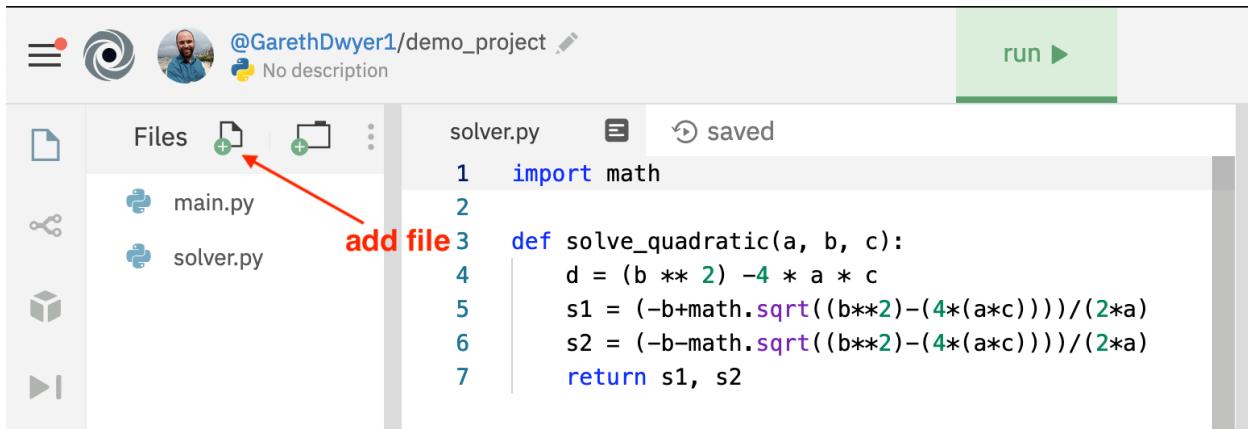


Image 6: Adding a new file

Open the `solver.py` file and add the following code to it.

```

1 import math
2
3 def solve_quadratic(a, b, c):
4     d = (b ** 2) - 4 * a * c
5     s1 = (-b + math.sqrt(d)) / (2 * a)
6     s2 = (-b - math.sqrt(d)) / (2 * a)
7     return s1, s2

```

Note that this won't solve all quadratic equations as it doesn't handle cases where `d`, the discriminant, is 0 or negative. However, it'll do for now.

Navigate back to the `main.py` file. Delete all the code we had before and add the following code instead.

```

1 from solver import solve_quadratic
2
3 answer = solve_quadratic(5, 6, 1)
4 print(answer)

```

Note how we use Python's import functionality to import the code from our new solver script into the `main.py` file so that we can run it. Python looks for `.py` (Python) files automatically, so you omit the `.py` suffix when importing code. Here we import the `solve_quadratic` function (which we just defined) from the `solver.py` file.

Run the code and you should see the solution to the equation, as shown below.

```

from solver import solve_quadratic
answer = solve_quadratic(5, 6, 1)
print(answer)

```

Image 7: The solution to the quadratic equation

Congratulations! You've written your first useful program.

## Sharing your application with others

Coding is more fun with friends or as part of a team. If you want to share your code with others, it's as easy as copying the URL and sending it. In this case, the URL is <https://repl.it/@GarethDwyer1/demoproject>, but yours will be different based on your Repl.it username and the project name you chose.

You can copy the link and open it in an incognito tab (or a different web browser) to see how others would experience your project if you were to share it. By default, they'll be able to see all of your files and code and run your code, but not make any changes. If someone else tries to make changes to your repl, it'll automatically get copied ("forked") to their account, or an anonymous account if they haven't signed up for Repl.it. Any changes your friends make will only happen in their copies, and won't affect your code at all.

To understand this, compare the three versions of the same repl below.

- As you see it, with all of the controls
- As your friend would see it, a read-only version
- As your friend would see it after forking it, on an anonymous account

Image 8: The owner's view of a repl

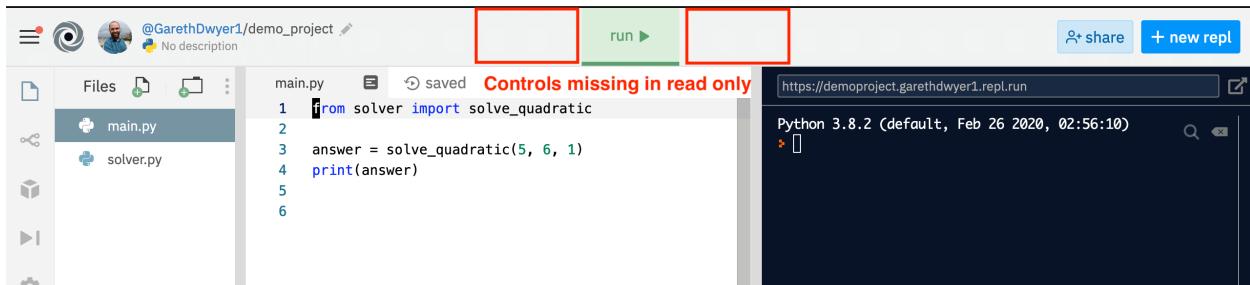


Image 9: A guest's 'read-only' view of a repl

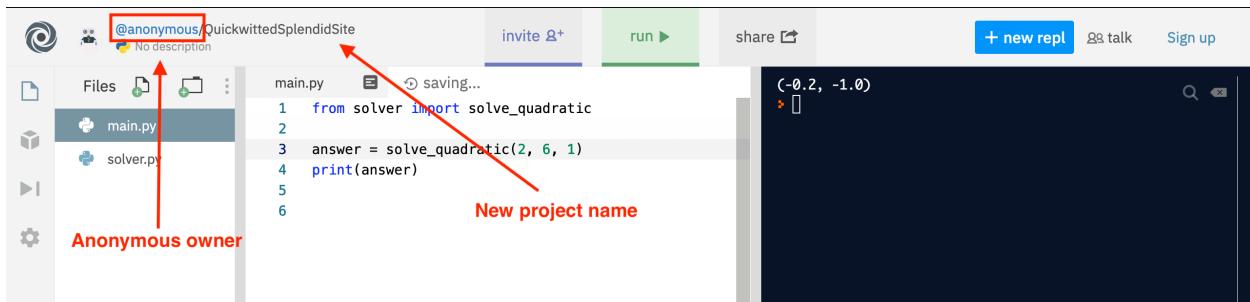


Image 10: An anonymous owner's view of a copy of a repl

What does this mean? Because no one else can edit your repl, you can share it far and wide. But because anyone can read your repl, you should be careful that you don't share anything private or secret in it.

## Sharing write-access: Multiplayer

Of course, sometimes you might want others to have write access to your repl so that they can contribute, or help you out with a problem. In these cases, you can use Repl.it's "multiplayer" functionality.

If you invite someone to your repl, it's different from sharing the URL with them. You can invite someone by clicking Share in the top right and sending them the secret link that starts with "<https://repl.it/join>". This link will give people edit access to your repl.

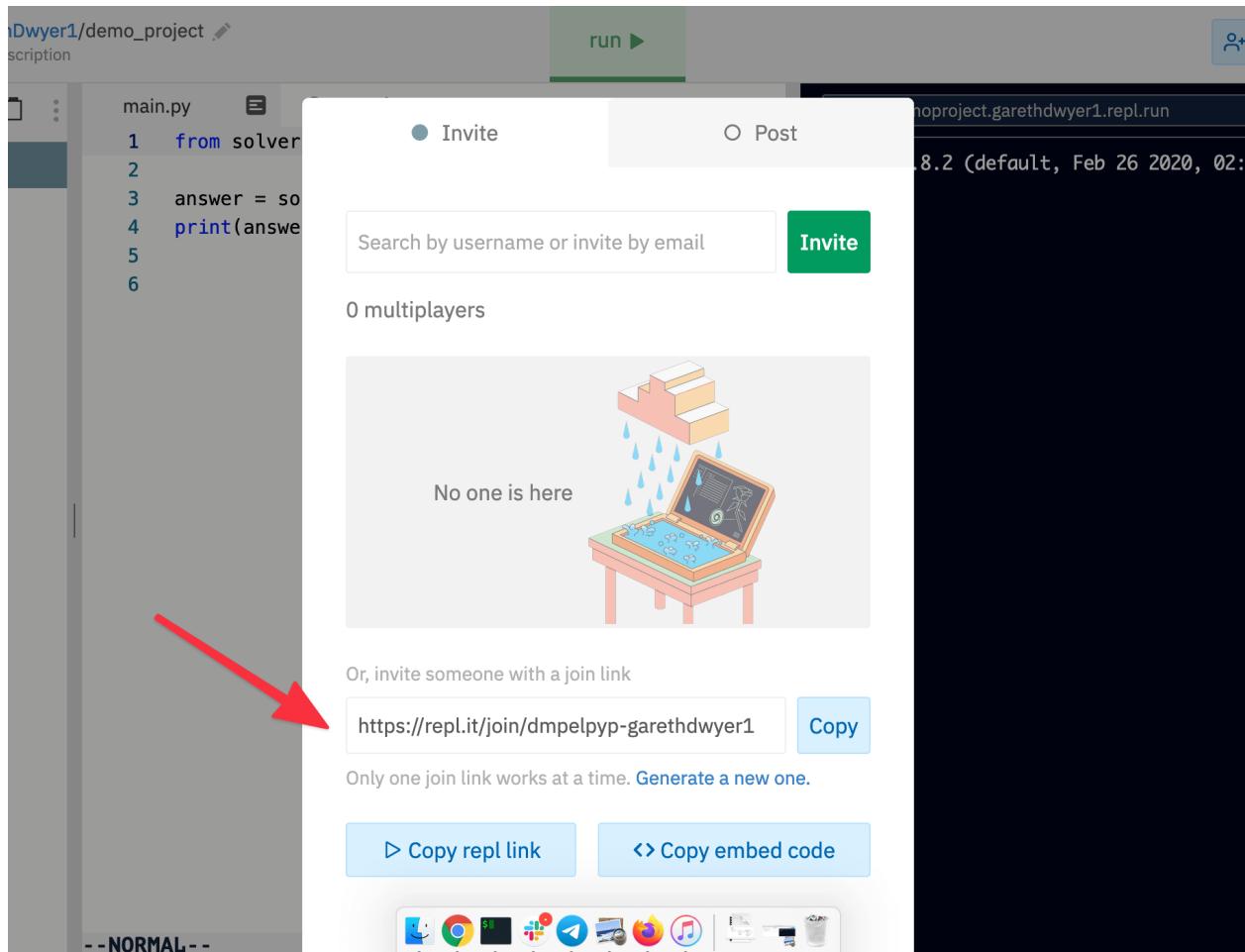


Image 11: *Inviting someone to your repl*

If you have a friend handy, send it to them to try it out. If not, you can try out multiplayer anyway using a separate incognito window again. Below is our main Repl.it account on the left and a second account which opened the multiplayer invite link on the right. As you can see, all keystrokes can be seen by all parties in real time.

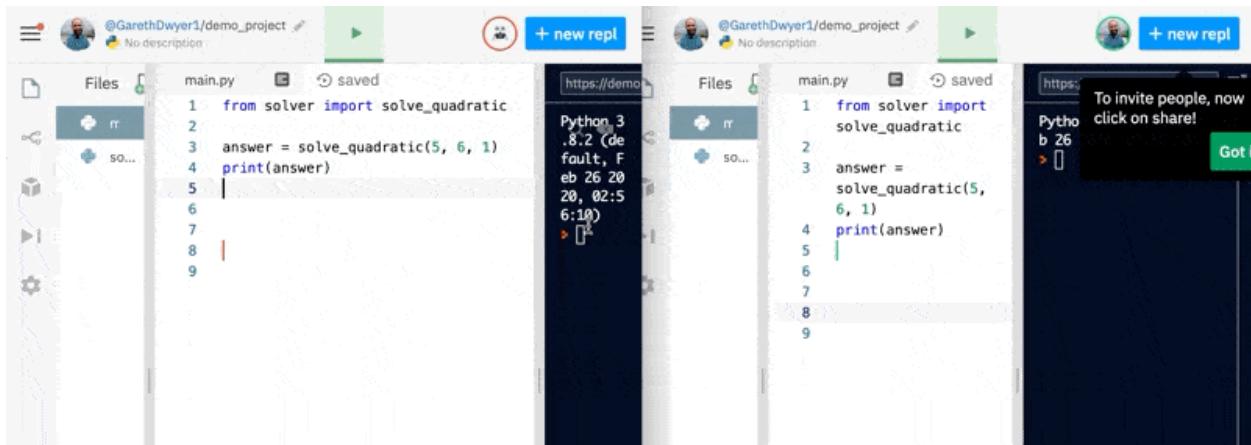


Image 12: Using multiplayer

## Make it your own

If you followed along, you'll already have your own version of the repl to extend. If not, start from ours. Fork it from the embed below.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-01-quadratic-equations?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

You can now create basic programs on your own or with friends, and you are familiar with the most important Repl.it features. There's a lot more to learn though. In the next nine lessons, you'll work through a series of projects that will teach you more about Repl.it features and programming concepts along the way.

If you get stuck, you can get help from the [Repl.it community](#)<sup>7</sup> or on the [Repl.it Discord server](#)<sup>8</sup>.

---

<sup>7</sup><https://repl.it/talk/all>

<sup>8</sup><https://repl.it/discord>

# Working with Files using Repl.it

In this lesson, you'll gain experience with using and manipulating files using Repl.it. In the previous lesson you saw how to add new files to a project, but there's a lot more you can do.

Files can be used for many different things. In programming, you'll primarily use them to store data or code. Instead of manually creating files and entering data, you can also use your programs to create files and automatically write data to these.

Repl.it also offers functionality to mass import or export files from or to the IDE; this is useful in cases when your program writes data to multiple files and you want to export all of these for use in another program.

## Working with files using Python

Create a new Python repl and call it `working-with-files`.

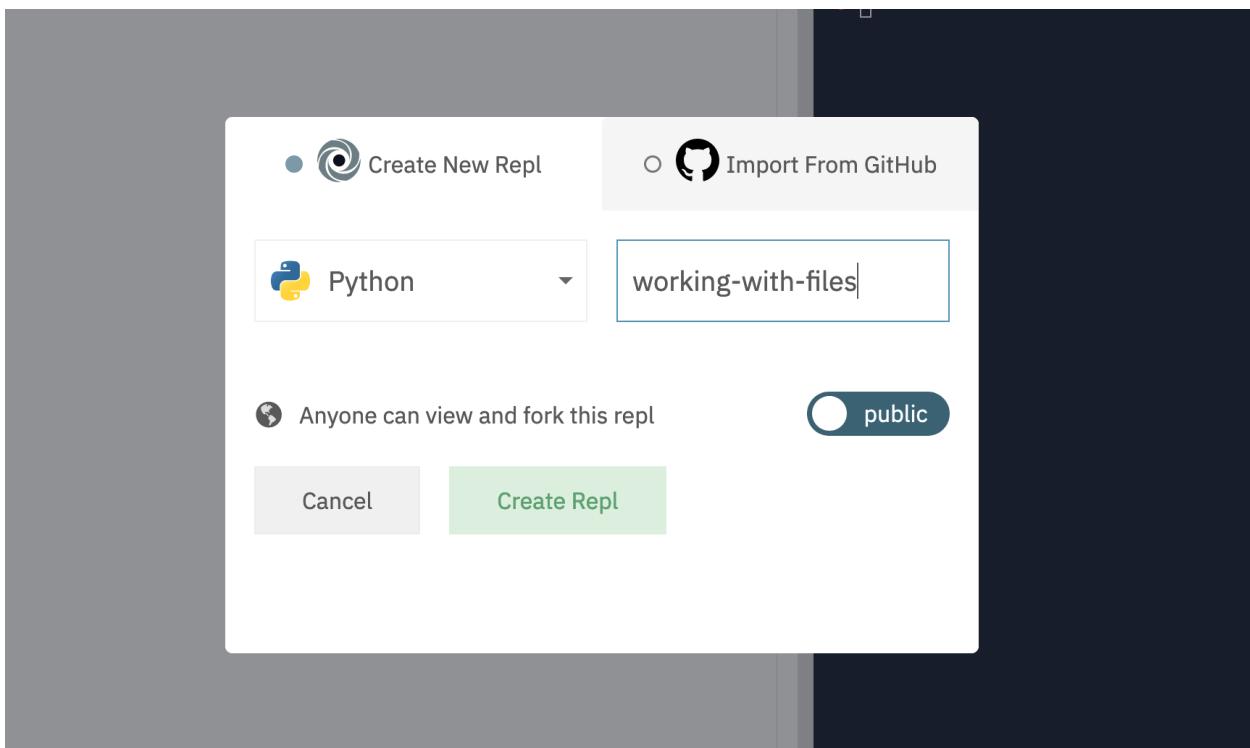


Image 1: Creating our files project

As before, you'll get a default repl project with a `main.py` file. We need a data file to practise reading data programmatically.

1. Create a new file using the `add file` button.
2. Call the file `mydata.txt`. Previously we created a Python file (`.py`) but in this case we are creating a plain text file (`.txt`).
3. Type a line of text into the file.

You should now have something similar to what you see below.

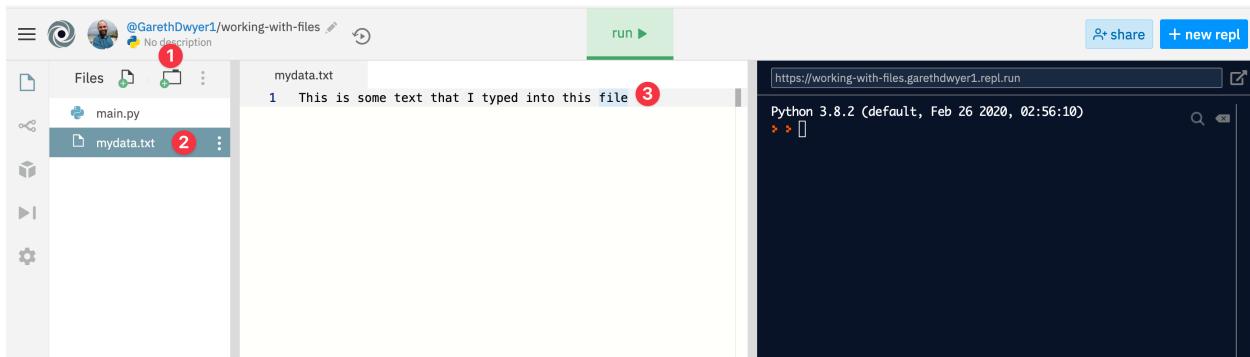


Image 2: Adding data manually to our text file

Go back to the `main.py` file and add the following code.

```
1 f = open("mydata.txt")
2 contents = f.read()
3 f.close()
4 print(contents)
```

This opens the file programmatically, reads the data in the file into a variable called `contents`, closes the file, and prints the file contents to our output pane.

Press the Run button. If everything went well, you should see output as shown in the image below.

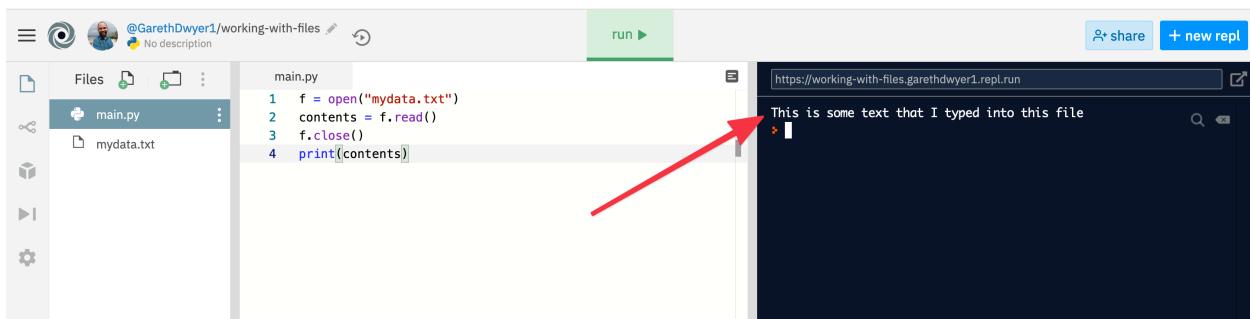


Image 3: Reading data from a file and printing it out

## Creating files using Python

Instead of manually creating files, you can also use Python to do this. Add the following code to your `main.py` file.

```

1 f = open("createdfile.txt", "w")
2 f.write("This is some data that our Python script created.\n")
3 f.close()

```

Note the `w` argument that we pass into the `open` function now. This means that we want to open the file in “write” mode. Be careful: if the file already exists, Python will delete it and create a new one. There are many different ‘modes’ to work with files in different ways which you can read about in the [Python documentation](#)<sup>9</sup>.

Run the code again and you should see a new file pop up in the files pane. If you click on the file, you’ll find the data that the Python script wrote to it, as shown below.



Image 4: Writing data to a file and viewing it

## Building a weather logging system using Python and Repl.it

Now that you can read from files and write to them, let’s build a mini-project that records historical weather temperatures. Our program will

- Get the current temperature for a specified set of cities.
- Write this data to file with today’s date.
- Print a summary of this data to the console.

## Creating a WeatherStack Account and getting an API key

We’ll get weather data from the WeatherStack API. You’ll need to sign up at [weatherstack.com](#)<sup>10</sup> and follow the instructions to get your own access key. Choose the “free tier” option which is limited to 1000 calls per month.

After sign-up, you should see a page similar to the following containing the API access key.

<sup>9</sup><https://docs.python.org/3/tutorial/inputoutput.html#reading-and-writing-files>

<sup>10</sup><https://weatherstack.com>

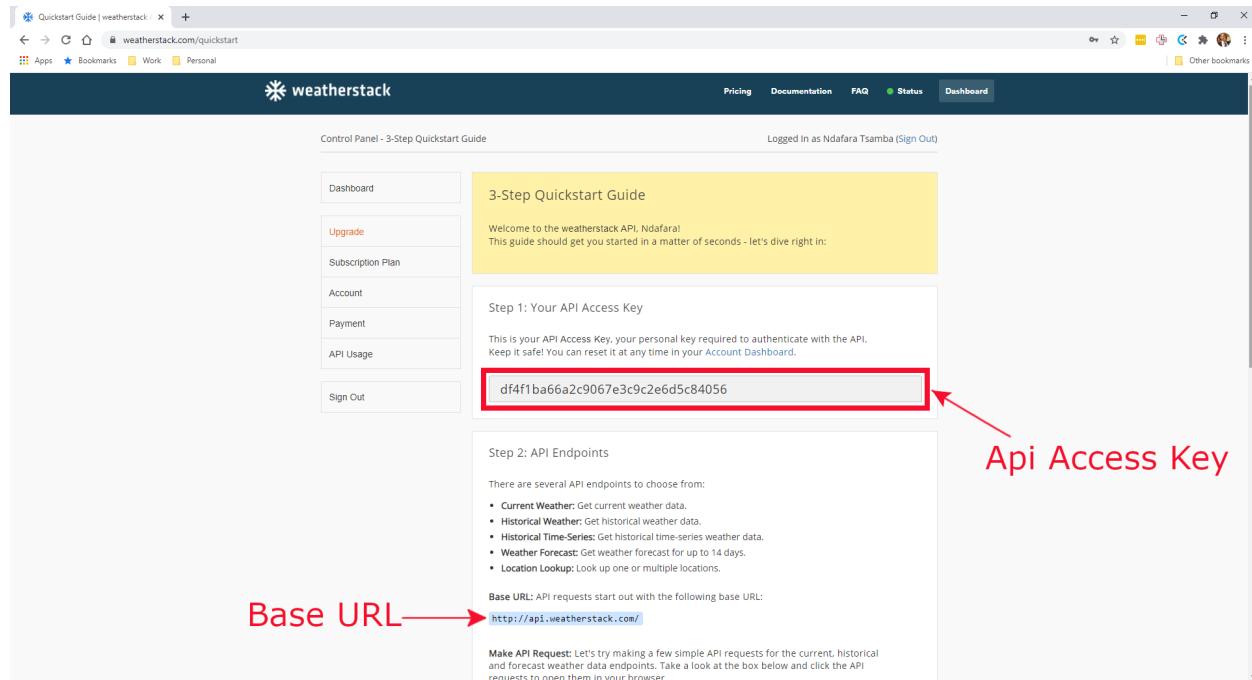


Image 5: Getting an API key from WeatherStack

You should keep this key secret to stop other people using up all of your monthly calls.

## Creating our weather reporting project

You can continue to use the `working-with-files` project that you created earlier if you want to, but for the demonstration we'll create a new Python repl called `weather_report`.

In the `main.py` file add the following code, but replace the string for `API_KEY` with your own one.

```

1 import requests
2
3 # change the following line to use your own API key
4 API_KEY = "baaf201731c0cbc4af2c519cb578f907"
5 WS_URL = "http://api.weatherstack.com/current"
6
7 city = "London"
8
9 parameters = { 'access_key': API_KEY, 'query': city}
10
11 response = requests.get(WS_URL, parameters)
12 js = response.json()
13 print(js)
14 print()
```

This code asks WeatherStack for the current temperature in London, gets the [JSON<sup>11</sup>](#) version of this and prints it out. You should see something similar to what is shown below.

The screenshot shows a Repl.it interface. On the left, there's a file browser with 'main.py' selected. The code in 'main.py' is:

```

1 import requests
2
3 API_KEY = "baaf201731c0cbc4af2c519cb578f907"
4 WS_URL = "http://api.weatherstack.com/current"
5
6 city = "London"
7
8 parameters = {'access_key': API_KEY, 'query': city}
9
10 response = requests.get(WS_URL, parameters)
11 js = response.json()
12 print(js)
13 print()
14

```

On the right, the terminal window shows the JSON response from WeatherStack. The JSON object has several fields, with some specific values highlighted by red boxes and numbers:

```

{
  "request": {
    "type": "City",
    "query": "London, United Kingdom",
    "language": "en",
    "unit": "m"
  },
  "location": {
    "name": "London, Greater London",
    "region": "Greater London",
    "country": "United Kingdom",
    "lat": "51.5074",
    "lon": "0.1278",
    "timezone": "Europe/London",
    "localtime": "2020-07-16 10:25"
  },
  "current": {
    "temperature": 17,
    "weather_code": 12,
    "humidity": 72,
    "wind_degree": 310,
    "wind_dir": "NW",
    "pressure": 1020,
    "precip": 0,
    "cloudcover": 100,
    "feelslike": 17,
    "uv_index": 4,
    "visibility": 10,
    "is_day": "yes"
  }
}

```

Image 6: Getting the current weather in JSON format

WeatherStack returns a lot of data, but we are mainly interested in

- The location:** To see if we found the correct London and not one of the [29 other places<sup>12</sup>](#) called London.
- The date:** We'll record this when we save this data to a file.
- The current temperature:** This is specified by default in Celsius, but [can be customised<sup>13</sup>](#) if you prefer Fahrenheit.

Add the following code below the existing code to extract these values into a format that's easier to read.

```

1 temperature = js['current']['temperature']
2 date = js['location']['localtime']
3 city = js['location']['name']
4 country = js['location']['country']
5
6 print(f"The temperature in {city}, {country} on {date} is {temperature} degrees Cels\
ius")

```

If you run the code again, you'll see a more human-friendly output, as shown below.

<sup>11</sup>[https://www.w3schools.com/js/js\\_json\\_intro.asp](https://www.w3schools.com/js/js_json_intro.asp)

<sup>12</sup><https://www.wanderlust.co.uk/content/londons-around-the-world/>

<sup>13</sup><https://weatherstack.com/documentation>

```

import requests
API_KEY = "baaf201731c0cbc4af2c519cb578f907"
WS_URL = "http://api.weatherstack.com/current"
city = "London"
parameters = {'access_key': API_KEY, 'query': city}
response = requests.get(WS_URL, parameters)
js = response.json()
print(js)
print()
temperature = js['current']['temperature']
date = js['location']['localtime']
city = js['location']['name']
country = js['location']['country']
print(f"The temperature in {city}, {country} on {date} is {temperature} degrees celsius")

```

Image 7: Seeing a human-readable summary of the data

This is great for getting the current weather, but now we want to extend it a bit to record weather historically.

We'll create a file called `cities.txt` containing the list of cities we want to get weather data for. Our script will request the weather for each city, and save a new line with the weather and timestamp.

Add the `cities.txt` file, as in the image below (of course, you can change which cities you would like to get weather info for).

Image 8: Creating the cities.txt file

Now remove the code we currently have in `main.py` and replace it with the following.

```

import requests
API_KEY = "baaf201731c0cbc4af2c519cb578f907"
WS_URL = "http://api.weatherstack.com/current"
cities = []
with open("cities.txt") as f:
    for line in f:
        cities.append(line.strip())
print(cities)

```

```

11
12 for city in cities:
13     parameters = { 'access_key': API_KEY, 'query': city}
14     response = requests.get(WS_URL, parameters)
15     js = response.json()
16
17     temperature = js['current']['temperature']
18     date = js['location']['localtime']
19
20     with open(f"{city}.txt", "w") as f:
21         f.write(f"{date},{temperature}\n")

```

This is similar to the code we had before, but now we

- Read the city names from our `cities.txt` file and put each city into a Python list.
- Loop through the cities and get the weather data for each one.
- Create a new file with the same name as each city and write the date and temperature (separated by a comma) to each file.

In our previous examples we explicitly closed files using `f.close()`. In this example, we instead open our files in a `with` block. This is a common idiom in Python and is usually how you will open files. You can read more about this in the [files section of the Python docs<sup>14</sup>](#).

If you run this code, you'll see it creates one file for each city.

The screenshot shows the Repl.it interface. On the left, there's a file tree with `main.py`, `cities.txt`, and four files created: `London.txt`, `Paris.txt`, `San Francisco.txt`, and `Sydney.txt`. The `London.txt` file is highlighted with a red box. The main area shows the code for `main.py`, which reads city names from `cities.txt` and writes them to individual files. The output window on the right shows the list of cities: ['London', 'Paris', 'San Francisco', 'Sydney'].

```

main.py
1 import requests
2
3 API_KEY = "baaf201731c0cbc4af2c519cb578f907"
4 WS_URL = "http://api.weatherstack.com/current"
5
6 cities = []
7 with open("cities.txt") as f:
8     for line in f:
9         cities.append(line.strip())
10 print(cities)
11

```

Image 9: *The script creates one file for each city*

If you open up one of the files, you'll see it contains the date and temperature that we fetched from WeatherStack.

<sup>14</sup><https://docs.python.org/3/tutorial/inputoutput.html#reading-and-writing-files>

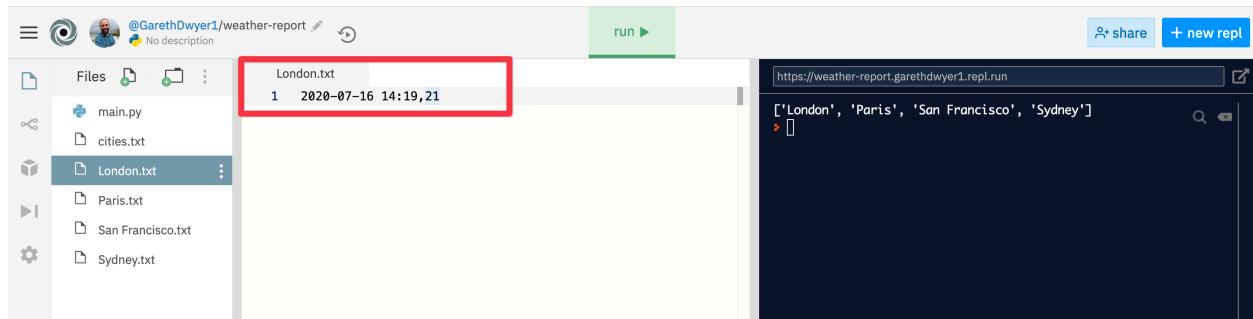


Image 10: Example data recorded for London

If you run the script multiple times, each file will still only contain one line of data: that from the most recent run. This is because when we open a file with in “write” mode (“w”), it overwrites it with the new data. Because we want to create historical weather logs, we need to change the second last line to use “append” mode instead (“a”).

Change

```
1     with open(f"{city}.txt", "w") as f:
```

to

```
1     with open(f"{city}.txt", "a") as f:
```

and run the script again. If you open one of the city files again, you’ll see it has a new line instead of the old data being overwritten. Newer data is appended to the end of the file. WeatherStack only updates its data every 5 minutes or so, so you might see exact duplicate lines if you run the script multiple times in quick succession.



Image 10: Adding new data to the end of each file

## Exporting our weather data files

If you run this script every day for a few months, you’ll have a nice data set that could be useful in other contexts too. If you want to download all of the data from Repl.it, you can use the Download as zip functionality to export all of the files in a repl (including the code and data files).

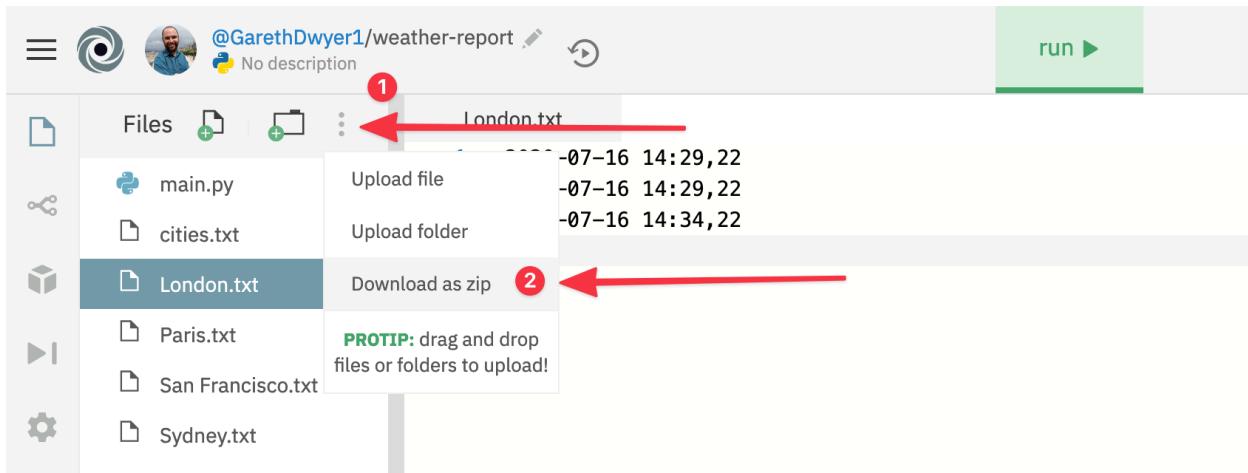


Image 11: Downloading all of our files from Repl.it

Once you've downloaded the .zip file you can extract it in your local file system and find all of the data files which can now be opened with other programs as required.

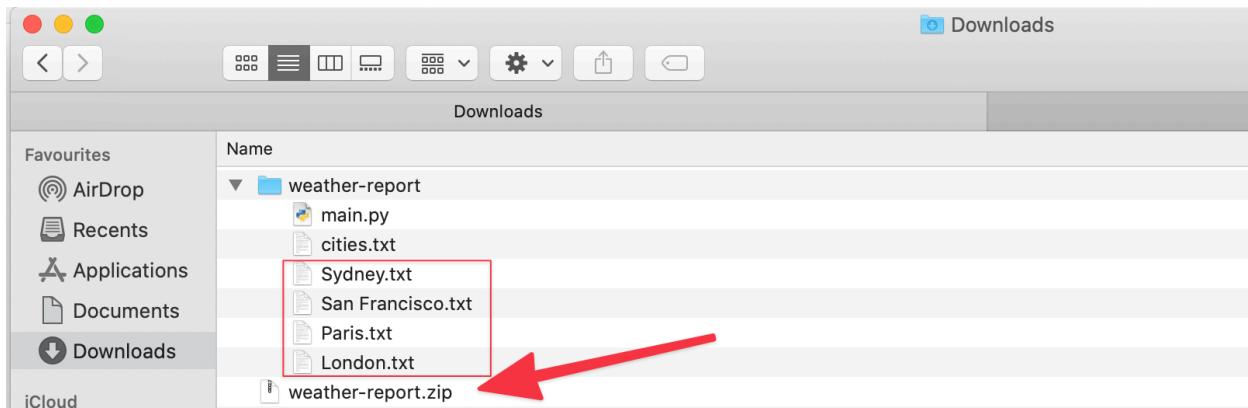


Image 12: The created data files on our local file system

From the same menu, you can also choose upload file or upload folder to import files into your repl. For example, if you cleaned the files using external software and then wanted your repl to start appending new data to the cleaned versions, you could re-import them.

Repl.it will warn you about overwriting your existing files if you haven't changed the names.

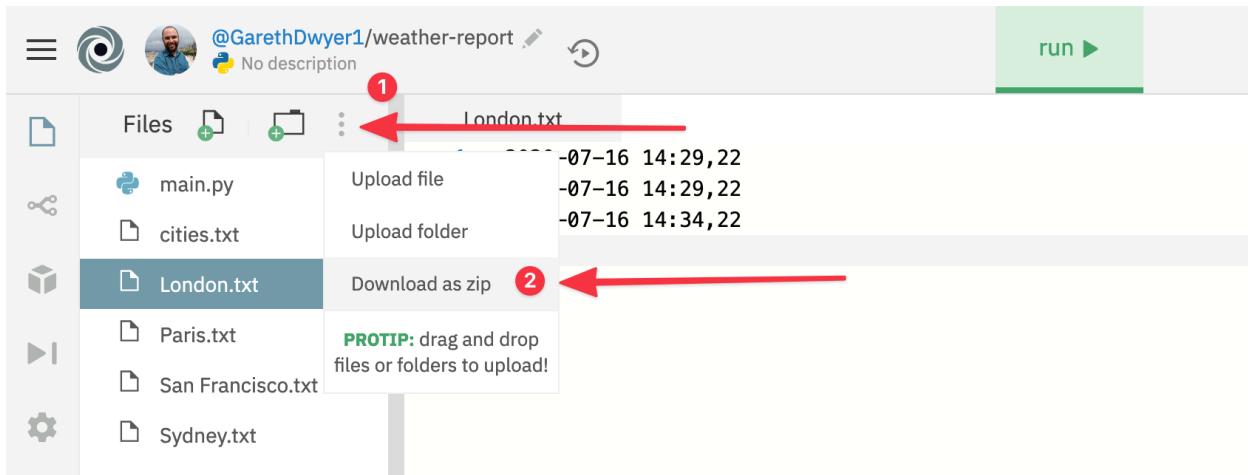


Image 13: Be careful about overwriting your precious data

## Make it your own

If you followed along, you'll already have your own version of the repl to extend. If not, start from ours. Fork it from the embed below.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-02-weather-report?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

That's it for our weather reporting project. You learned how to work with files in Python and Repl.it, including different modes (read, write, or append) in which files can be opened.

You also worked with an external library, `requests`, for fetching data over the internet. This module is not actually part of Python, and in the next article you'll learn more about how to manage external modules or dependencies.

# Managing dependencies using Repl.it

Nearly all useful programs rely to some extent on pre-existing code in various forms. The existing code that your code relies on is known as a dependency. You have already come across some dependencies in previous tutorials: you used the `math` module to calculate quadratic equations in the first tutorial and you used the `requests` module to fetch weather data in the second tutorial.

In the first tutorial, you also wrote the `solver.py` module, and imported this into the `main.py` file.

We can think of dependencies as falling into three broad categories:

- Internal dependencies: other code that you or your organisation wrote and which you fully control, e.g. the `solver.py` file.
- Standard dependencies: code that exists as part of the standard language libraries, e.g. the `math` module.
- External dependencies: code that is written by third-party developers, e.g. the `requests` module.

In this tutorial, you'll gain more experience with all three categories of dependencies. Specifically, you'll write an NLP (natural language processing) program to analyse sentences, using `spaCy`<sup>15</sup>, a third-party dependency.

Dependency management is a hugely complicated area, and there is a large ecosystem of related tools to help manage packaging and installing Python programs. We won't be covering all of the options and background, but you can read an overview of the different tools [here](#)<sup>16</sup>.

## Understanding Repl.it's magic import tool and the universal package manager

In nearly all programming environments, you have to explicitly install third-party dependencies. Let's say you wanted to use the `requests` library (which is not included in Python by default) on your local machine. If you try to import it, you would get a `ModuleNotFoundError` error, as shown below.

```
Type "help", "copyright", "credits" or "license" for more information.
>>> import requests
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'requests'
>>> 
```

Image 1: Trying to use a dependency without installing it

<sup>15</sup><https://spacy.io/>

<sup>16</sup><https://modelpredict.com/python-dependency-management-tools>

In order to use this library, you would first have to install it using a command similar to `pip install requests`, and only then would the import statement run correctly.

Repl.it, by contrast, can often do the installation for you completely automatically, using the [Universal Package Manager](#)<sup>17</sup>. The moment you run the `import requests` line of code, the package manager will go find the correct package and install it, or in some cases Repl.it will even have pre-installed the package. Either way, your code will “just work”.

```
main.py
1 import requests
2 print(requests.get("https://google.com").text)
3
```

The browser preview shows the HTML response from Google's homepage, indicating that the package was successfully imported and used.

Image 2: Seamless external package use on a new repl

This is super convenient, but sometimes you need more control. For example, you might need a specific version of a package, or the universal package manager might not be able to automatically install all of your dependencies. In these cases, you can use more advanced ways to install packages.

## Installing packages through the GUI

If you’re not sure exactly which package you need, you can use Repl.it’s built-in package manager GUI to search for packages. In the example below, we are looking for a package called `beautifulsoup4`.

To use this, you need to

1. Click on the packages tab from the left toolbar.
2. Search for a package by typing in part or all of its name.
3. Select the package you want from the search results.

<sup>17</sup><https://blog.repl.it/upm>

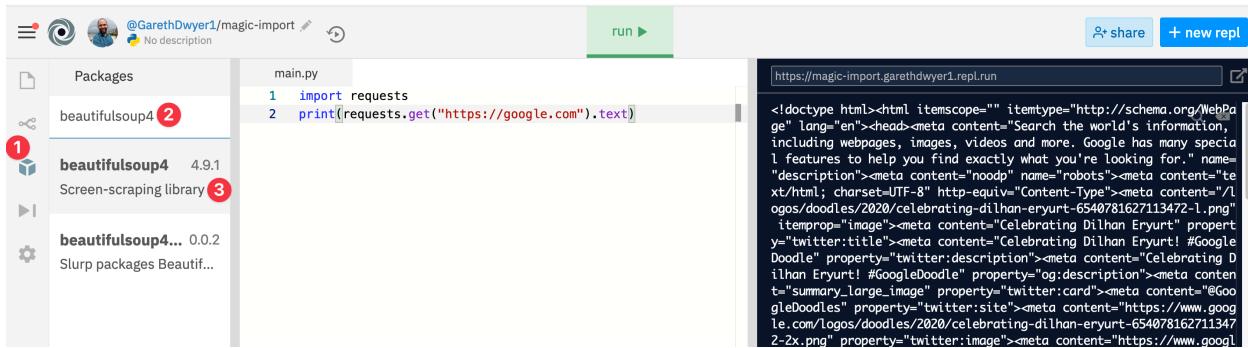


Image 3: Using the GUI package manager to find a package

This will take you to a page showing an overview and summary of the selected package. You can install it to your repl by using the + button, as shown below.

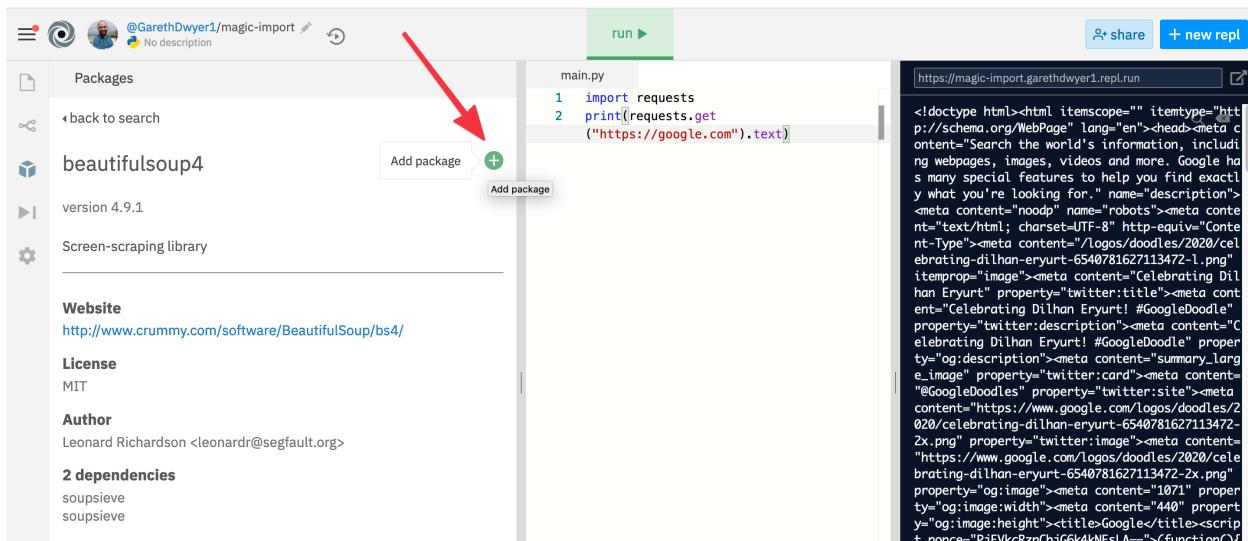


Image 4: Installing a package from the overview page

Once the package is installed, we can use it in our code. Run the example shown below to extract the “Google Search” text from the main button on the homepage.

```

1 import requests
2 from bs4 import BeautifulSoup
3
4 r = requests.get("https://google.com").text
5 soup = BeautifulSoup(r, "html.parser")
6
7 print([x.get("title") for x in soup.findAll("input") if x.get("title")])

```

This code uses the requests library to scrape the HTML from google.com and then uses the beautifulsoup4 library to get the title of the button off the page and print it to the console.

Because `requests` is one of the most commonly used Python libraries, Repl.it probably installed it in a slightly different way from most packages. However, `beautifulsoup4` is less common and this will have been installed in the standard way using `poetry`<sup>18</sup>.

If you go back to the files tab, you'll see two new files `poetry.lock` and `pyproject.toml` which were created automatically by the installer. Take a look inside the `pyproject.toml` file.

```

pyproject.toml
1 [tool.poetry]
2 name = "magic-import"
3 version = "0.1.0"
4 description = ""
5 authors = ["Your Name <you@example.com>"]
6
7 [tool.poetry.dependencies]
8 python = "3.8"
9 beautifulsoup4 = "^4.9.1" ← Red arrow here
10
11 [tool.poetry.dev-dependencies]
12
13 [build-system]
14 requires = ["poetry>=0.12"]
15 build-backend = "poetry.masonry.api"
16

```

Image 5: The `pyproject.toml` file lists all dependencies and their versions.

In this case, line 9 says that our project relies on the `beautifulsoup4` package and needs at least version 4.9.1. If we look at [the beautifulsoup page on PyPi](#)<sup>19</sup>, we'll see that the latest stable version is 4.9.1, so if this project is run in the future and there is a new version available, it will automatically use the updated package.

## Building an NLP project using spaCy

So far, we have installed packages that are easy for the Repl.it universal dependency manager to install automatically, behind the scenes. Some packages are more complicated though. `spaCy`<sup>20</sup>, for example, is an NLP library that relies on a large external data file. When installing this library, you usually have to install this data file as a separate step.

### Installing the spaCy language model

To get this to work on Repl.it, we'll have to manually modify the `pyproject.toml` file.

Create a new repl, `SpacyExample`, then click on the Packages icon and search for "spacy".

<sup>18</sup><https://python-poetry.org/docs/basic-usage/>

<sup>19</sup><https://pypi.org/project/beautifulsoup4/>

<sup>20</sup><https://spacy.io/>

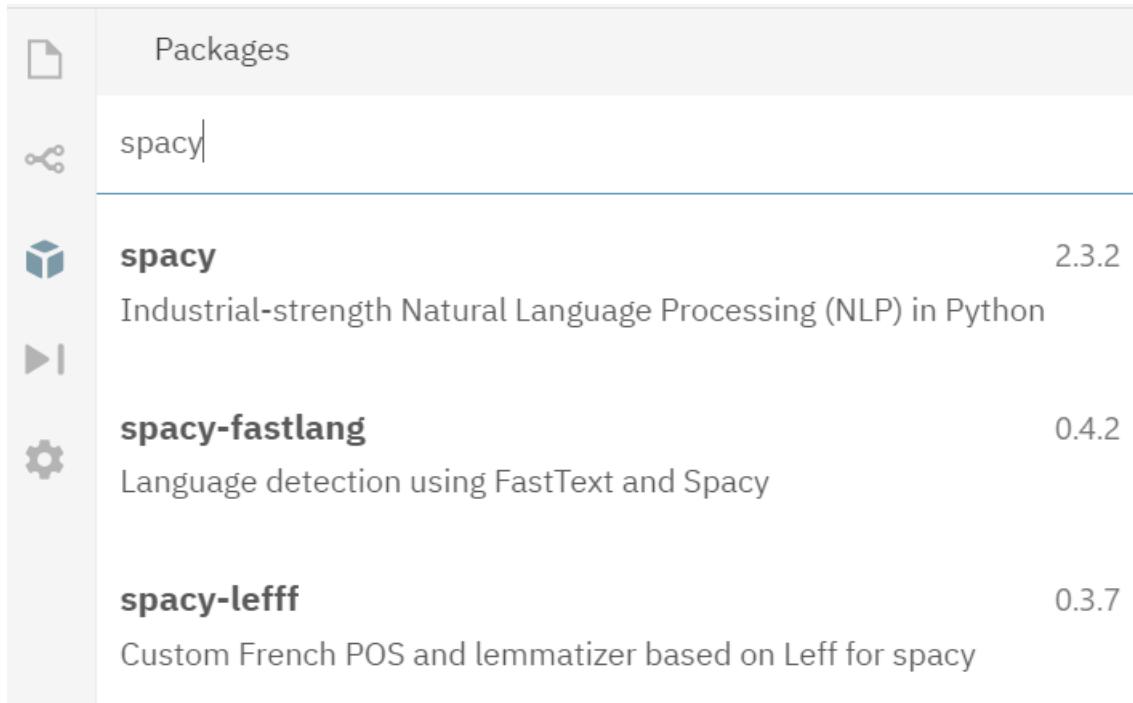


Image 6: We can access the `spacy` package via the package index"

Select the version at the top and hit the + button to add this package to your application. Once this is complete, head across to your `main.py` and enter the following code:

```
1 import spacy
2 print(spacy.__version__)
```

This should output the version of spaCy that we are using, which means that spaCy has been added as a dependency correctly.

If you take a look at your `pyproject.toml` file now, you should see that it has specified spaCy as a dependency.

```
1 [tool.poetry]
2 name = "spacy-example"
3 version = "0.1.0"
4 description = ""
5 authors = ["Your Name <you@example.com>"]
6
7 [tool.poetry.dependencies]
8 python = "^3.8"
9 spacy = "^2.3.2"
10
11 [tool.poetry.dev-dependencies]
```

```
12  
13 [build-system]  
14 requires = ["poetry>=0.12"]  
15 build-backend = "poetry.masonry.api"
```

An important component of spaCy is a set of pretrained statistical models that support NLP. These do not come with spaCy by default, nor are they indexed on PyPi. One of these models is `en_core_web_sm`.

In your `main.py` file, replace your current code with the following:

```
1 import spacy  
2  
3 nlp = spacy.load("en_core_web_sm")  
4 doc = nlp("The quick brown fox jumps over the lazy dog.")  
5 for token in doc:  
6     print(token.text)
```

This code should simply break our short sentence into tokens (words), and print each one out.

However, at this point, if you run your code you will get an error, as Python cannot find the `en_core_web_sm` model:

```
1 OSError: [E050] Can't find model 'en_core_web_sm'. It doesn't seem to be a shortcut \  
2 link,  
3 a Python package or a valid path to a data directory.
```

We will now explicitly tell our application how to access this dependency. To do this, we need to find where the model is stored online.

First, we need to find the spaCy documentation for this model. This can be accessed [here<sup>21</sup>](#).

---

<sup>21</sup><https://spacy.io/models/en>

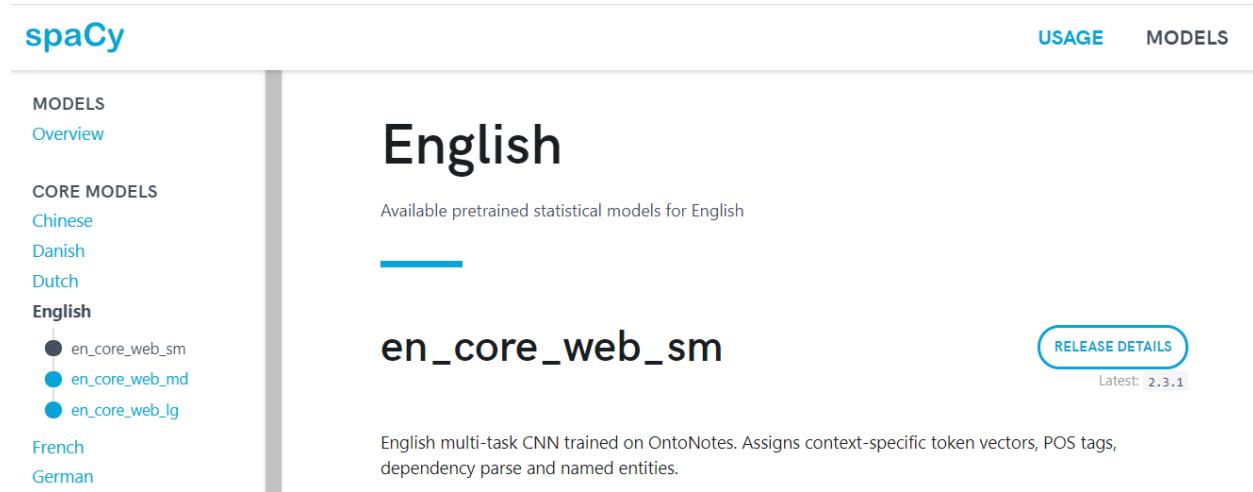


Image 7: *The spaCy documentation gives us information about the model*

Selecting the RELEASE DETAILS button will guide us to where the model is stored online, on GitHub<sup>22</sup>. GitHub is a very common place to store code and related components online.

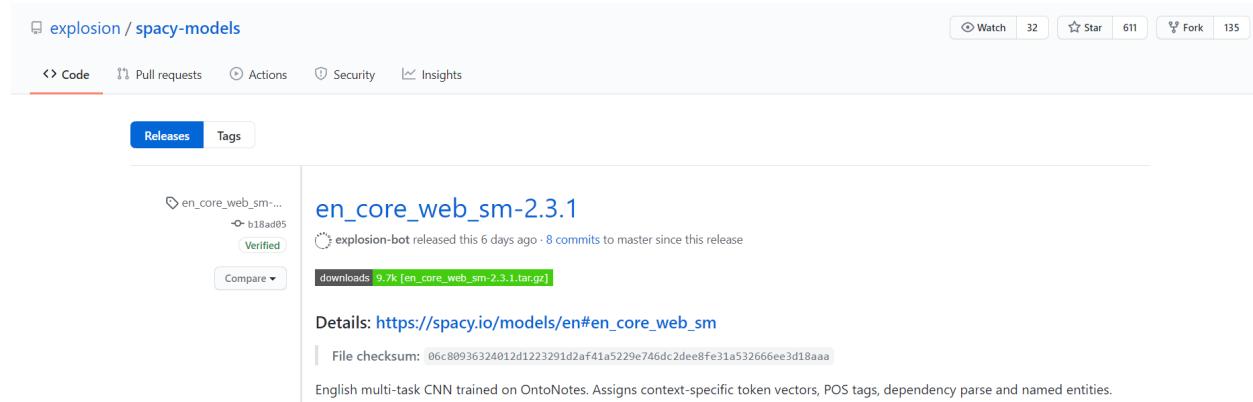


Image 8: *This is the en\_core\_web\_sm GitHub page*

The GitHub page also lets us know what version of spaCy is needed to make sure the model runs correctly.

<sup>22</sup>[https://github.com/explosion/spacy-models/releases//tag/en\\_core\\_web\\_sm-2.3.1](https://github.com/explosion/spacy-models/releases//tag/en_core_web_sm-2.3.1)

Feature	Description
Name	en_core_web_sm
Version	2.3.1
spaCy	>=2.3.0,<2.4.0
Model size	11 MB
Pipeline	tagger, parser, ner
Vectors	0 keys, 0 unique vectors (0 dimensions)
Sources	OntoNotes 5
License	MIT
Author	Explosion

Image 9: The GitHub page provides information about the requirements and features of the model

Here we see that spaCy version should be greater than or equal to 2.3.0, but less than 2.4.0. We should make a note of this for later, so we can check that we have pinned an appropriate spaCy version.

If we scroll right to the bottom of the page, you will see an “Assets” section, and under this you will see the same Package icon we used in Repl.it with “en\_core\_web\_sm-2.3.1.tar.gz” next to it. This is what we have been looking for: the file containing the model.

Assets 3	
 en_core_web_sm-2.3.1.tar.gz	11.5 MB
 Source code (zip)	
 Source code (tar.gz)	

Image 10: The model can be found under the “Assets” heading

Right-click on this file and select copy link address. We will need this shortly, as this is the URL of the file.

We now need to modify our `pyproject.toml` file in Repl.it. Open this file and add the following section to it

```

1 [tool.poetry.dependencies.en_core_web_sm]
2 url = "https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-2.3.1.tar.gz"

```

The `url` should be the one that you copied from GitHub in the previous step. Your whole `pyproject.toml` file should now look like the one below.



```

Files main.py Packager files poetry.lock pyproject.toml

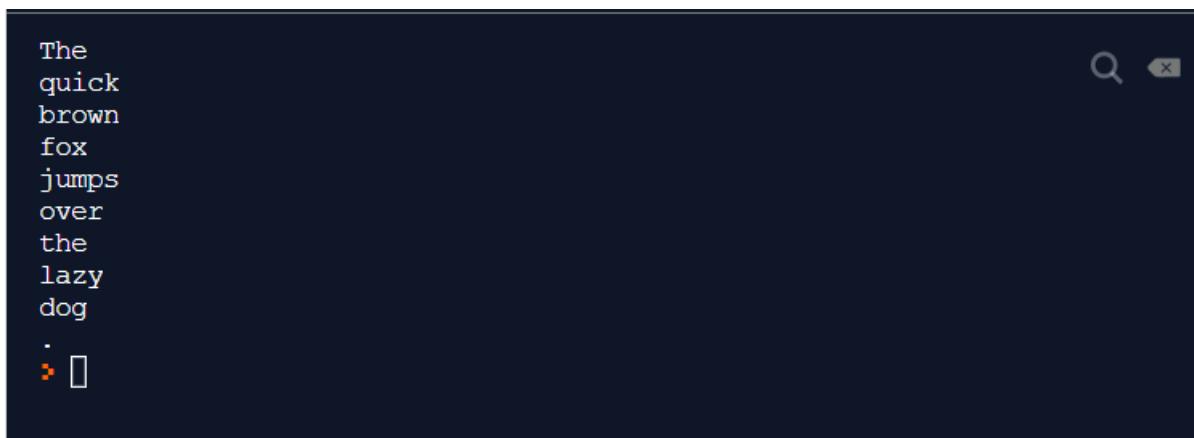
pyproject.toml
1 [tool.poetry]
2 name = "spacyexample"
3 version = "0.1.0"
4 description = ""
5 authors = ["Your Name <you@example.com>"]
6
7 [tool.poetry.dependencies]
8 python = "^3.8"
9 spacy = "^2.3.1"
10
11 [tool.poetry.dev-dependencies]
12
13 [tool.poetry.dependencies.en_core_web_sm]
14 url = "https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-2.3.1.tar.gz"
15
16 [build-system]
17 requires = ["poetry>=0.12"]
18 build-backend = "poetry.masonry.api"
19
20
21
22

```

Image 11: *Modifying `pyproject.toml` to explicitly point to the model allows our application to find it and use it*

At this point that we should also check that we are using an appropriate version of spaCy. We are using version 2.3.2, which is in the allowed range for the model release ( $\geq 2.3.0, < 2.4.0$ ) , so we do not need to modify this.

Finally, hit the `run` button. This will cause your configuration files to be updated and then will run your application. If everything has gone correctly, you should see the following in the output pane once it completes.



```

The
quick
brown
fox
jumps
over
the
lazy
dog
.

```

Image 12: *spaCy and the necessary components are all found as dependencies, so the application runs successfully*

## Extracting names from headlines using spaCy

We've now seen how to install common packages like `requests` simply by importing them, how to find and install slightly more complicated packages like `beautifulsoup` using the GUI package manager, and how to manually install even more complicated packages like `spaCy` (which have their own dependencies) by manually writing sections of the `pyproject.toml` file.

Let's put everything together and use all three packages to extract people's names from today's headlines. We'll use the plaintext version of CNN at [lite.cnn.com](https://lite.cnn.com)<sup>23</sup> as it's easier to extract text from.

Replace the code in your `main.py` with the following.

```

1 import spacy
2 import requests
3 from bs4 import BeautifulSoup
4 from collections import Counter
5
6 nlp = spacy.load("en_core_web_sm")
7 response = requests.get("http://lite.cnn.com/en")
8 soup = BeautifulSoup(response.text, "html.parser")
9
10 # https://stackoverflow.com/questions/1936466/beautifulsoup-grab-visible-webpage-text
11 [s.extract() for s in soup(['style', 'script', '[document]', 'head', 'title'])]
12 text = soup.getText()
13 doc = nlp(text)
14
15 names = []
16 for ent in doc.ents:
17     if ent.label_ == "PERSON":
18         names.append(ent.lemma_)
19
20 print("These people are in the headlines today")
21 print(Counter(names).most_common(10))

```

This pulls the HTML from the lite version of CNN, extracts the HTML, removes non-visible text such as CSS styles and JavaScript, and parses the resulting text using `spaCy`.

Then we loop through all of the `named entities`<sup>24</sup> that `spaCy` detects as part of its standard parse, and print out any that look like people.

If you run this code, you should see a list of people making headlines today. At the time of writing, John Lewis is mentioned in the most headlines. (Note that named entity recognition is a difficult

---

<sup>23</sup><https://lite.cnn.com>

<sup>24</sup>[https://en.wikipedia.org/wiki/Named\\_entity](https://en.wikipedia.org/wiki/Named_entity)

task and here spaCy considers the possessive form John Lewis' to be a separate entity. We can see that John Lewis was mentioned a total of 7 times though.)

The screenshot shows a Repl.it interface. On the left, the file structure includes 'main.py', 'poetry.lock', and 'pyproject.toml'. The 'main.py' code is as follows:

```

main.py
1 import spacy
2 import requests
3 from bs4 import BeautifulSoup
4 from collections import Counter
5
6 nlp = spacy.load("en_core_web_sm")
7 response = requests.get("http://lite.cnn.com/en")
8 soup = BeautifulSoup(response.text, "html.parser")
9
10 #
11 https://stackoverflow.com/questions/1936466/beautifulsoup-grab-visible-webpage-
12 text
13 [s.extract() for s in soup(['style', 'script', '[document]', 'head', 'title'])]
14 text = soup.getText()
15 doc = nlp(text)
16
17 names = []
18 for ent in doc.ents:
19     if ent.label_ == "PERSON":
20         names.append(ent.lemma_)
21
22 print("These people are in the headlines today")
23 print(Counter(names).most_common(10))

```

The right panel shows the output of the script: "These people are in the headlines today" followed by a list of names and their counts. A red arrow points to the output area.

Image 13: Using spaCy to extract people in today's news.

## Make it your own

If you followed along, you'll already have your own version of the repl to extend. If not, start from ours. Fork it from the embed below.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-03-nlp-spacy?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

spaCy is a very powerful NLP library and it can do far more than simply extract people's names. See what other interesting insights you can automatically extract from today's news.

Now you can use the Repl.it IDE, write programs that use files, and install third-party dependencies. Next up, we'll be taking a look at doing data science with Repl.it by visualising data using `matplotlib` and `seaborn`.

# Data science with Repl.it: Plots and graphs

So far, all the programs we have looked at have been entirely **text based**. They have taken text input and produced text output, on the console or saved to files.

While text is flexible and powerful, sometimes a picture is worth a thousand words. Especially when analysing data, you'll often want to produce plots and graphs. There are three main ways of achieving this using Repl.it.

1. Creating a front-end only project and using only JavaScript, HTML and CSS.
2. Creating a full web application with something like [Flask<sup>25</sup>](#), analysing the data in Python and passing the results to a front end to be visualised.
3. Using Python code only, creating windows using [X<sup>26</sup>](#) and rendering the plots in there.

Option 1 is great if you're OK with your users having access to all of your data, you like doing data manipulation in JavaScript, and your data set is small enough to load into a web browser. Option 2 is often the most powerful, but can be overkill if you just want a few basic plots.

Here, we'll demonstrate how to do option 3, using Python and [Matplotlib<sup>27</sup>](#).

## Installing Matplotlib and creating a basic line plot

Matplotlib is a third-party library for doing all kinds of plots and graphs in Python. We can install it by using Repl.it's "magic import" functionality. Matplotlib is a large and powerful library with a lot of functionality, but we only need `pyplot` for now: the module for plotting.

Create a new Python repl and add the following code.

```
1 from matplotlib import pyplot as plt  
2  
3 plt.plot([1,2,3,4,5,6], [6,3,6,1,2,3])  
4 plt.show()
```

---

<sup>25</sup><https://flask.palletsprojects.com/>

<sup>26</sup>[https://en.wikipedia.org/wiki/X\\_Window\\_System](https://en.wikipedia.org/wiki/X_Window_System)

<sup>27</sup><https://matplotlib.org/>

There are many traditions in the Python data science world about how to import libraries. Many of the libraries have long names and get imported as easier-to-type shortcuts. You'll see that nearly all examples import pyplot as the shorter plt before using it, as we do above. We can then generate a basic line plot by passing two arrays to plt.plot() for X and Y values. In this example, the first point that we plot is (1, 6) (the first value from each array). We then add all of the plotted points joined into a line graph.

Repl.it knows that it needs an X server to display this plot (triggered when you call plt.show()), so after running this code you'll see "Starting X" in the main output console and a new graphical window will appear.

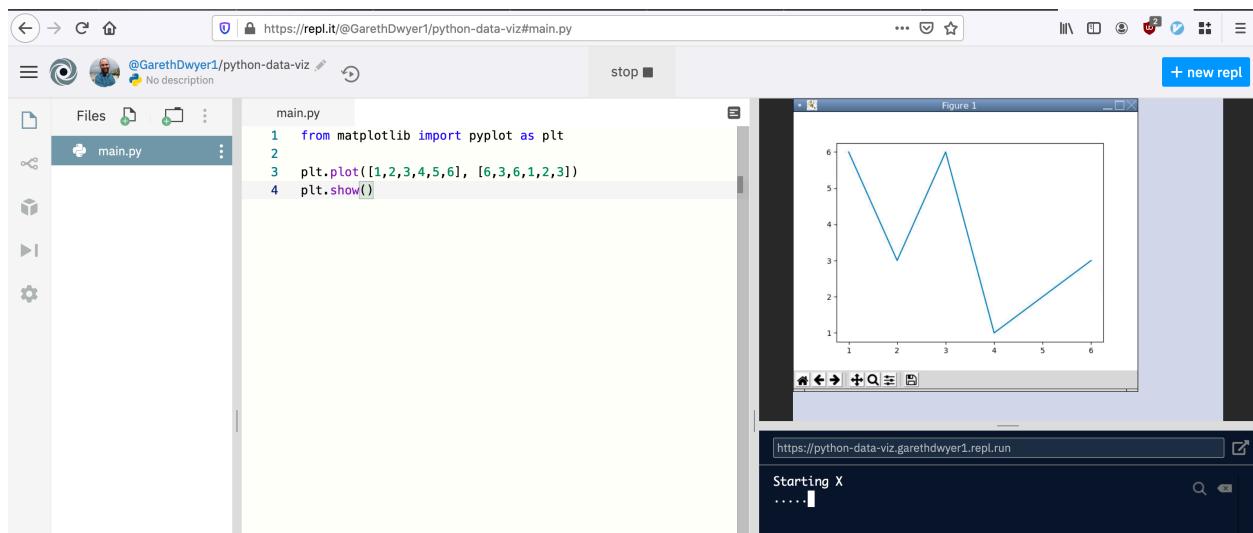


Image 1: We can plot a basic line plot by passing in the X and Y values

The X server is very limited compared to a full operating system GUI. Beneath the plot, you'll see some controls to pan and zoom around the image, but if you try to use them you'll see that the experience is not that smooth.

Line plots are cool, but we can do more. Let's plot a real data set.

## Making a scatter plot of US cities by state

Scatter plots are often used to plot 2D data and look for correlations and other patterns. However, they can also loosely be used to plot geographical X-Y coordinates (in reality, the field of plotting geographical points is [far more complicated<sup>28</sup>](#)). We'll use a subset from the [city data from simplemaps<sup>29</sup>](#) to generate our next plot. Each row of the data set represents one city in the USA, and gives us its latitude, longitude, and two-letter state code.

To download the data and plot it, replace the code in your main.py file with the following.

<sup>28</sup><https://www.gislounge.com/what-is-gis/>

<sup>29</sup><https://simplemaps.com/data/us-cities>

```
1 from matplotlib import pyplot as plt
2 import requests
3 import random
4
5 data_url = "https://raw.githubusercontent.com/sixhobbits/ritza/master/data/us-cities\
6 .txt"
7
8 r = requests.get(data_url)
9
10 with open("us-cities.txt", "w") as f:
11     f.write(r.text)
12
13 lats = []
14 lons = []
15 colors = []
16 state_colors = {}
17
18 # matplotlib uses single letter shortcuts for common colors
19 # blue, green, red, cyan, magenta, yellow, black
20 all_colors = ['b', 'g', 'r', 'c', 'm', 'y', 'k']
21
22 with open("us-cities.txt") as f:
23     for i, line in enumerate(f):
24         state, lat, lon = line.split()
25         lats.append(float(lat))
26         lons.append(float(lon))
27
28         # we assign each state a random colour, but once we've picked
29         # a colour we always use it for all cities in that state.
30         if state not in state_colors:
31             state_colors[state] = random.choice(all_colors)
32             colors.append(state_colors[state])
33 plt.scatter(lons, lats, c=colors)
34 plt.show()
```

If you run this, you'll notice it takes a little bit longer than the six point plot we created before, as it now has to plot nearly 30 000 data points. Once it's done, you should see something similar to the following (though, as the colours were chosen randomly, yours might look different).

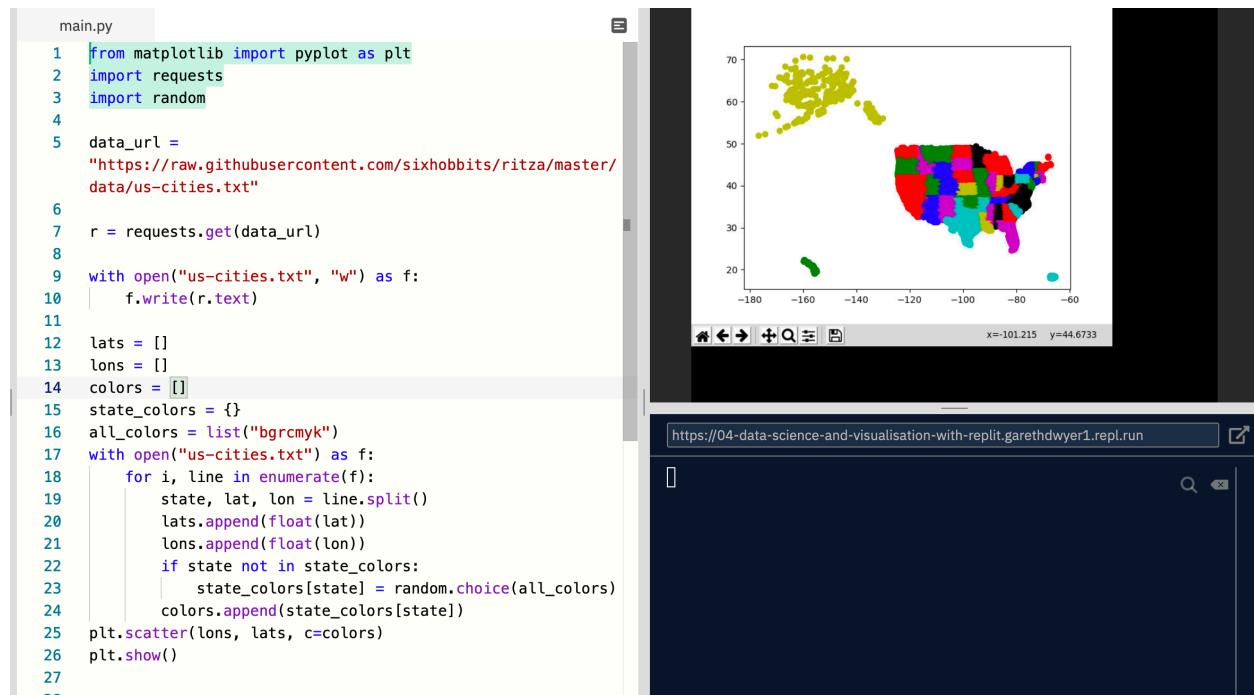


Image 2: All the cities in the US plotted by state as a scatterplot

You'll also notice that while it's recognisable as the US, the proportions are not right. Mapping a 3D sphere to a 2D plane is surprisingly difficult and there are many different ways of doing it.

## More advanced plotting with seaborn and pandas

Plotting X-Y points is a good start, but in most cases you'll want to do a little bit more. [seaborn<sup>30</sup>](#) is a plotting library built on top of Matplotlib that makes it easier to create good-looking visualisations.

Let's do another scatter plot based on GDP and life expectancy data to see if people live longer in richer countries.

Replace the code in `main.py` with the following. Remember how we mentioned earlier that data scientists have traditions about how to import certain libraries? Here you see a few more of these "short names". We'll use `seaborn` for plotting but import it as `sns`, `pandas31` for reading the CSV file but import it as `pd` and `NumPy32` for calculating the correlation but import it as `np`.

<sup>30</sup><https://seaborn.pydata.org/>

<sup>31</sup><https://pandas.pydata.org/>

<sup>32</sup><https://numpy.org/>

```
1 import requests
2 import seaborn as sns
3 import pandas as pd
4 from matplotlib import pyplot as plt
5 import numpy as np
6
7 data_url = "https://raw.githubusercontent.com/sixhobbits/ritza/master/data/us-cities\
8 .txt"
9
10 r = requests.get(data_url)
11
12 with open("gdp-life.txt", "w") as f:
13     f.write(r.text)
14
15 df = pd.read_csv("gdp-life.txt")
16 print(df.head())
17
18 print("___")
19 print("The correlation is: ", np.corrcoef(df["gdpPercap"], df["lifeExp"])[0, 1])
20 print("___")
21
22 sns.lmplot("gdpPercap", "lifeExp", df).set_axis_labels(
23     "Life expectancy", "GDP per capita"
24 )
25
26 plt.title("People live longer in richer countries")
27 plt.tight_layout()
28 plt.show()
```

If you run this, you'll see it plots each country in a similar way to our previous scatter plot, but also adds a line showing the correlation.

In the output pane below you can also see that the correlation coefficient between the two variables is 0.67 which is a fairly strong positive correlation.

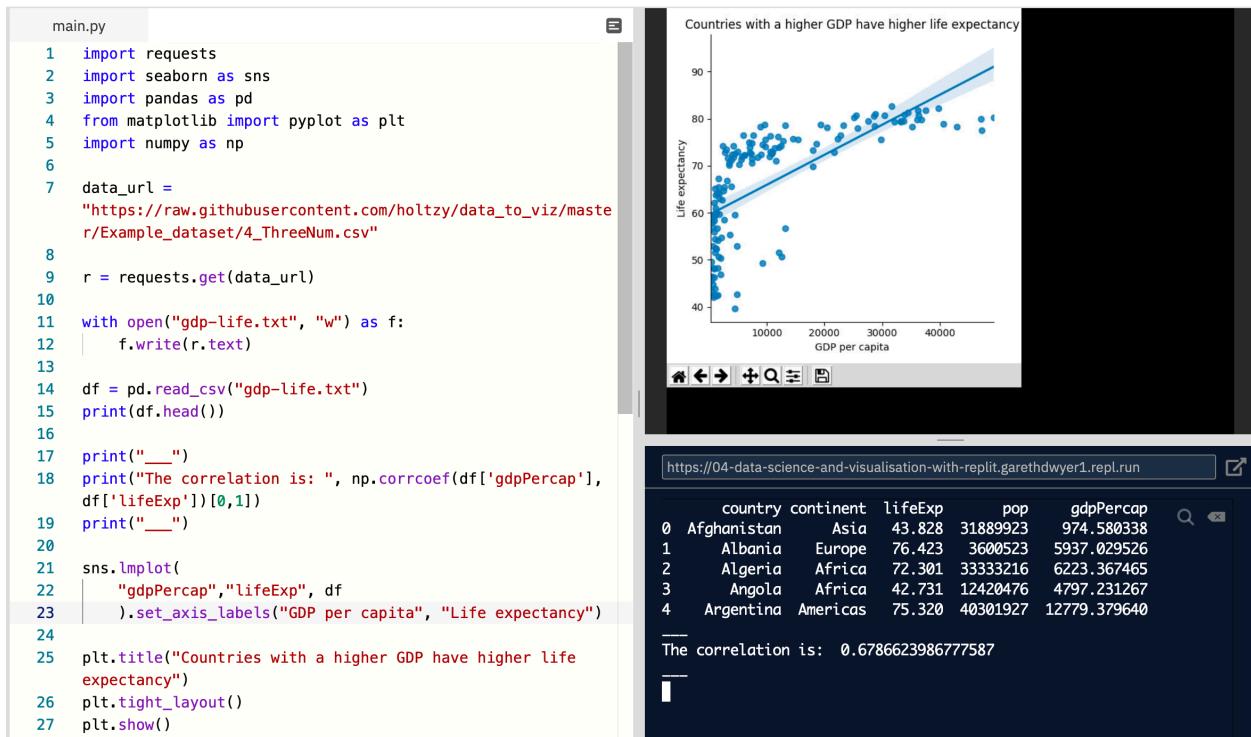


Image 3: Using `seaborn` to create a scatter plot with a best fit line to see correlation

Data science and data visualisation are huge topics, and there are dozens of Python libraries that can be used to plot data. For a good overview of all of them and their strengths and weaknesses, you should watch [Jake Vanderplas's talk](#)<sup>33</sup>.

## Saving plots to PNG files

While visualising data right after you analyse it is often useful, sometimes you need to save the figures to embed into reports. You can save your graphs by calling `plt.savefig()`. Change the last line (`plt.show()`) to

```
1 plt.savefig("GDPlife.png")
```

Rerun the code. Instead of seeing the plot appear in the right-hand pane, you'll see a new file in the files pane. Clicking on it will show you the PNG file in the editing pane.

<sup>33</sup><https://www.youtube.com/watch?v=FytuB8nFHPQ>

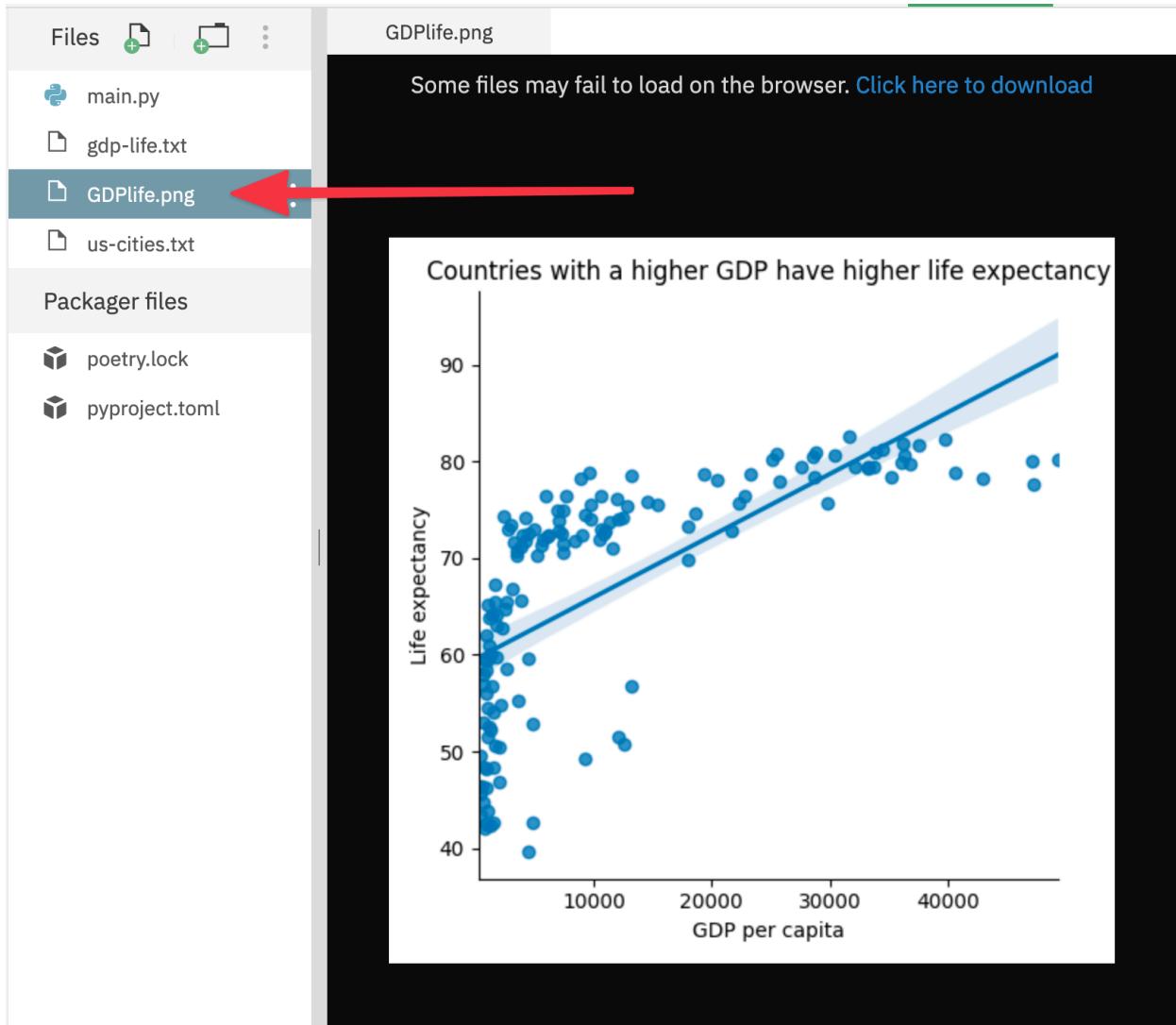


Image 4: Saving a PNG file for later use

## Make it your own

If you followed along, you'll already have your own version of the repl to extend. If not, start from ours. Fork it from the embed below.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-04-matplotlib-plotting?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

You've learned how to make some basic plots using Python and Repl.it. There are millions of freely available data sets on the internet, waiting for people to explore them. You can find many of these using [Google's Dataset Search<sup>34</sup>](#) service. Pick a topic that you're interested in and try to find out more about it through data visualisations.

Next up, we'll explore the mutiplayer functionality of Repl.it in more detail so that you can code collaboratively with friends or colleagues.

---

<sup>34</sup><https://datasetsearch.research.google.com/>

# Multiplayer: Pair programming with Repl.it

Software developers have a reputation for being loners, but they don't always code by themselves. [Pair programming<sup>35</sup>](#) is used by many programmers to

- Write bug-free code more efficiently (for example, one person might watch for mistakes while the other codes).
- Share knowledge (a less-experienced programmer might 'follow along' while a more experienced programmer develops something, learning from each step of the process).
- Assess expertise (if you're considering a new hire, watching them code first can be helpful to assess how good a coder they are, but coding with them allows you to also see their experience in teamwork and communication).

Pair programming intuitively sounds like it would be inefficient: after all, the two developers could instead be working on different projects simultaneously. But on top of catching more bugs, two people working together often display more creativity as well. You might think of an idea based on something your buddy said that wouldn't have come to you alone.

If you have a friend handy, work through this tutorial together to gain real pair programming experience. If you're alone, fire up two browsers (or use incognito mode) to sign into two Repl.it accounts simultaneously.

## Extending our data science article using pair programming: Getting help

Imagine that you are a developer who has come across the [previous tutorial on plotting and graphing<sup>36</sup>](#). You want to adapt the graphs shown a bit, but you haven't used Python much, so you decide to ask your friend for help.

In this scenario, you are "@Lean3Viljoen94" and the friend that you're asking for help is "@GarethDwyer1".

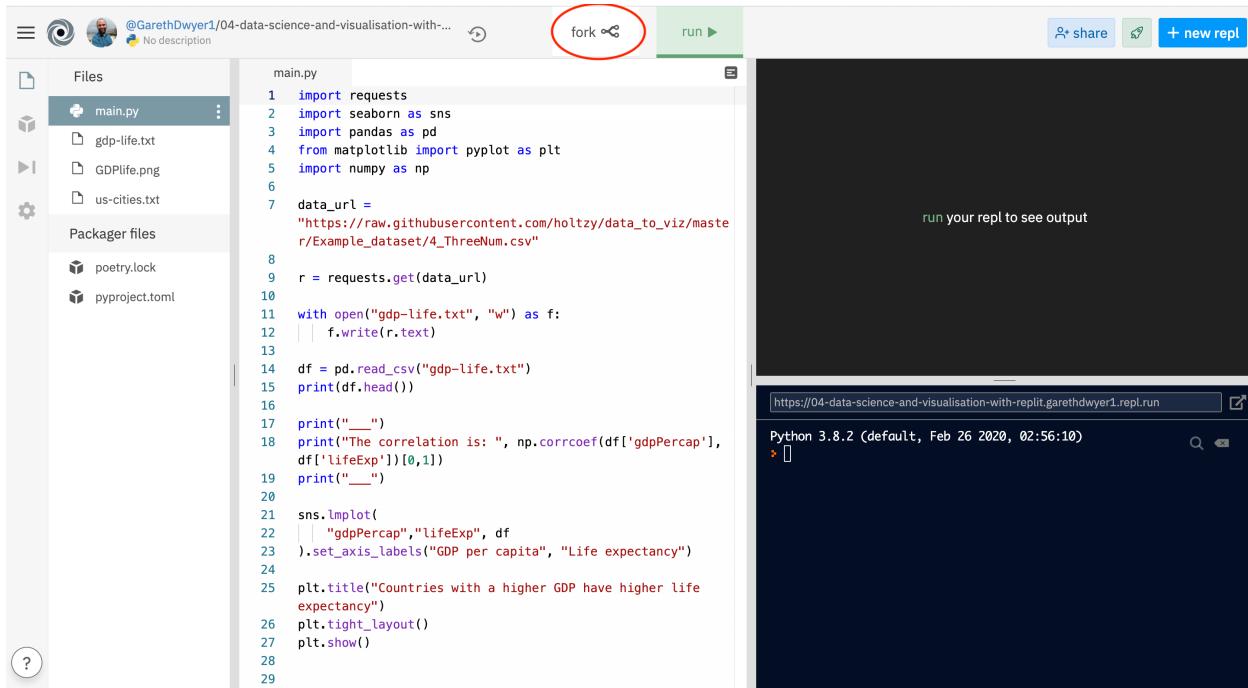
Start by forking [the data science repl<sup>37</sup>](#) and making sure that you can run it.

---

<sup>35</sup>[https://en.wikipedia.org/wiki/Pair\\_programming](https://en.wikipedia.org/wiki/Pair_programming)

<sup>36</sup><http://www.codewithrepl.it/04-data-science-and-visualisation-with-repl-it.html>

<sup>37</sup><https://repl.it/@GarethDwyer1/04-data-science-and-visualisation-with-replit>



The screenshot shows a Repl.it interface for a project titled '04-data-science-and-visualisation-with-repl'. The left sidebar lists files: main.py, gdp-life.txt, GDPLife.png, and us-cities.txt. Under 'Packager files', there are poetry.lock and pyproject.toml. The main area displays the code for main.py:

```

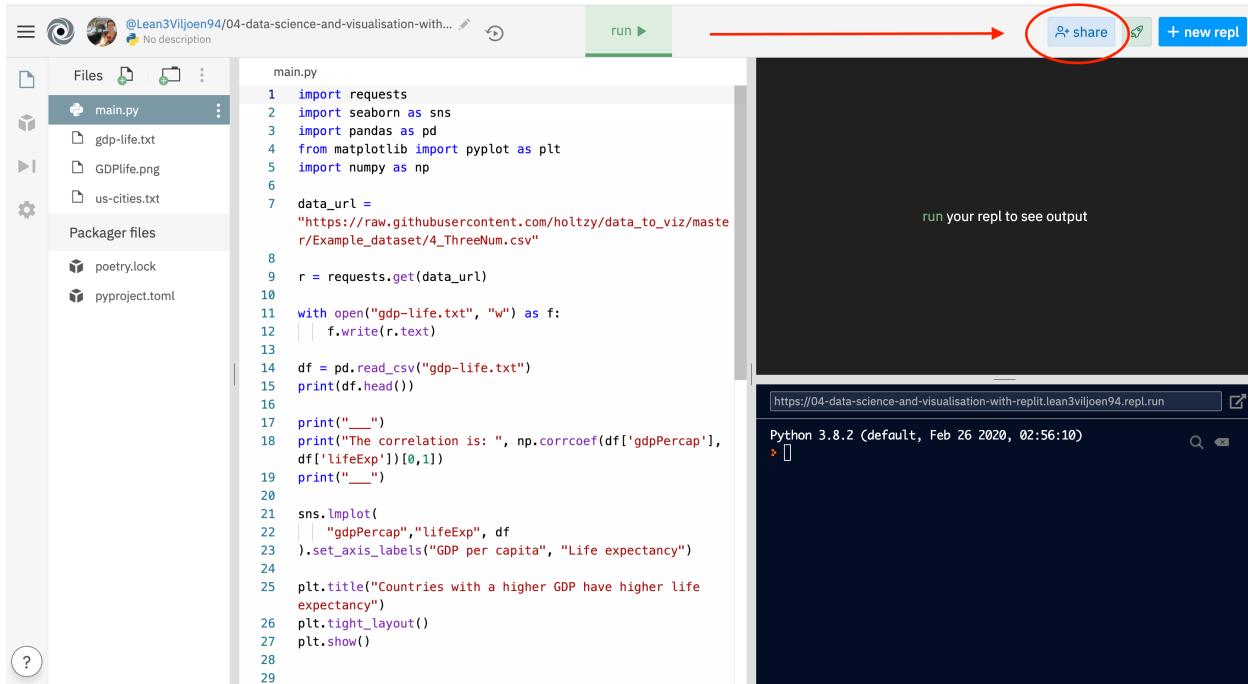
1 import requests
2 import seaborn as sns
3 import pandas as pd
4 from matplotlib import pyplot as plt
5 import numpy as np
6
7 data_url =
8 "https://raw.githubusercontent.com/holtzy/data_to_viz/master/Example_dataset/4_ThreeNum.csv"
9 r = requests.get(data_url)
10
11 with open("gdp-life.txt", "w") as f:
12     f.write(r.text)
13
14 df = pd.read_csv("gdp-life.txt")
15 print(df.head())
16
17 print("___")
18 print("The correlation is: ", np.corrcoef(df['gdpPerCap'],
19 df['lifeExp'])[0,1])
19 print("___")
20
21 sns.lmplot(
22     "gdpPerCap", "lifeExp", df
23 ),set_axis_labels("GDP per capita", "Life expectancy")
24
25 plt.title("Countries with a higher GDP have higher life
expectancy")
26 plt.tight_layout()
27 plt.show()
28
29

```

The right side shows a dark terminal window with the URL <https://04-data-science-and-visualisation-with-repl.garethdwyer1.repl.run> and the Python version Python 3.8.2 (default, Feb 26 2020, 02:56:10). A red circle highlights the 'fork' button at the top right of the interface.

Image 1: Forking another user's project

Now from your own fork, press the share button, as shown below.

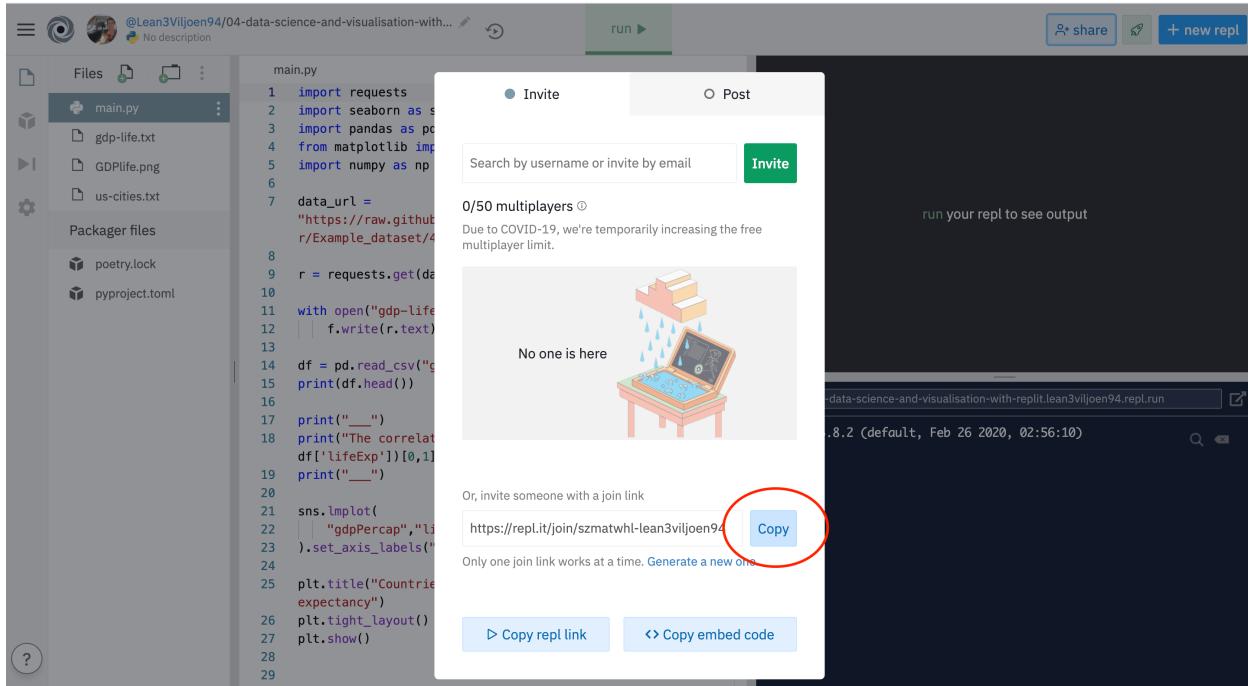


The screenshot shows a Repl.it interface for a project titled '04-data-science-and-visualisation-with-repl'. The left sidebar lists files: main.py, gdp-life.txt, GDPLife.png, and us-cities.txt. Under 'Packager files', there are poetry.lock and pyproject.toml. The main area displays the same code as in Image 1. The right side shows a dark terminal window with the URL <https://04-data-science-and-visualisation-with-repl.lean3viljoen94.repl.run> and the Python version Python 3.8.2 (default, Feb 26 2020, 02:56:10). A red arrow points to the 'share' button at the top right of the interface.

Image 2: Sharing your project with another user

Copy the invite link, and note that this is different from the normal link to your repl. If you copy the link from your URL bar, you can give people read access to your repl, but by copying the invite

link you'll give them write access.



**Image 3: Sharing options modal**

If you knew your friend's Repl.it username or the email associated with the Repl.it account, you could instead use the Invite box at the top. Share the link with your friend and wait for them to join.

As soon as they do, you will see that a chat box pops up in the bottom right corner. Their profile picture or letter will be at the top of the chat box, so you can always know who is currently active.

Remember, you forked the repl in a previous step, so you are the owner of this fork and the "host" of this multiplayer session. If you invite multiple people and then leave, they can continue collaborating without you, but they won't be able to rejoin if the host is no longer in the session.

You can use the team chat feature, as shown below.

The screenshot shows the Repl.it interface. On the left, the code editor displays a Python script named `main.py`. The code imports libraries like requests, seaborn, pandas, matplotlib, and numpy, reads data from files and URLs, and performs a correlation analysis between GDP per capita and life expectancy. On the right, there's a terminal window showing a Python session running on version 3.8.2. A red box highlights the terminal area, which contains a chat window. The chat shows two users: `garethdwyer3` and `Lean3Viljoen94`. `garethdwyer3` says "hey" and `Lean3Viljoen94` replies "Hi! Thanks for joining! I'm a bit stuck and would appreciate your help." Below the chat is a text input field with "Start typing here..." placeholder text.

```

main.py
1 import requests
2 import seaborn as sns
3 import pandas as pd
4 from matplotlib import pyplot as plt
5 import numpy as np
6
7 data_url =
8 "https://raw.githubusercontent.com/holtzy/data_to_viz/master/Example_dataset/4_ThreeNum.csv"
9 r = requests.get(data_url)
10
11 with open("gdp-life.txt", "w") as f:
12     f.write(r.text)
13
14 df = pd.read_csv("gdp-life.txt")
15 print(df.head())
16
17 print("___")
18 print("The correlation is: ", np.corrcoef(df['gdpPerCap'],
19 df['lifeExp'])[0,1])
19 print("___")
20
21 sns.lmplot(
22     "gdpPerCap","lifeExp", df
23 ),set_axis_labels("GDP per capita", "Life expectancy")
24
25 plt.title("Countries with a higher GDP have higher life
26 expectancy")
26 plt.tight_layout()
27 plt.show()
28
29

```

Image 4: Starting chat with another user

In the previous tutorial, we looked at GDP by country. Imagine that you are now interested in how this is broken down by continent too. You still want to plot each country as a separate data point, but you want them in different colours, one for each continent. You're not sure how to do this, so you ask for help.

You can see a typing indicator to help decide if you should wait around for a reply or go make coffee.

The screenshot shows a Repl.it session. On the left, the code editor displays `main.py` with Python code for reading data from URLs and plotting GDP per capita versus life expectancy. In the center, a scatter plot titled "Countries with a higher GDP have higher life expectancy" shows a positive correlation. On the right, a chat interface shows two users, Lean3Viljoen94 and garethdwyer3, discussing how to visualize data by continent.

```

main.py
1 import requests
2 import seaborn as sns
3 import pandas as pd
4 from matplotlib import pyplot as plt
5 import numpy as np
6
7 data_url =
8 "https://raw.githubusercontent.com/holtzy/data_to_viz/master/Example_dataset/4_ThreeNum.csv"
9 r = requests.get(data_url)
10 with open("gdp-life.txt", "w") as f:
11     f.write(r.text)
12
13 df = pd.read_csv("gdp-life.txt")
14 print(df.head())
15
16 print("___")
17 print("The correlation is: ", np.corrcoef(df['gdpPerCap'],
18 df['lifeExp'])[0,1])
19 print("___")
20
21 sns.lmplot(
22     "gdpPerCap","lifeExp", df
23 ),set_axis_labels("GDP per capita", "Life expectancy")
24
25 plt.title("Countries with a higher GDP have higher life
expectancy")
26 plt.tight_layout()
27 plt.show()
28
29

```

**Countries with a higher GDP have higher life expectancy**

life expectancy

GDP per capita

x=32643.6 y=60.9375

2/2 online feedback

Lean3Viljoen94 I want to visualise GDP / life expectancy of the continents instead of all countries. How would you suggest I do this?

garethdwyer3 you could group the data by continent and colour each point based on that

garethdwyer3 is typing...

Start typing here...

Image 5: Chat box showing that user is typing

Your friend tells you about the `hue` argument and points out that you already have this data in the `continent` column in your data frame. You add `hue="continent"` to the graph and re-run it, but it doesn't quite work out how you expect.

The screenshot shows a Repl.it environment. On the left, the file structure includes `main.py`, `gdp-life.txt`, `GDPlife.png`, and `us-cities.txt`. The `main.py` file contains code for reading data from a URL and plotting GDP vs Life expectancy by continent. A red circle highlights the line `hue="continent"` in the `sns.lmplot` call. The central area displays a scatter plot titled "Countries with a higher GDP have higher life expectancy". The x-axis is "GDP per capita" and the y-axis is "Life expectancy". Data points are colored by continent: Asia (blue), Europe (orange), Africa (green), Americas (red), and Oceania (purple). A regression line and shaded confidence interval are shown. Below the plot is a table of data. On the right, a chat window shows two users, Lean3Viljoen94 and garethdwyer3, discussing the code and its output.

```

main.py
1 import numpy as np
2
3 data_url =
4     "https://raw.githubusercontent.com/holtzy/data_to_viz/master/Example_dataset/4_ThreeNum.csv"
5
6 r = requests.get(data_url)
7
8 with open("gdp-life.txt", "w") as f:
9     f.write(r.text)
10
11 df = pd.read_csv("gdp-life.txt")
12 print(df.head())
13
14 print("___")
15 print("The correlation is: ", np.corrcoef(df['gdpPerCap'], df['lifeExp'])[0,1])
16 print("___")
17
18 sns.lmplot(
19     "gdpPerCap", "lifeExp", df, hue="continent"
20 ).set_axis_labels("GDP per capita", "Life expectancy")
21
22 plt.title("Countries with a higher GDP have higher life expectancy")
23 plt.tight_layout()
24 plt.show()
25
26
27
28
29
30
31
32

```

	country	continent	lifeExp	pop
0	Afghanistan	Asia	43.828	3188
1	Albania	Europe	76.423	366
2	Algeria	Africa	72.301	333
3	Angola	Africa	42.731	124
4	Argentina	Americas	75.320	4930

The correlation is: 0.678662398677758

Start typing here...

2/2 online feedback

Lean3Viljoen94: that continent is a categorical variable and assign a colour to each item in the group

OK cool, let me try that quick..

garethdwyer3: add it right after the 'df' argument

Image 6: *Changing the plot from grouping data by country to grouping by continent*

Your friend suggests maybe a scatter plot without the correlation line might look better, but when you try that it results in an error. The error message is hidden by the chat box, so you move it to the other side of the screen.

The screenshot shows the Repl.it interface. On the left is a file explorer with files like main.py, gdp-life.txt, GDPLife.png, us-cities.txt, and poetry.lock. The main area is a code editor with the following Python script:

```

main.py
1 import requests
2 import seaborn as sns
3 import pandas as pd
4 from matplotlib import pyplot as plt
5 import numpy as np
6
7 data_url =
8 "https://raw.githubusercontent.com/holtzy/data_to_viz/master/Example_dataset/4_ThreeNum.csv"
9 r = requests.get(data_url)
10
11 with open("gdp-life.txt", "w") as f:
12     f.write(r.text)
13
14 df = pd.read_csv("gdp-life.txt")
15 print(df.head())
16
17 print("___")
18 print("The correlation is: ", np.corrcoef(df['gdpPercap'],
19 df['lifeExp'])[0,1])
20 print("___")
21
22 ax = sns.scatterplot(
23     "gdpPercap", "lifeExp", df, hue="continent"
24 )
25
26 ax.set_axis_labels("GDP per capita", "Life expectancy")
27
28 plt.title("Countries with a higher GDP have higher life
29 expectancy")
30 plt.tight_layout()
31 plt.show()

```

The terminal window on the right shows the output of the script, including a correlation value and a traceback for a scatterplot error.

The chat interface on the left has two messages:

- "Oops, that didn't work as I expected?" by @Lean3Viljoen94
- "oh scatterplot works a bit differently to lmplot" by garethdwyer3

Below the chat is a message from Lean3Viljoen94: "let me make some changes quick". A red arrow points from the top of the chat box to the left edge of the IDE.

Image 7: You can move the chat box to the left of the IDE to see errors better

This is getting a bit more complicated than you bargained for. Sometimes showing is easier than telling, so your friend starts editing the code directly instead of telling you how to do so using chat. The code

```

1 sns.scatterplot(
2     "gdpPercap", "lifeExp", df, hue="continent"
3 ).set_axis_labels("GDP per capita", "Life expectancy")

```

changes to

```

1 ax = sns.scatterplot(
2     "gdpPercap", "lifeExp", df, hue="continent"
3 )
4 ax.set(xlabel="GDP per capita", ylabel="Life expectancy")

```

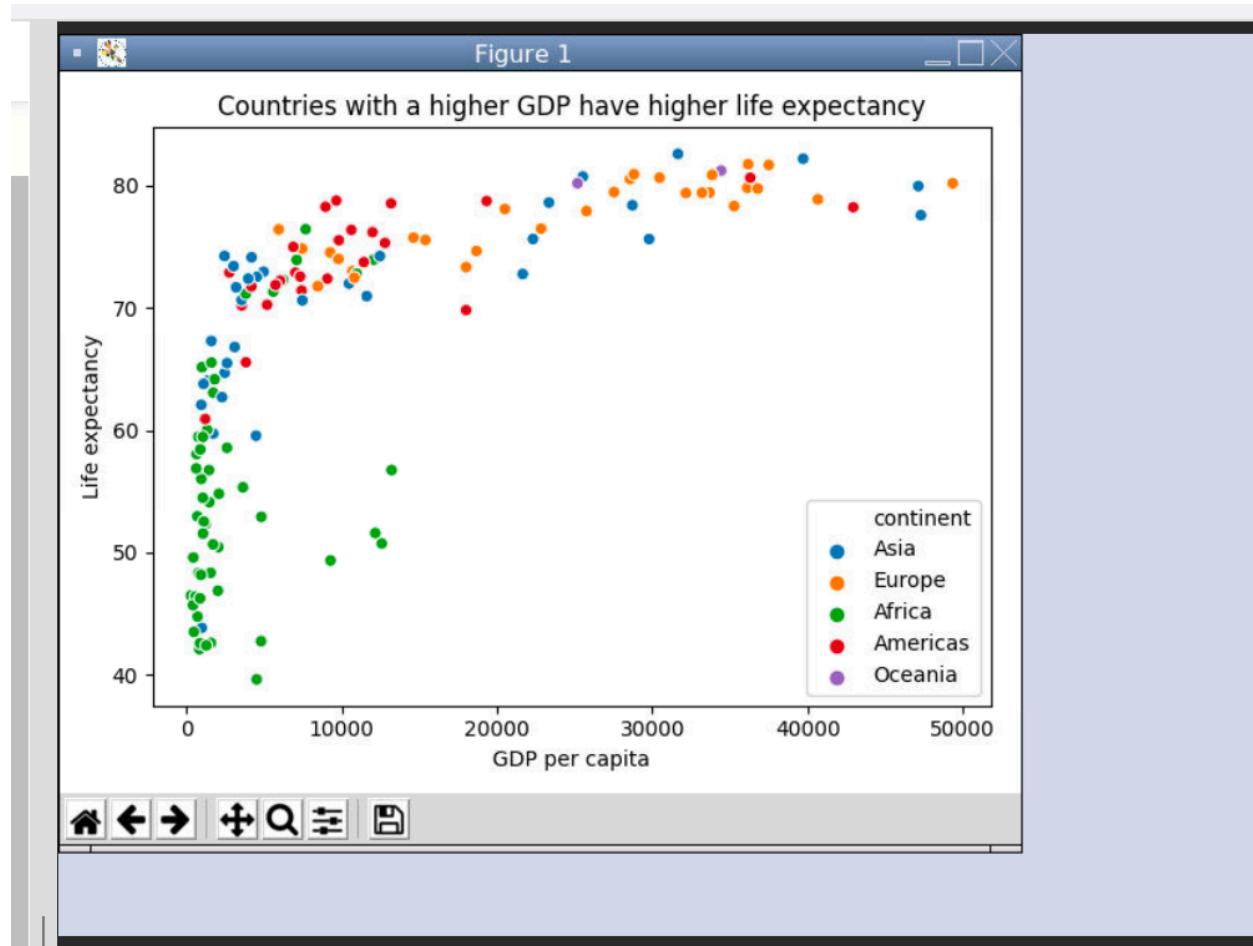


Image 8: In our new plot, we can see that African countries tend to have low life expectancy and low GDP, but the correlation looks weaker for the other continents

## Make it your own

If you followed along, you'll already have your own version of the repl to extend. If not, start from ours. Fork it from the embed below.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-05-multiplayer?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

Getting help on a single file in a program is only one use for multiplayer, but there are many scenarios where it can be useful to see your teammates' changes in real time. For example, if you're a back-end developer you could work closely with a front-end developer, ironing out any issues with

data communication between the back- and front-end in real time, instead of waiting for multiple iterations of several days.

That brings us to the end of part 1 of this series and you should now be familiar with all of the basic features of Repl.it.

In part 2, we'll cover more advanced features, such as running projects from GitHub, storing secrets securely, and productivity hacks.

# Repl.it and GitHub: Using and contributing to open-source projects

You've probably heard of [GitHub<sup>38</sup>](#), which hosts millions of coding projects that you can use or learn from.

In this tutorial, we'll see how to:

- Import open-source projects from GitHub to Repl.it so that we can run them or modify them
- Integrate your own GitHub account with your Repl.it account so that you can work on your private projects
- Push changes back to open-source projects as pull requests.

We'll use a basic Flask `hello world` app for the demonstrations. You can use this same project to follow along, or pick any other project on GitHub that interests you.

## Importing a project from GitHub and running it on Repl.it

We'll use the [Flask<sup>39</sup>](#) application available at

---

<sup>38</sup><https://github.com>

<sup>39</sup><https://flask.palletsprojects.com/en/1.1.x/>

<sup>40</sup><https://github.com/ritza-co/flask-hello-world>

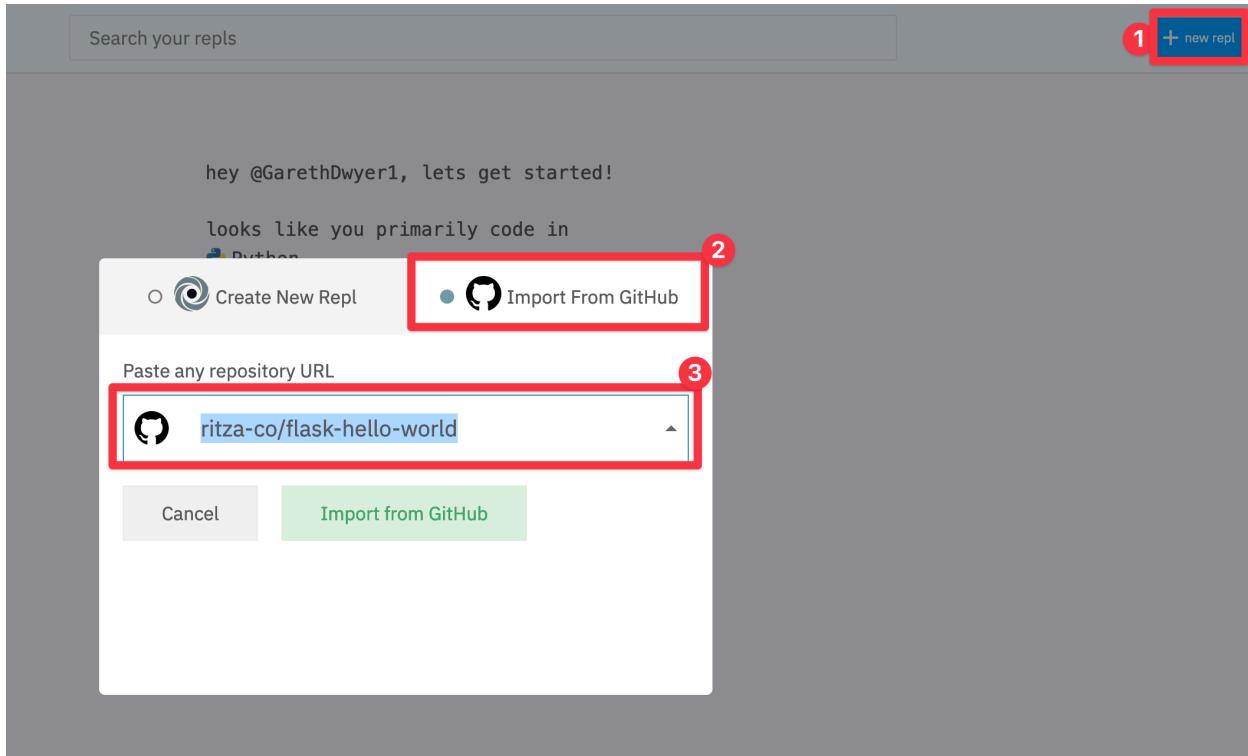


Image 1: Importing a repository from GitHub to Repl.it.

Press the green `Import from GitHub` button and you'll see Repl.it clone the repository and turn it into a repl. In all of our previous projects, we used the `main.py` file that Repl.it automatically creates for all new Python projects, and which it runs automatically when you press the `run` button. Note how in this GitHub project, we have no `main.py` file, and our code is instead in `mydemoapp.py`. Therefore, Repl.it will need some help from you to define how to run the project. This is configured through another special file named `.replit`. Because there was no `main.py` file, Repl.it automatically created this file and will prompt you to configure it.

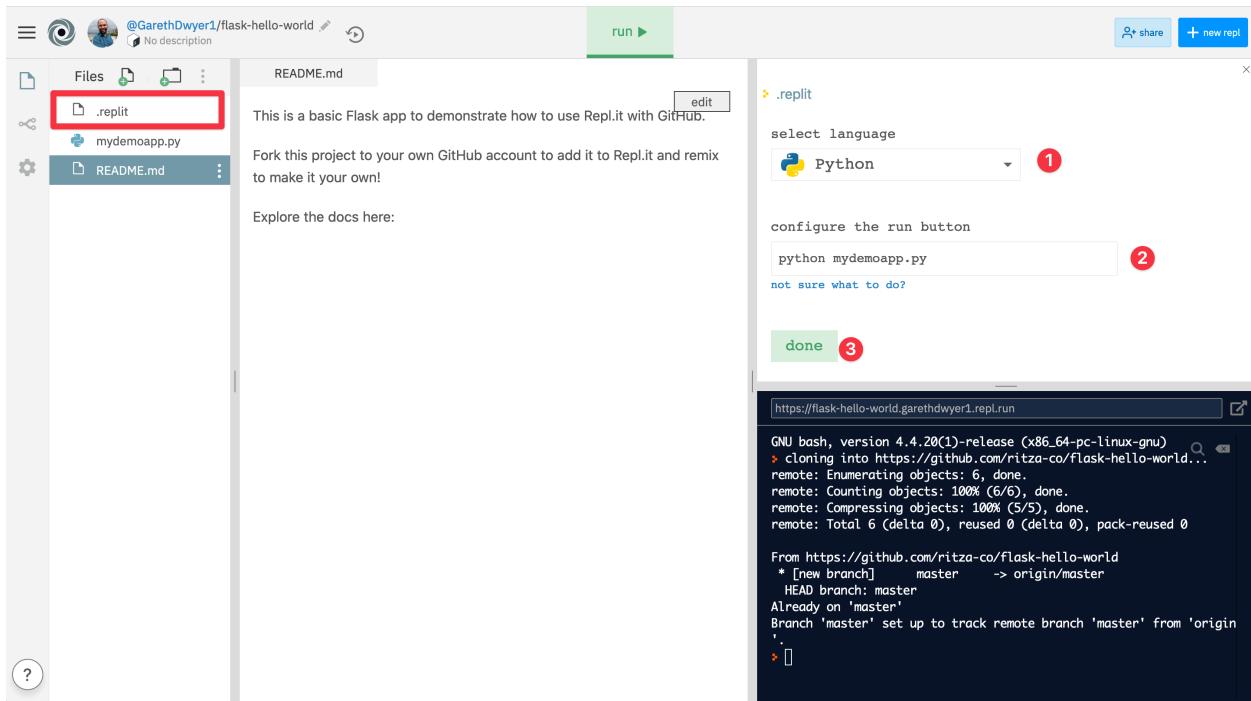
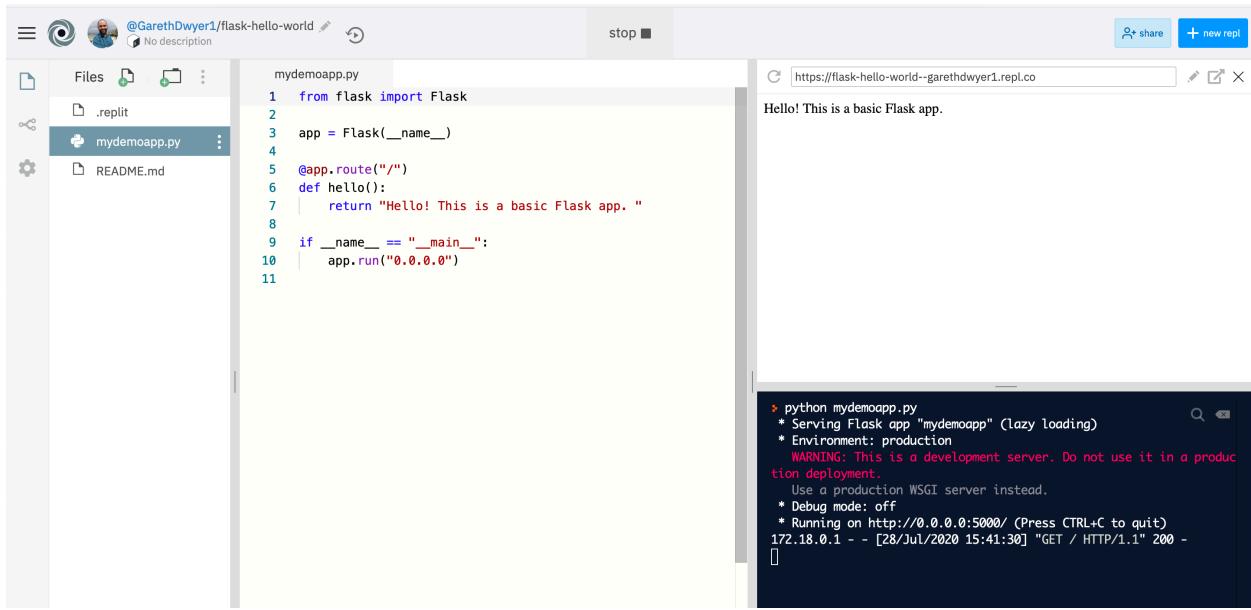


Image 2: Adding a `.replit` file to indicate how the project should be run.

Select the language (Python) from the first dropdown and type `python mydemoapp.py` in the “configure the run button” input. Every time you press the run button, Repl.it will execute the command given here. If you prefer, you can also edit the `.replit` file directly. If you click on it, you’ll see it now contains the following configuration, which matches what we provided through the GUI panel.

```
1 language = "python3"
2 run = "python mydemoapp.py"
```

If you hit the run button, you should see the app start. As you can see, the web application is very basic: all it can do is display a welcome message. If the configuration panel doesn’t pop up automatically, you can manually create a file called `.replit` and add the configuration above to get the same result.

A screenshot of the Repl.it web interface. On the left, there's a sidebar with a user icon, the repository name '@GarethDwyer1/flask-hello-world', and a 'No description' link. Below that is a 'Files' section showing a .replit file, mydemoapp.py, and a README.md file. The main area shows the code for mydemoapp.py:

```
1 from flask import Flask
2
3 app = Flask(__name__)
4
5 @app.route("/")
6 def hello():
7     return "Hello! This is a basic Flask app."
8
9 if __name__ == "__main__":
10     app.run("0.0.0.0")
```

To the right, there's a preview window showing the output of the application: "Hello! This is a basic Flask app.". Below that is a terminal window showing the command `python mydemoapp.py` being run, followed by the Flask application's startup logs.

Image 3: *Running the Flask application on Repl.it.*

Some GitHub projects are very large and complicated, and you might not be able to run everything you need directly on Repl.it, but in many cases it just works. Open-source projects can be read and run by anyone, but still have restrictions on who can push changes to them. Next we'll improve this project and request that the owner merges our changes into the original.

## Looking at the version control panel in Repl.it

Repl.it includes a version control tab which shows you information about the GitHub repository and in some cases allows you to push your changes made in Repl.it back to GitHub.

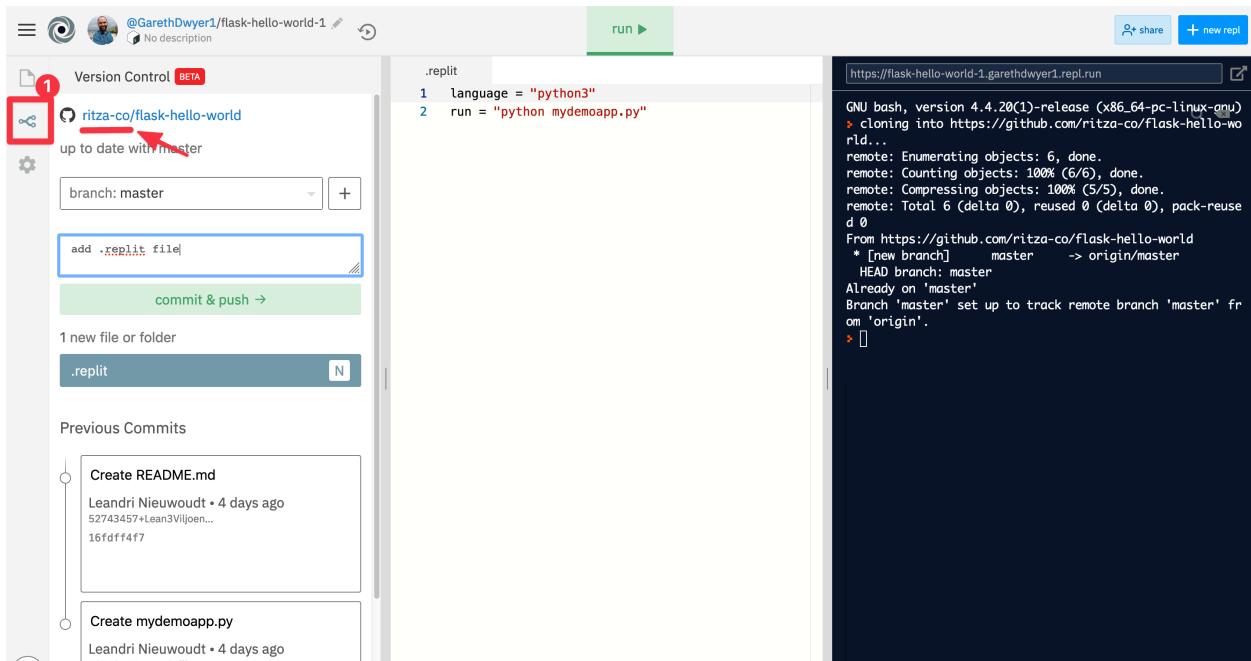


Image 4: Viewing details about the repository in the version control tab.

If you select this version control tab from the menu on the left, you'll see a summary of the linked repository. Note that it's already figured out what changes we've made, and it shows that the `.replit` file is new. It would be nice for other people who use this repository with GitHub to have the file automatically, so we might want to push the changes we made back to GitHub.

Note that the owner of the repository is `ritza-co` though, so you won't have write permissions for this repository. If you press the `commit & push` button, you'll see an error as shown below.

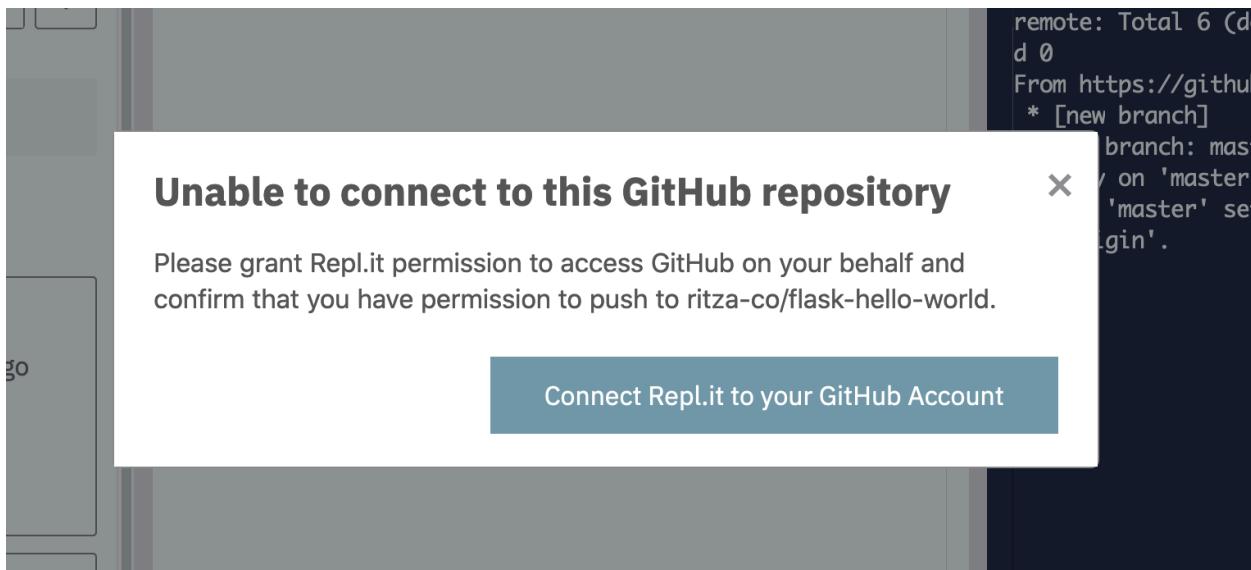


Image 5: You need permission to push to repositories on GitHub.

## Forking the project to your own GitHub account

Usually when contributing to open-source projects, you'll first create a "fork" of the original project. This means that you make your own version of the project and, as it's yours, you can make any changes to it that you want. If you think these changes would be useful to others too and are an obvious improvement over the original project, you can make a "pull request", which asks the owner of the original project to merge your changes into the main canonical project.

Create an account on [github.com](https://github.com)<sup>41</sup> or log in to your existing one and navigate back to the original project (<https://github.com/ritza-co/flask-hello-world>). In the top right corner, press the Fork button to create a copy of the project in your own GitHub account.

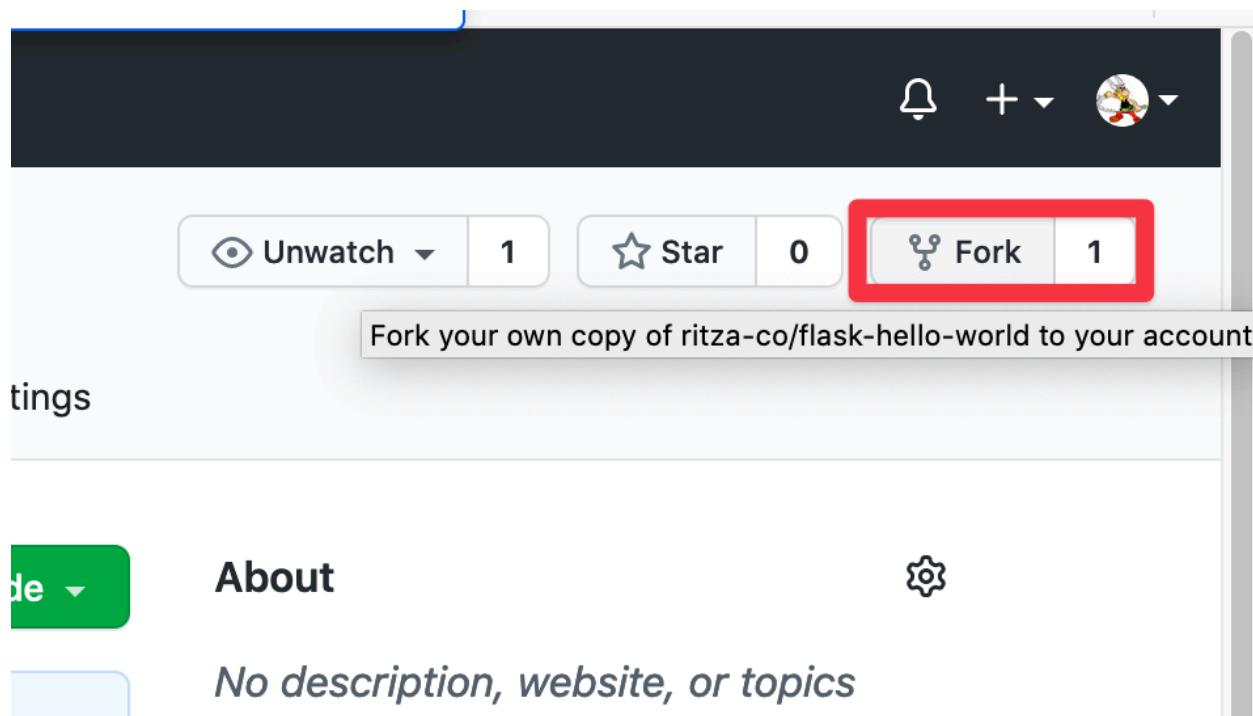


Image 6: Forking a repository in GitHub.

You should be taken to a new page in GitHub that looks very similar to the old one but which is owned by your own GitHub username. My GitHub username is `sixhobbits` so the new URL for me is <https://github.com/sixhobbits/flask-hello-world> (but yours will be different).

Now, instead of cloning the original project into Repl.it, create a new repl and import this fork of the project instead. Instead of going through the import UI again, you can also just create and load the relevant import URL. These URLs are in the format `https://repl.it/github/<githubproject>` so in my case I open <https://repl.it/github/sixhobbits/flask-hello-world> in my browser (you need to substitute your own GitHub username for this to work).

Configure the `.replit` file again and open the version control tab, as before. Under "what did you

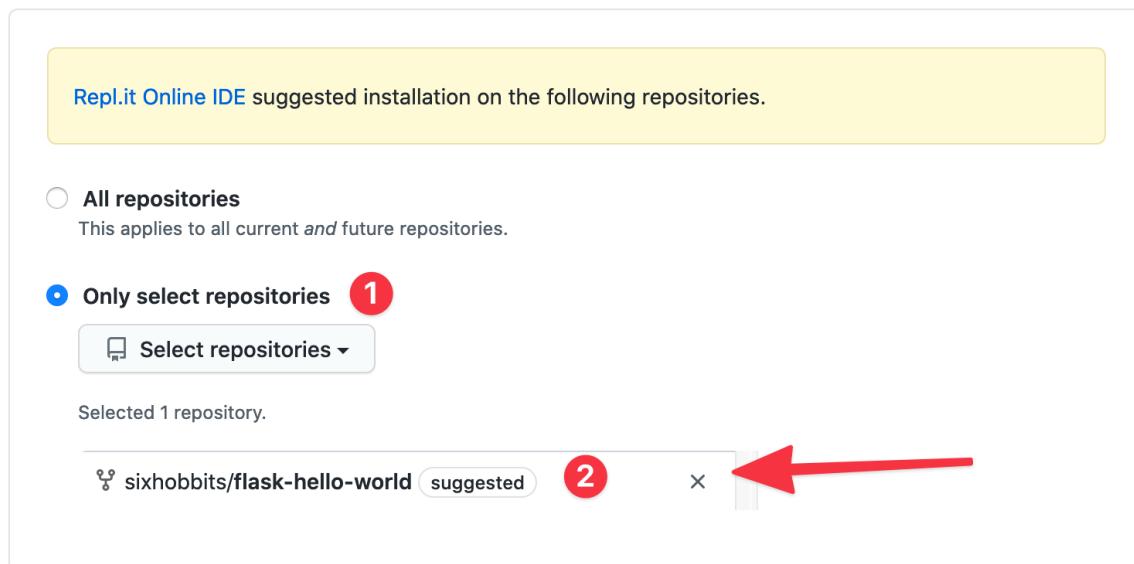
<sup>41</sup><https://github.com>

change” enter “add .repl file” or a similar message to describe what contributions you’re making, and press commit and push.

You’ll see the error again and be presented with the option to connect your Repl.it account to GitHub to prove that you authorise Repl.it to make these changes to GitHub on your behalf.

You can give Repl.it access to all of your repositories (useful if you want to use this integration a lot), but by default it will only get permission for the specific repository that we’re working with.

## Repository access



## Danger zone



Image 7: Giving Repl.it permission to access your GitHub data.

Press the green approve button and you’ll be directed back to Repl.it. Press the commit & push button again on Repl.it and this time everything should work without any errors.

Navigate back to your fork of the GitHub project, and you should see that the changes are reflected in GitHub too.

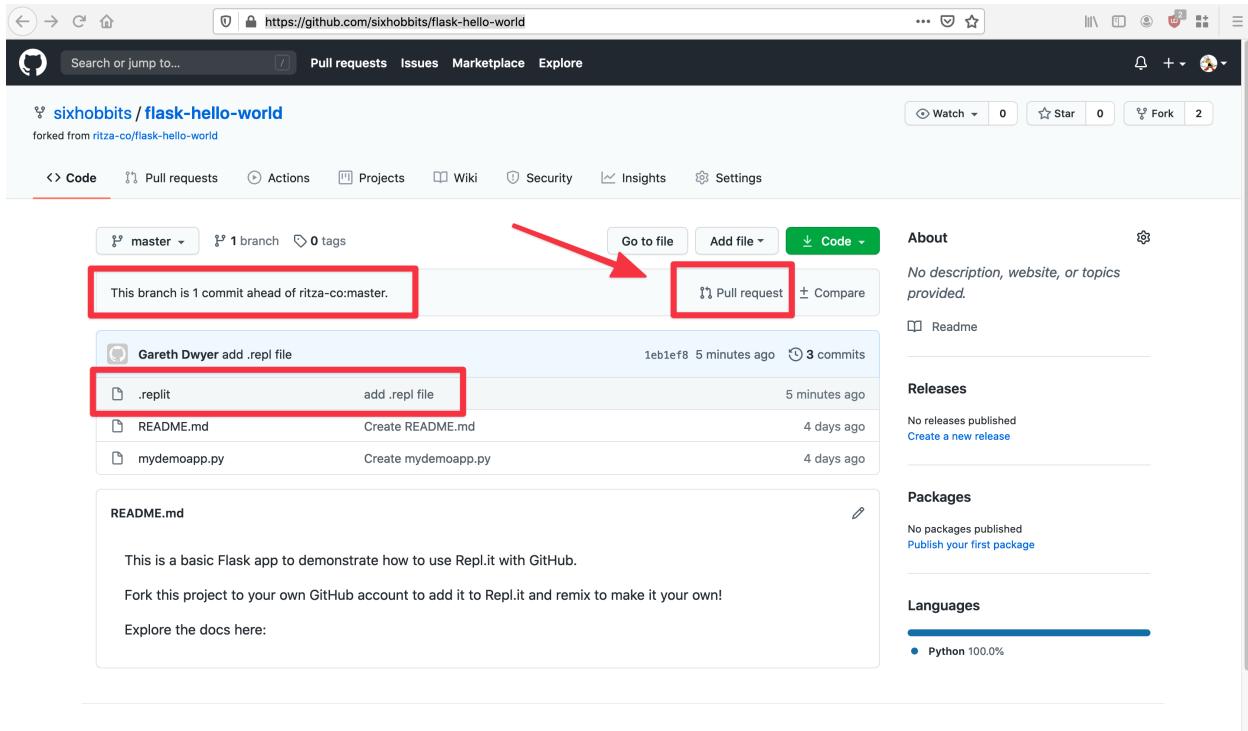


Image 8: *Seeing our changes reflected in our GitHub fork.*

As you can see, the new `.replit` file is visible and GitHub prompts you to make a pull request back into the original repository. Press `Pull request`, `Create pull request`, add a comment explaining why your changes should be merged into the original repository, and click `Create pull request` again.

## Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also [compare across forks](#).

The screenshot shows the GitHub interface for creating a pull request. At the top, it says 'base repository: ritza-co/flask-hello-world' and 'head repository: sixhobbits/flask-hello-world'. The 'base' dropdown is set to 'master'. The 'head' dropdown is set to 'master'. A green checkmark indicates 'Able to merge'. Below this, there's a text input field with the placeholder 'Add a .replit file to allow this repository to run seamlessly on https://repl.it |'. This input field is highlighted with a red border. There are tabs for 'Write' and 'Preview'. Above the preview area are various text editing tools like bold, italic, and code block buttons. Below the preview area is a note: 'Attach files by dragging & dropping, selecting or pasting them.' To the right of the preview area is a 'Create pull request' button, which is also highlighted with a red border. There are other buttons for 'Allow edits by maintainers' and a note about GitHub Community Guidelines.

Image 9: *Creating a pull request from GitHub to merge our changes back into the original repository.*

The owner of the repository will get a notification about your proposal and can choose to add your changes or reject them (in this case, don't be too hopeful about your changes being accepted as the .replit file being missing is important to follow along the earlier steps of this tutorial ☹).

## Make it your own

For this tutorial, it's important that you do the steps yourself so that everything is correctly linked to your own GitHub account, but if you want an example to play with, use the one below. It's the same as the Flask app but it can greet individual users dynamically instead of having a static welcome message.

Visit <https://cwr-06-github--garethdwyer1.repl.co/Gareth<sup>42</sup>> to see it in action (replace the last part of the URL with your own name to receive a personalised greeting).

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-06-github?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

Can you figure out how it works?

<sup>42</sup><https://cwr-06-github--garethdwyer1.repl.co/Gareth>

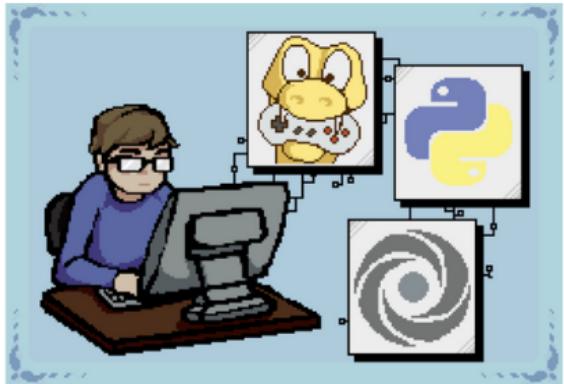
## Where next?

Cloning repositories and being able to immediately run them is useful in many scenarios, from just wanting to try out a cool project that you found to running production workflows.

Open-source software only exists because of people who build, maintain, and improve it. There's no global committee that decides who gets to be an open-source software developer and you can be one too. Find a project that you like, look at the "Issues" tab on GitHub to see what problems exist, and try to fix one. Many issues on GitHub are tagged with "Good first issue" to help direct newer developers to places where they can get started.

In the next tutorial, we'll do something a bit different and build a 2D game using PyGame.

# Building a game with PyGame and Repl.it



BUILDING  
A GAME  
..... WITH .....

**PYGAME AND REPL.IT**

So far, we've mainly seen how to write text-based programs, or those with a basic web front end. In this tutorial, we'll instead build a 2D game using PyGame. You'll use animated sprites and learn how to:

- Make these sprites move
- Recognise when a sprite is clicked with the mouse.

The basic premise of the game is as follows. You're a juggler, learning to juggle. Balls will fall down from the top of the screen, and you'll need to click them to 'throw' them up again. After several successful throws without dropping any balls, more balls will be added to make the game harder.

## Creating a PyGame repl

Although [PyGame<sup>43</sup>](#) is a standard Python library, Repl.it provides it installed as a separate language. Create a new repl and select PyGame from the language dropdown.

<sup>43</sup><https://en.wikipedia.org/wiki/Pygame>

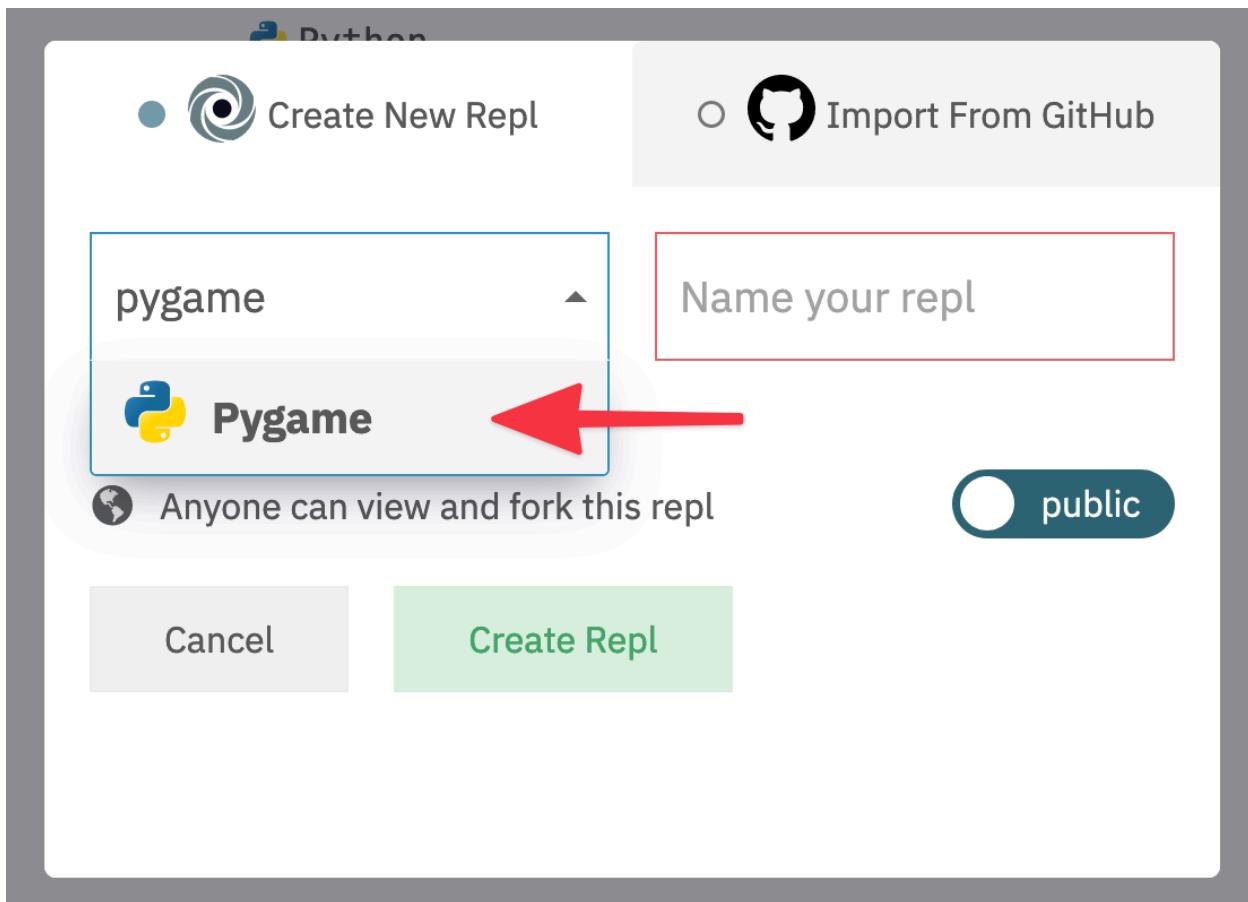


Image 1: Choosing PyGame from the Create New Repl screen.

You'll see "Python3 with PyGame" displayed in the default console and a separate pane in the Repl.it IDE where you will be able to see and interact with the game you will create.

The first thing we need is a so-called "sprite", which is a basic image file that we will use in our game. Download the tennis ball file available [here](#)<sup>44</sup> and save it to your local machine.

Now upload it to your repl using the `upload file` button and you should be able to see a preview of the image by clicking on it in the files pane.

<sup>44</sup>[https://raw.githubusercontent.com/ritza-co/public-images/master/small\\_tennis.png](https://raw.githubusercontent.com/ritza-co/public-images/master/small_tennis.png)

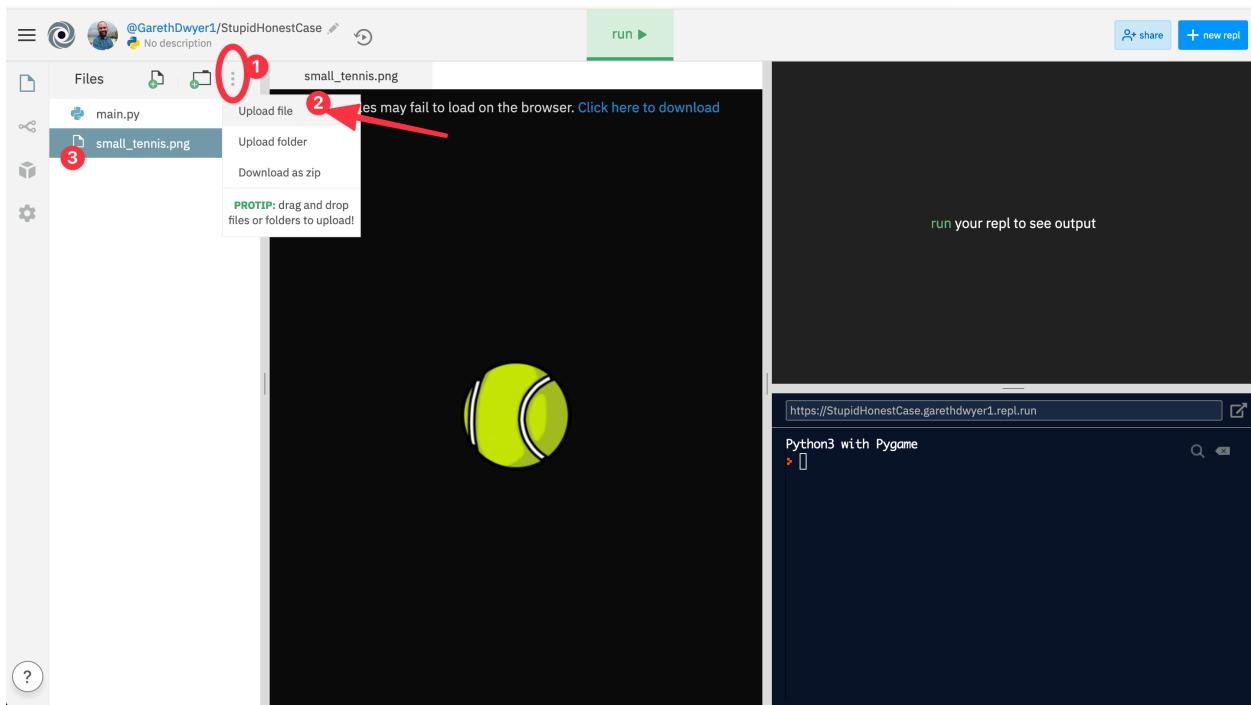


Image 2: Viewing our sprite after uploading it.

## Displaying the sprite using PyGame

Our first goal is to display the tennis ball in a game environment using PyGame. To do this, go back to the `main.py` file and add the following code.

```

1 import pygame
2
3 WIDTH = 800
4 HEIGHT = 600
5 BACKGROUND = (0, 0, 0)
6
7 class Ball:
8     def __init__(self):
9         self.image = pygame.image.load("small_tennis.png")
10        self.rect = self.image.get_rect()
11
12 def main():
13     pygame.init()
14     screen = pygame.display.set_mode((WIDTH, HEIGHT))
15     clock = pygame.time.Clock()
16     ball = Ball()

```

```
17
18     while True:
19         screen.fill(BACKGROUND)
20         screen.blit(ball.image, ball.rect)
21         pygame.display.flip()
22         clock.tick(60)
23
24 if __name__ == "__main__":
25     main()
```

This code looks a bit more complicated than it needs to be because in addition to drawing the ball to the screen, it also sets up a game loop. While basic 2D games appear to move objects around the screen, they usually actually simulate this effect by redrawing the entire screen many times per second. To account for this we need to run our logic in a `while True:` loop.

We start by importing PyGame and setting up some global variables: the size of our screen and the background color (black). We then define our `Ball`, setting up an object that knows where to find the image for the ball and how to get the default coordinates of where the image should be drawn.

We then set up PyGame by calling `init()` and starting the screen as well as a clock. The clock is necessary because each loop might take a different amount of time, based on how much logic needs to run to calculate the new screen. PyGame has built-in logic to calculate how much time elapses between calls to `clock.tick()` to draw frames faster or slower as necessary to keep the game experience smooth.

We start the game loop and call `blit` on our ball. [Blitting](#)<sup>45</sup> refers to moving all of the pixels from our sprite file (the tennis ball) to our game environment. The `flip()` function updates our screen and the `tick(60)` call means that our game will redraw the screen around 60 times per second.

If you run this code, you should see the ball pop up in the top right pane, as shown below.

---

<sup>45</sup>[https://en.wikipedia.org/wiki/Bit\\_blit](https://en.wikipedia.org/wiki/Bit_blit)

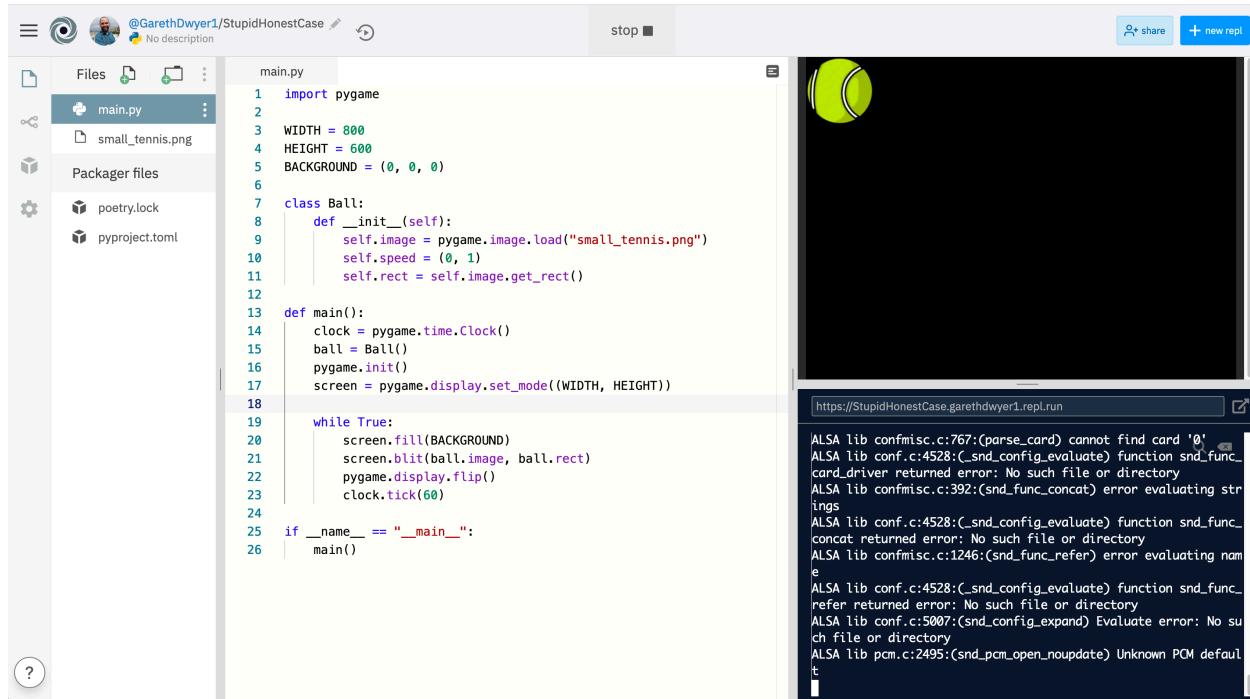


Image 3: Drawing the tennis ball in our PyGame environment.

## Making our tennis ball move with each frame

Although PyGame has a lot of built-in logic for handling common game operations, you still need to get your hands dirty with calculating some of the basic movements. For every loop, we need to tell our game the new X and Y coordinates to draw the ball. As we want our ball to move at a constant speed, we'll move the X and Y coordinates each loop.

Add two methods to your `Ball` class: `update` and `move`, and add a `speed` attribute. The new code for your `Ball` class should look as follows.

```

1 class Ball:
2     def __init__(self):
3         self.image = pygame.image.load("small_tennis.png")
4         self.speed = [0, 1]
5         self.rect = self.image.get_rect()
6
7     def update(self):
8         self.move()
9
10    def move(self):
11        self.rect = self.rect.move(self.speed)

```

Now modify your game loop to include a call to the new `update()` method. The loop should look as follows.

```

1  while True:
2      screen.fill(BACKGROUND)
3      screen.blit(ball.image, ball.rect)
4      ball.update()
5      pygame.display.flip()
6      clock.tick(60)

```

The `(0, 1)` tuple causes the ball to move its Y coordinate by 1 each loop and keep a constant X coordinate. This has the effect of making the ball drop slowly down the screen. Run your code again to check that this works.

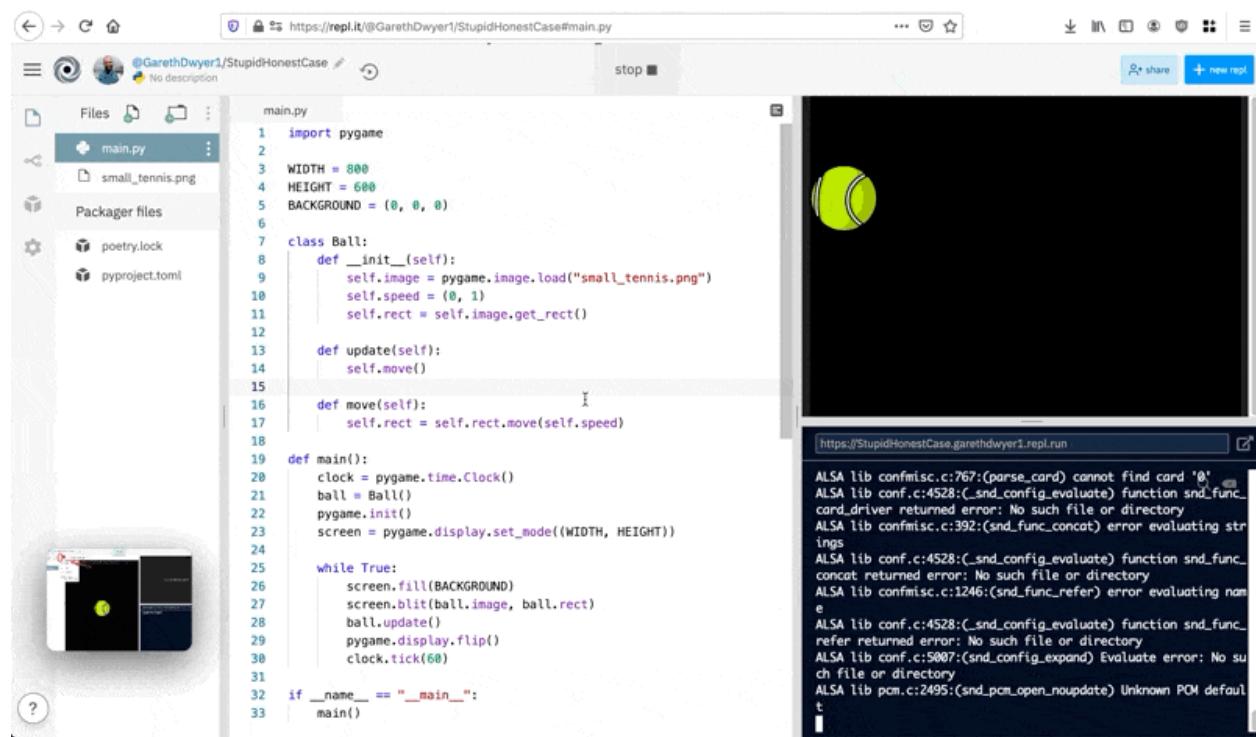


Image 4: The ball falling at a constant rate.

When the ball gets to the bottom of the screen, it'll just keep falling but that's OK for now. Let's see how we can add click detection.

## Processing events: Detecting mouse clicks

PyGame records all “events”, including mouse clicks, and makes these available through `pygame.event.get()`. We need to check what events happened in each game loop and see if any of them were important.

If the user clicks on an empty space, that will still be recorded but we will simply ignore it. If the user clicks on a falling ball, we want it to change direction.

Right after the `while True` loop, add the following lines of code.

```
1   for event in pygame.event.get():
2       if event.type == pygame.MOUSEBUTTONDOWN:
3           if ball.rect.collidepoint(pygame.mouse.get_pos()):
4               ball.speed = [0, -1]
```

With this code, we loop through all events and check for left click (`MOUSEBUTTONDOWN`) events. If we find one, we check if the click happened on top of the ball (using `collidepoint()` which checks for overlapping coordinates), and in this case we reverse the direction of the ball (still no x-axis or horizontal movement, but we make the ball move negatively on the y-axis, which is up.)

If you run this code again, you should now be able to click on the ball (let it fall for a while first) and see it change direction until it goes off the top of the screen.

## Making the ball bounce off the edges and move randomly

To simulate juggling, we want the ball to bounce off the “roof” (top edge of the screen) and “walls” (left and right edge). If the ball touches the “floor” (bottom edge) we want to kill it and remove it from the game as a dropped ball.

To achieve this, we’ll add logic to our `update()` method (this is why we kept it separate from our `move()` method before). Add two lines of code to `update()` to make it look as follows.

```
1   def update(self):
2       if self.rect.top < 0:
3           self.speed = [0, 1]
4       self.move()
```

This checks to see if the top of the ball is above the top of the screen. If it is, we set the speed back to `(0, 1)` (moving down).

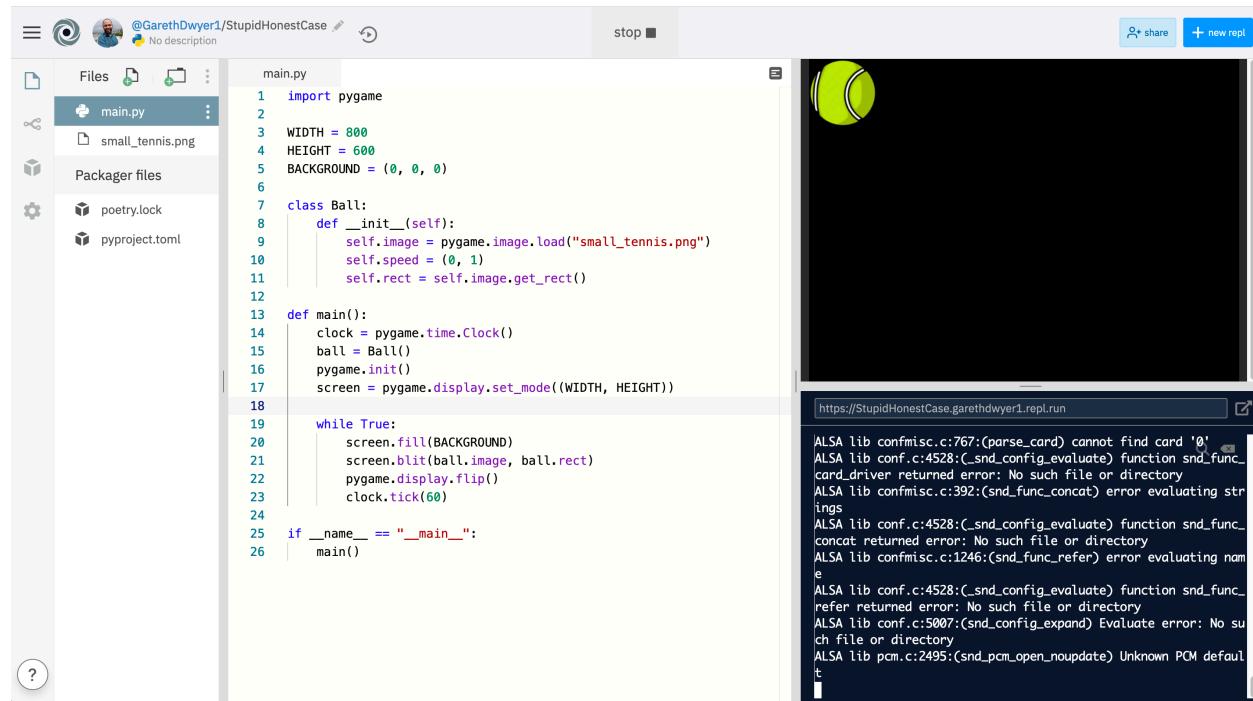


Image 5: Now we can bounce the ball off the ceiling.

So far, we have restricted the ball to moving vertically, but we can apply the same principles and move it horizontally or diagonally too. Let's also add some randomness into the mix so that it's less predictable (and harder for the player to press). The ball will randomly change its horizontal movement when it bounces off the ceiling and each time we throw it.

Import the `random` module at the top of your file and use the `random.randrange()` function to specify the range of acceptable horizontal movement. Also modify the `update()` function to detect if the ball is falling off the left or right edges and reverse its horizontal movement in this case.

Finally, modify the collision detection section to add randomness there too.

Your full code should now look as follows.

```

1 import pygame
2 import random
3
4 WIDTH = 800
5 HEIGHT = 600
6 BACKGROUND = (0, 0, 0)
7
8 class Ball:
9     def __init__(self):
10         self.image = pygame.image.load("small_tennis.png")
11         self.speed = [random.uniform(-4,4), 2]
12         self.rect = self.image.get_rect()

```

```

13
14     def update(self):
15         if self.rect.top < 0:
16             self.speed[1] = -self.speed[1]
17             self.speed[0] = random.uniform(-4, 4)
18         elif self.rect.left < 0 or self.rect.right > WIDTH:
19             self.speed[0] = -self.speed[0]
20         self.move()
21
22     def move(self):
23         self.rect = self.rect.move(self.speed)
24
25 def main():
26     clock = pygame.time.Clock()
27     ball = Ball()
28     pygame.init()
29     screen = pygame.display.set_mode((WIDTH, HEIGHT))
30
31     while True:
32         for event in pygame.event.get():
33             if event.type == pygame.MOUSEBUTTONDOWN:
34                 if ball.rect.collidepoint(pygame.mouse.get_pos()):
35                     ball.speed[0] = random.uniform(-4, 4)
36                     ball.speed[1] = -2
37                     screen.fill(BACKGROUND)
38                     screen.blit(ball.image, ball.rect)
39                     ball.update()
40                     pygame.display.flip()
41                     clock.tick(60)
42
43     if __name__ == "__main__":
44         main()

```

If you run your code again, you should be able to juggle the ball around by clicking on it and watch it randomly bounce off the ceiling and walls.

## Adding more balls

Juggling with one ball is no fun, so let's add some more. Because we used Object Oriented Programming (OOP), we can create more balls by instantiating more `Ball()` objects. We'll need to keep track of these so we'll add them to an array. For each iteration of the game loop, we'll need to update the position of each ball, so we'll need one more loop to account for this.

We also want to start keeping track of which of our balls is “alive” (that is, hasn’t hit the ground), so add an attribute for this to the `Ball` class too, in the `__init__` function.

```
1     self.alive = True
```

In the `main()` function, directly before the `while True:` line, add the following code.

```
1     ball1 = Ball()
2     ball2 = Ball()
3     ball3 = Ball()
4
5     balls = [ball1, ball2, ball3]
```

Now remove the `ball=Ball()`, `ball.update()` and `screen.blit(...)` lines and replace them with a loop that updates all of the balls and removes the dead ones (even though we haven’t written the logic yet to stop the balls from ever being alive.)

```
1     for i, ball in enumerate(balls):
2         if ball.alive:
3             screen.blit(ball.image, ball.rect)
4             ball.update()
5         if not ball.alive:
6             del balls[i]
```

You’ll also need to account for multiple balls in the the event detection loop. For each event, loop through all of the balls and check if the mouse click collided with any of them.

At this point, the full `main()` function should look as follows.

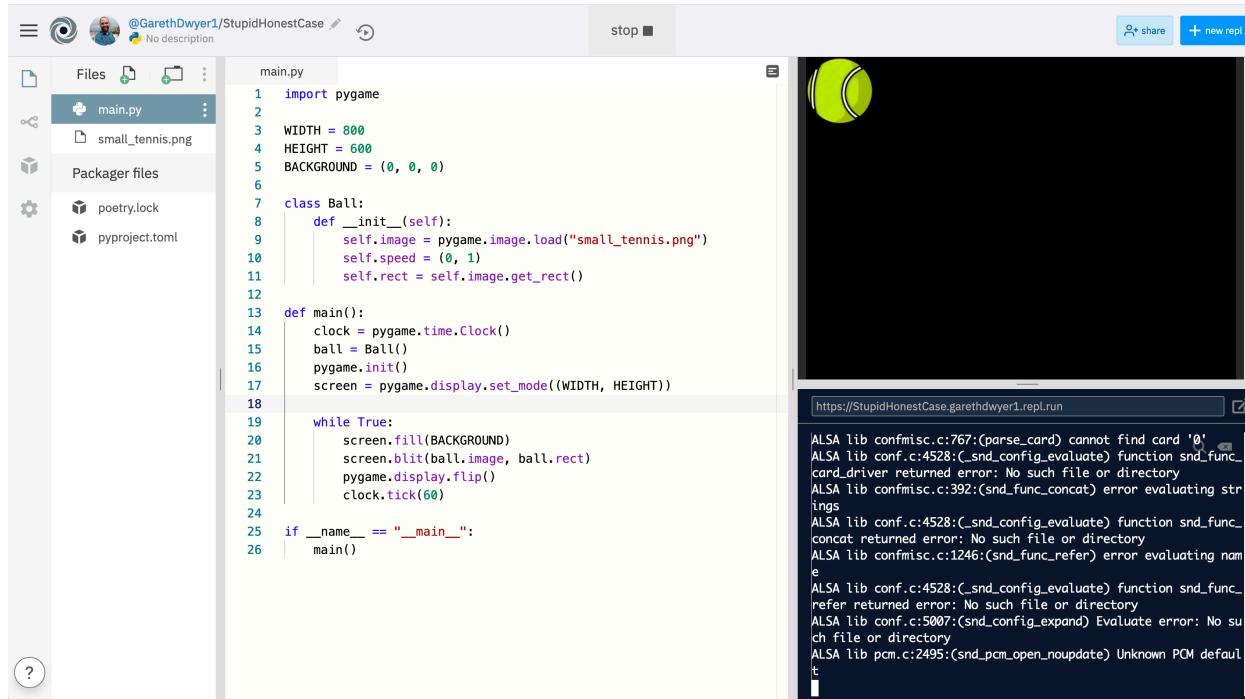
```
1 def main():
2     clock = pygame.time.Clock()
3     pygame.init()
4     screen = pygame.display.set_mode((WIDTH, HEIGHT))
5
6     ball1 = Ball()
7     ball2 = Ball()
8     ball3 = Ball()
9
10    balls = [ball1, ball2, ball3]
11
12    while True:
13        for event in pygame.event.get():
```

```
14     if event.type == pygame.MOUSEBUTTONDOWN:
15         for ball in balls:
16             if ball.rect.collidepoint(pygame.mouse.get_pos()):
17                 ball.speed[0] = random.randrange(-4, 4)
18                 ball.speed[1] = -2
19                 break
20         screen.fill(BACKGROUND)
21         for i, ball in enumerate(balls):
22             if ball.alive:
23                 screen.blit(ball.image, ball.rect)
24                 ball.update()
25             if not ball.alive:
26                 del balls[i]
27         pygame.display.flip()
28         clock.tick(60)
```

To kill balls when they fall through the floor, we can add another check to the `update()` function as follows.

```
1     elif self.rect.bottom > HEIGHT:
2         self.alive = False
```

Run the code again and you should be able to juggle three balls. See how long you can keep them in the air.



The screenshot shows a Repl.it workspace for a Python game. The code in main.py initializes a single ball and handles its movement. The terminal output shows ALSA errors related to sound configuration.

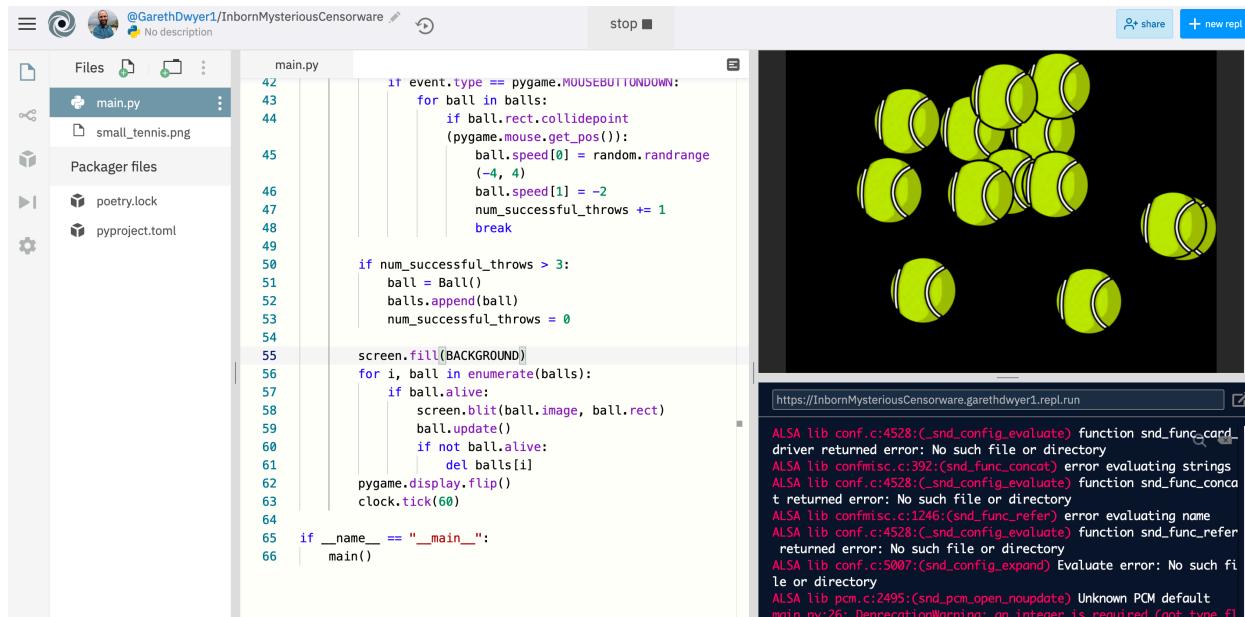
```

main.py
1 import pygame
2
3 WIDTH = 800
4 HEIGHT = 600
5 BACKGROUND = (0, 0, 0)
6
7 class Ball:
8     def __init__(self):
9         self.image = pygame.image.load("small_tennis.png")
10        self.speed = (0, 1)
11        self.rect = self.image.get_rect()
12
13 def main():
14     clock = pygame.time.Clock()
15     ball = Ball()
16     pygame.init()
17     screen = pygame.display.set_mode((WIDTH, HEIGHT))
18
19     while True:
20         screen.fill(BACKGROUND)
21         screen.blit(ball.image, ball.rect)
22         pygame.display.flip()
23         clock.tick(60)
24
25 if __name__ == "__main__":
26     main()

```

Image 6: Juggling three balls.

If you want a harder version, add a counter to keep track of how many successful throws the player has achieved and add a new ball for every three successful throws.



The screenshot shows a Repl.it workspace for a modified game where multiple balls are tracked and updated. The code adds logic for handling mouse input and creating new balls. The terminal output shows ALSA errors.

```

main.py
42     if event.type == pygame.MOUSEBUTTONDOWN:
43         for ball in balls:
44             if ball.rect.collidepoint(
45                 pygame.mouse.get_pos()):
46                 ball.speed[0] = random.randrange(
47                     (-4, 4)
48                     ball.speed[1] = -2
49                     num_successful_throws += 1
50                     break
51
52                     if num_successful_throws > 3:
53                         ball = Ball()
54                         balls.append(ball)
55                         num_successful_throws = 0
56
57                     screen.fill(BACKGROUND)
58                     for i, ball in enumerate(balls):
59                         if ball.alive:
60                             screen.blit(ball.image, ball.rect)
61                             ball.update()
62                             if not ball.alive:
63                                 del balls[i]
64
65                     pygame.display.flip()
66                     clock.tick(60)
67
68 if __name__ == "__main__":
69     main()

```

Image 7: Adding more balls.

Now the game is to see how many balls you can juggle with. If it's too easy, modify the speeds and angles of the balls.

## Make it your own

If you followed along, you'll already have your own version of the repl to extend. If not, start from ours. Fork it from the embed below.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-07-juggling-with-pygame?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

You've learned how to make 2D games using PyGame. If you want to make more games but are stuck for ideas, check out [PyGame's extensive collection of examples<sup>46</sup>](#).

You could also extend the juggling game more. For example, make the balls accelerate as they fall, or increase the speed of all balls over time.

---

<sup>46</sup><https://www.pygame.org/docs/ref/examples.html>

# Staying safe: Keeping your passwords and other secrets secure

While developing software fully in public has many benefits, it also means that we need to be extra careful about leaking sensitive information. Because all of our repls are public by default, we shouldn't store passwords, access keys, personal information, or anything else sensitive in them.

Even if you're coding offline or only in private repls, it's good practice to keep your code separate from any private information in any case.

In this tutorial, we'll look at how to use the special `.env` file that Repl.it provides to set [environment variables](#)<sup>47</sup>. We can use these to store sensitive information and Repl.it will make sure that this file isn't included when others fork our repl.

## Understanding `.env` files

Similarly to the `.replit` file that we saw in a previous tutorial, the `.env` file is a special Repl.it file. If you call a file exactly `.env`, Repl.it will

- Not include this file in any forks of the repl
- Attempt to parse key-value pairs out of this file and make them available to the underlying operating system.

This can be used to store all kinds of configuration, but it's commonly used for passwords, API keys, and database credentials.

## The structure of a `.env` file

A `.env` file consists of keys-value pairs, one per line, separated by an `=` sign. Environment variables are traditionally in ALL\_CAPS and separated by underscores. For example, you might have a `.env` file with the following.

```
1 SECRET_PASSWORD=ThisIsMySuperSecretP@ssword!!
```

With this file present, your scripts can load the variable `SECRET_PASSWORD` from the operating system environment directly.

Unlike in Python, where `x = 1` and `x= 1` are the same, in `.env` files, spaces matter and you should be careful to not add any extra ones.

---

<sup>47</sup>[https://en.wikipedia.org/wiki/Environment\\_variable](https://en.wikipedia.org/wiki/Environment_variable)

## Refactoring our weather project to keep our API key secure

In the [working with files tutorial<sup>48</sup>](#) we used the WeatherStack API to fetch weather data and save it to disk. Part of this involved getting an API key from WeatherStack. Because WeatherStack limits the number of calls each key can make per day, it would be bad if someone else used up the quota for our key and broke our app as a result.

Let's refactor the WeatherStack project to prevent our key from being made public.

Visit [The image shows a screenshot of a repl.it interface. At the top, there is a header with a user profile, the repository name '@GarethDwyer1/cwr-02-weather-report', and a 'Run' button. Below the header, there is a file tree on the left and a code editor on the right. In the code editor, there is a line of Python code that includes an API key. A red arrow labeled '1' points to the pencil icon in the top right corner of the header. Another red arrow labeled '2' points to the 'fork' button in the bottom left corner of the interface. The code in the editor looks like this:

```

16
17     temperature = js\['current'\]\['temperature'\]
18     date = js\['location'\]\['localtime'\]

```](https://repl.it/@GarethDwyer1/cwr-02-weather-report<sup>49</sup></a> (or your own version of this if you followed along previously) and create a new fork by pressing the pencil icon and then fork.</p>
</div>
<div data-bbox=)

Image 1: Forking our repl before refactoring it.

We have `API_KEY` defined near the top of `main.py`, and this is the value that we want to keep secret. Let's move it to a `.env` file instead.

Click on the add file icon and call your new file `.env`. Be sure to add the initial full stop and don't add any spaces.

<sup>48</sup><http://www.codewithrepl.it/02-managing-files-using-repl-it.html>

<sup>49</sup><https://repl.it/@GarethDwyer1/cwr-02-weather-report>

The screenshot shows a Repl.it workspace. On the left, there's a sidebar with icons for files, imports, variables, and settings. The main area shows a file tree with several text files: cities.txt, London.txt, Paris.txt, San Francisco.txt, Sydney.txt, and a newly created .env file. A red arrow points from a circled '1' to the 'Add file' button. Another red arrow points from a circled '2' to the .env file in the file tree. The code editor on the right contains a Python script named main.py:

```
1 import requests
2
3 API_KEY = "baaf201731c0cbc4af2c519cb578f907"
4 WS_URL = "http://api.weatherstack.com/current"
5
6 cities = []
7 with open("cities.txt") as f:
8     for line in f:
9         cities.append(line.strip())
10 print(cities)
11
12 for city in cities:
13     parameters = {'access_key': API_KEY, 'query': city}
```

Image 2: Creating the .env file to store sensitive information.\*

Now remove the API\_KEY variable from the main.py file and add it to the .env file, removing all quotation marks ("") and spaces.

Your .env file should look as follows (but use your own WeatherStack API instead of the example given here).

1 API\_KEY=baaf201731c0cbc4af2c519cb578f907

## Testing that the file is not copied into others' forks

If you copy your project's URL into an incognito window (or use a separate browser), you'll see all of the other files as usual, but the .env file will not be there. Your API key, and the entire .env file, can only be seen when you're logged into the Repl.it account that created it.

```

1 import requests
2
3 WS_URL = "http://api.weatherstack.com/current"
4
5 cities = []
6 with open("cities.txt") as f:
7     for line in f:
8         cities.append(line.strip())
9 print(cities)
10
11 for city in cities:
12     parameters = {'access_key': API_KEY, 'query': city}
13     response = requests.get(WS_URL, parameters)
14     js = response.json()
15
16     temperature = js['current']['temperature']
17     date = js['location']['localtime']
18
19     with open(f"{city}.txt", "a") as f:
20         f.write(f"{date},{temperature}\n")
21

```

Image 3: Checking that the .env file isn't included in public versions of our repl.

## Using environment variables in our script

Our API key is now securely defined and available to the project, but we still need to tell our code where to find it. Because this data is stored in environment variables, we need to use the operating system (os) module to access it.

At the top of your `main.py` file, add an import for `os` and load the `API_KEY` into a variable as follows.

```

1 import requests
2 import os
3
4 API_KEY = os.getenv("API_KEY")

```

The `getenv` function looks for an environment variable of a specific name. Now our code (and anyone who sees it) only needs to know the name of the key that stores our private API key, instead of the API key itself. You should be able to run your code again at this point to verify that new weather entries are correctly added to the relevant files.

There are many other environment variables that make various parts of an operating system work correctly. For example, you could also take a look at the `LANG` and `PATH` environment variables, which will show you that Repl.it has their servers configured to use US English and 8-bit unicode character encoding, and have some default places where the system looks for executable programs.



A screenshot of the Repl.it interface. On the left, a code editor shows a Python script named 'main.py' with the following code:

```

1 import requests
2 import os
3
4 print(os.getenv("LANG"))
5 print(os.getenv("PATH"))
6
7

```

Two red arrows point from the lines `print(os.getenv("LANG"))` and `print(os.getenv("PATH"))` to the right-hand terminal window. The terminal window displays the output of these commands:

```

en_US.UTF-8
/home/runner/.apt/usr/bin:/opt/virtualenvs/python3/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin
> []

```

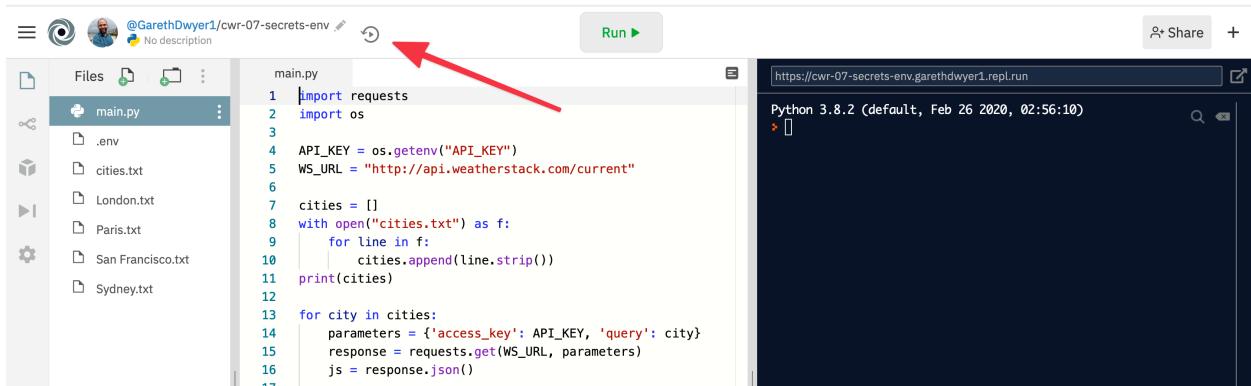
Image 4: Looking at other environment variables.

## Time travelling to find secrets

We removed the sensitive information from our project, but it's not actually completely gone. It's securely placed in our `.env` file, but it's also still saved in the repl's history.

Repl.it saves every change you make to a project so that you can always go back to previous versions if you make a mistake or need to check what has changed.

Click on the history button in the top bar, as shown below.



A screenshot of the Repl.it interface showing the history viewer. A red arrow points to the 'History' button in the top navigation bar. The left sidebar shows the project structure with files like `main.py`, `.env`, and several `.txt` files. The main area shows the code for `main.py`:

```

1 import requests
2 import os
3
4 API_KEY = os.getenv("API_KEY")
5 WS_URL = "http://api.weatherstack.com/current"
6
7 cities = []
8 with open("cities.txt") as f:
9     for line in f:
10         cities.append(line.strip())
11 print(cities)
12
13 for city in cities:
14     parameters = {'access_key': API_KEY, 'query': city}
15     response = requests.get(WS_URL, parameters)
16     js = response.json()
17

```

The right-hand terminal window shows the Python version and a blank command prompt.

Image 5: Diving into the history of our repl.

You should see a bunch of entries from each change you've made to this project. Click through them and find the one where you deleted your API key. As you can see, the history viewer shows not only which lines have been changed, but also what was there before.

The screenshot shows a GitHub repository interface. On the left is a sidebar with navigation links: Home, My Repls, Talk, Notifications, Languages, Templates, Tutorials, and Teams (BETA). The main area displays a list of commits for a file named 'repl.py'. A red arrow points from the bottom of the commit list to the commit message of the bottom-most commit, which is timestamped 'Jul 29, 2020 10:06 PM'. This commit message contains Python code. A red box highlights the line where an API key is defined:

```

@@ -1,7 +1,8 @@
1 import requests
2 + import os
3 - API_KEY = "baaf201731c0cbc4af2c519cb578f907"
4 + API_KEY = os.getenv("API_KEY")
5
6     WS_URL = "http://api.weatherstack.com/current"
7
8     cities = []
9
10    with open("cities.txt") as f:

```

*Image 6: Finding the credentials in the change logs.*

Luckily history is not included when other people fork your repl so this is not a huge problem, but it's important to keep in mind where people might find your credentials.

In our case, the worst case scenario is that someone finds our WeatherStack API key and uses up the quota, which is not the end of the world. A far more painful (and very common) scenario involves real money. For example, a developer signs up for a free trial on an expensive service like AWS, links a credit card, and then accidentally pushes the credentials for the service to GitHub or similar. Even if they realise their mistake and delete these within seconds, the credentials themselves are still available in the history of their repository. Hackers have bots that regularly look out for mistakes like this and use the credentials to spin up thousands of servers (often to mine cryptocurrency or join a botnet attack), potentially costing the poor developer thousands of dollars before they notice.

## Rotating credentials

Even if there's a small chance that your API key has been exposed, it's important to rotate it. This involves creating a new key, ensuring the new key works with your service, and then disabling the old one.

In the case of WeatherStack, there is no option to create a new key while keeping the old one active, so we need to reset it and then copy the new key to our `.env` file (meaning that our app can't function between the time that we disable the old key and replace it with the new one).

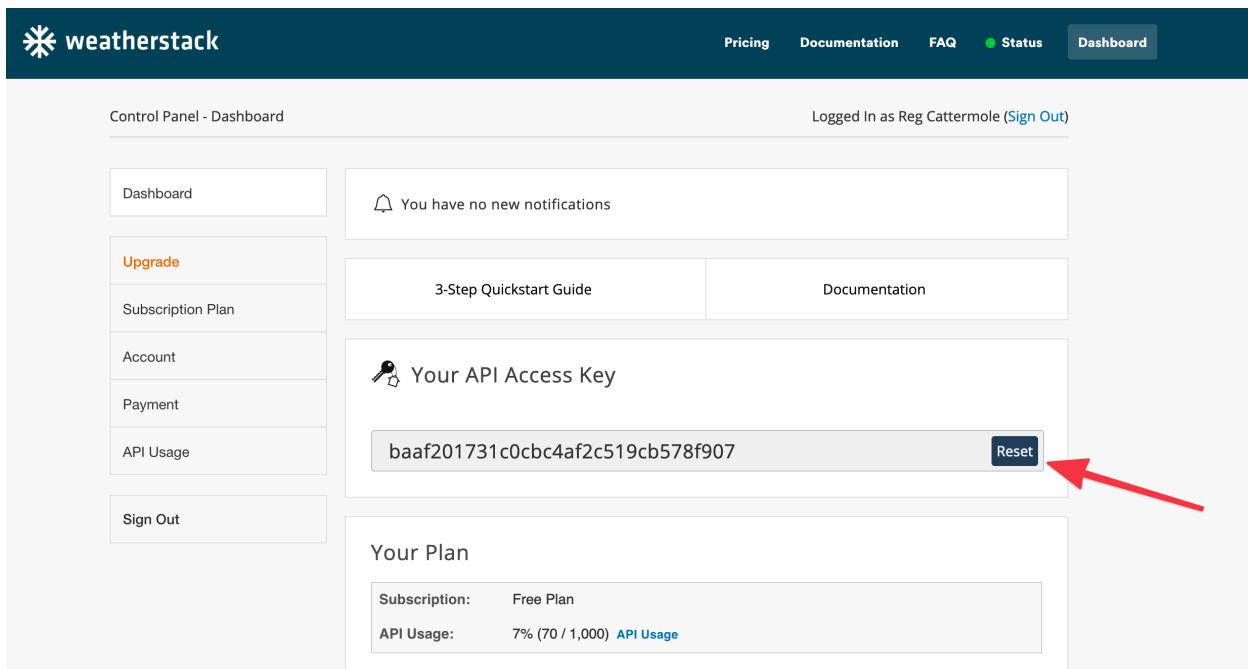


Image 7: Rotating our WeatherStack API key.

Visit your WeatherStack account and press the reset button to get your new API key.

## Make it your own

You can make a copy of the new repl below. If you fork it, the `.env` file will be missing, so you'll need to create it and add your WeatherStack API key before it will run.

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-08-secrets-env?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

There's a lot more that you can do with `.env` files. In a later tutorial, we'll use it to store database credentials (which are more important to keep safe than a free API key).

You could also keep other private information in environment variables. For example, if you code a hangman game, you could keep the word that people need to guess in there so that you can share your code without spoiling the game.

# An introduction to pytest and doing test-driven development with Repl.it

In this tutorial we'll introduce test-driven development and you'll see how to use [pytest<sup>50</sup>](#) to ensure that your code is working as expected.

pytest lets you specify inputs and expected outputs for your functions. It runs each input through your function and validates that the output is correct. pytest is a Python library and works just like any other Python library: you install it through your package manager and you can import it into your Python code. Tests are written in Python too, so you'll have code testing other code.

Test-driven development or TDD is the practice of writing tests before you write code. You can read more about TDD and why it's popular on [Wikipedia<sup>51</sup>](#).

Specifically you'll:

- See how to structure your project to keep your tests separate but still have them refer to your main code files
- Figure out the requirements for a function that can split a full name into first and last name components
- Write tests for this function
- Write the actual function.

## Creating a project structure for pytest

For large projects, it's useful to keep your testing code separate from your application code. In order for this to work, you'll need your files set up in specific places, and you'll need to create individual Python **modules** so that you can refer to different parts of the project easily.

Create a new Python repl called `namesplitter`. As always, it'll already have a `main.py` file, but we're going to put our name splitting function into a different module called `utils`, which can house any helper code that our main application relies on. We also want a dedicated place for our tests.

Create two new folders: one called `utils` and one called `tests`, using the `add folder` button. Note that when you press this button it will by default create a folder in your currently active folder, so select the `main.py` file after creating the first folder or the second folder will be created inside the first folder.

---

<sup>50</sup><https://docs.pytest.org/en/stable/>

<sup>51</sup>[https://en.wikipedia.org/wiki/Test-driven\\_development](https://en.wikipedia.org/wiki/Test-driven_development)

You want both the folders to be at the root level of your project.

Now add a file at the root level of the project called `__init__.py`. This is a special file that indicates to Python that we want our project to be treated as a “module”: something that other files can refer to by name and import pieces from. Also add an `__init__.py` file inside the `utils` folder and the `tests` folder. These files will remain empty, but it’s important that they exist for our tests to run. Their presence specifies that our main project should be treated as a module and that any code in our `utils` and `tests` folders should be treated as submodules of the main one.

Finally, create the files where we’ll actually write code. Inside the `utils` folder create a file called `name_helper.py` and inside the `tests` folder create one called `test_name_helper.py`. Your project should now look as follows. Make sure that you have all the files and folders with exactly these names, in the correct places.

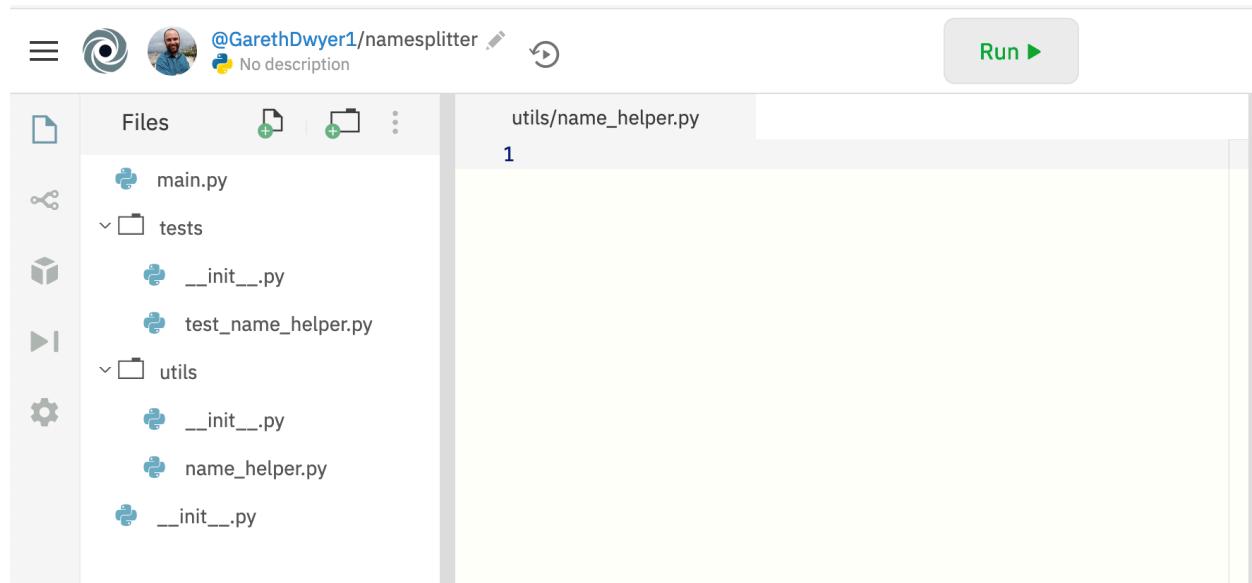


Image 1: Setting up our project structure for pytest.

## Defining examples for the name split function

Splitting names is useful in many contexts. For example, it is a common requirement when users sign up on websites with their full names and then companies want to send personalised emails addressing users by their first name only. You might think that this is as simple as splitting a name based on spaces as in the following example.

```

1 def split_name(name):
2     first_name, last_name = name.split()
3     return first_name, last_name
4
5 print(split_name("John Smith"))
6 # >>> ["John", "Smith"]

```

While this does indeed work in many cases, names are surprisingly complicated and it's very common to make mistakes when dealing with them as programmers, as discussed in [this classic article<sup>52</sup>](#). It would be a huge project to try and deal with *any* name, but let's imagine that you have requirements to deal with the following kinds of names:

- First Last, e.g. John Smith
- First Middle Last, e.g. John Patrick Smith (John Patrick taken as first name)
- First Middle Middle Last, e.g. John Patrick Thomson Smith (John Patrick Thomson taken as first name)
- First last last Last, e.g. Johan van der Berg (note the lowercase letters, Johan taken as first name, the rest as last)
- First Middle last last Last, e.g. Johan Patrick van der Berg (note the lowercase letters, Johan taken as first name, the rest as last)
- Last, e.g. Smith (we can assume that if we are given only one name, it is the last name)

Specifically, you can assume that once you find a name starting with a lowercase letter, it signifies the start of a last name, and that all other names starting with a capital letter are part of the first and middle names. Middle names can be combined with first names.

Of course, this does not cover all possibilities, but it is a good starting point in terms of requirements.

Using TDD, we always write *failing tests* first. The idea is that we should write a test about how some code *should* behave, check to make sure that it breaks in the way we expect (as the code isn't there). Only then do we write the actual code and check that the tests now pass.

## Writing the test cases for our names function

Now that we understand what our function should do, we can write tests to check that it does. In the `tests/test_name_helper.py` file, add the following code.

---

<sup>52</sup><https://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/>

```
1 from namesplitter.utils import name_helper
2
3 def test_two_names():
4     assert split_name("John Smith") == ["John", "Smith"]
```

Note that the `namesplitter` in the first line is taken from the name of your Repl.it project, which defines the names of the parent module. If you called your project something else, you'll need to use that name in the import line. It's important to not include special characters in the project name (including a hyphen, so names like `my-tdd-demo` are out) or the import won't work.

The `assert` keyword simply checks that a specific statement evaluates to `True`. In this case, we call our function on the left-hand side and give the expected value on the right-hand side, and ask `assert` to check if they are the same.

This is our most basic case: we have two names and we simply split them on the single space. Of course, we haven't written the `split_name` function anywhere yet, so we expect this test to fail. Let's check.

Usually you would run your tests by typing `py.test` into your terminal, but using Repl.it things work better if we import `pytest` into our code base and run it from there. This is because a) our terminal is always already activated into a Python environment and b) caching gets updated when we press the Run button, so invoking our tests from outside of this means that they could run on old versions of our code, causing confusion.

Let's run them from our `main.py` file for now as we aren't using it for anything else yet. Add the following to this file.

```
1 import pytest
2 pytest.main()
```

Press the Run button. `pytest` does automatic test discovery so you don't need to tell it which tests to run. It will look for files that start with `test` and for functions that start with `test_` and assume these are tests. (You can read more about exactly how test discovery works and can be configured [here<sup>53</sup>](#).)

You should see some scary looking red failures, as shown below. (`pytest` uses dividers such as ===== and ----- to format sections and these can get messy if your output pane is too narrow. If things look a bit wonky try making it wider and rerunning.)

---

<sup>53</sup>[https://docs.pytest.org/en/reorganize-docs/new-docs/user/naming\\_conventions.html](https://docs.pytest.org/en/reorganize-docs/new-docs/user/naming_conventions.html)

```
import pytest
pytest.main()
```

Image 2: *Reading the pytest error messages.*

If you read the output from the top down you'll see a bunch of different things happened. First, pytest ran test discovery and found one test. It ran this and it failed so you see the first red F above the FAILURES section. That tells us exactly which line of the test failed and how. In this case, it was an `AttributeError` as we tried to use `split_name` which was not defined. Let's go fix that.

Head over to the `utils/name_helper.py` file and add the following code.

```
def split_name(name):
    first_name, last_name = name.split()
    return first_name, last_name
```

This is the very simple version we discussed earlier that can only handle two names, but it will solve the name error and TDD is all about small increments. Press Run to re-run the tests and you should see a far more friendly green output now, as below, indicating that all of our tests passed.

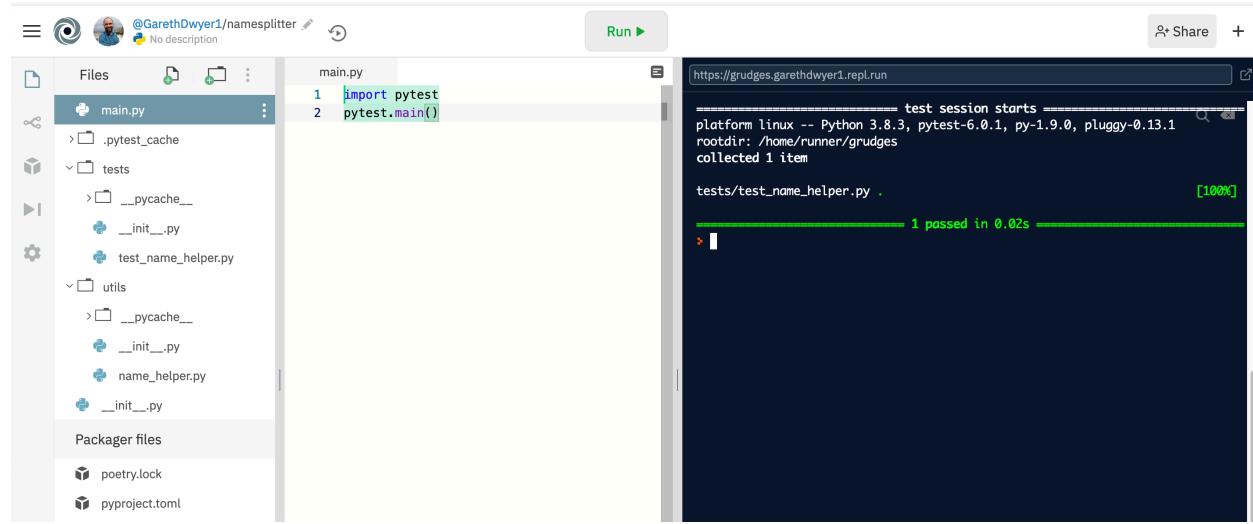


Image 3: Seeing our tests pass after updating the code.

Before fixing our function to handle more complex cases, let's first write the tests and check that they fail. Go back to `tests/test_name_helper.py` and add the following four test functions beneath the existing one.

```

1  from namesplitter.utils import name_helper
2
3  def test_two_names():
4      assert name_helper.split_name("John Smith") == ["John", "Smith"]
5
6  def test_middle_names():
7      assert name_helper.split_name("John Patrick Smith") == ["John Patrick", "Smith"]
8      assert name_helper.split_name("John Patrick Thomson Smith") == ["John Patrick Th\\
9 omson", "Smith"]
10
11 def test_surname_prefixes():
12     assert name_helper.split_name("John van der Berg") == ["John", "van der Berg"]
13     assert name_helper.split_name("John Patrick van der Berg") == ["John Patrick", "\\
14 van der Berg"]
15
16 def test_split_name_onename():
17     assert name_helper.split_name("Smith") == ["", "Smith"]
18
19 def test_split_name_nonames():
20     assert name_helper.split_name("") == ["", ""]

```

Rerun the tests and you should see a lot more output now. If you scroll back up to the most recent ===== test session starts ===== section, it should look as follows.

```
1 import pytest
2 pytest.main()
```

```
platform linux -- Python 3.8.3, pytest-6.0.1, py-1.9.0, pluggy-0.13.1
rootdir: /home/runner/grudges
collected 5 items

tests/test_name_helper.py .FFFFF [100%]

===== FAILURES =====
test_middle_names
def test_middle_names():
>     assert name_helper.split_name("John Patrick Smith") == ["John Patrick", "Smith"]
E     ValueError: too many values to unpack (expected 2)

tests/test_name_helper.py:7:

name = 'John Patrick Smith'

def split_name(name):
>     first_name, last_name = name.split()
E     ValueError: too many values to unpack (expected 2)

utils/name_helper.py:2: ValueError
                               test_surname_prefixes

def test_surname_prefixes():
>     assert name_helper.split_name("John van der Berg") == ("John", "van der Berg")
E     ValueError: too many values to unpack (expected 2)

tests/test_name_helper.py:11:

name = 'John van der Berg'

def split_name(name):
```

Image 4: Seeing more failures after adding more tests.

In the top section, the .FFFF is shorthand for “five tests were run, the first one passed and the next four failed” (a green dot indicates a pass and a red F indicates a failure). If you had more files with tests in them, you would see a line like this per file, with one character of output per test.

The failures are described in detail after this, but they all amount to variations of the same problem. Our code currently assumes that we will always get exactly two names, so it either has too many or too few values after running `split()` on the test examples.

## Fixing our `split_name` function

Go back to `name_helper.py` and modify it to look as follows.

```
1 def split_name(name):
2     names = name.split(" ")
3
4     if not name:
5         return ["", ""]
6
7     if len(names) == 1:
8         return ["", name]
9
10    if len(names) == 2:
```

```

11     firstname, lastname = name.split(" ")
12     return [firstname, lastname]

```

This should handle the case of zero, one, or two names. Let's run our tests again to see if we've made progress before we handle the more difficult cases. You should get a lot less output now and three green dots, as shown below.

```

https://grudges.garethdwyer1.repl.run
platform linux -- Python 3.8.3, pytest-6.0.1, py-1.9.0, pluggy-0.13.1
rootdir: /home/runner/grudges
collected 5 items

tests/test_name_helper.py: FF...
[100%]

===== FAILURES =====
test_middle_names
> assert name_helper.split_name("John Patrick Smith") == ["John Patrick", "Smith"]
E AssertionError: assert None == ['John Patrick', 'Smith']
E + where None = <function split_name at 0x7f1489a34ca0>('John Patrick Smith')
E + where <function split_name at 0x7f1489a34ca0> = name_helper.split_name

tests/test_name_helper.py:7: AssertionError
test_surname_prefixes

def test_surname_prefixes():
> assert name_helper.split_name("John van der Berg") == ["John", "van der Berg"]
E AssertionError: assert None == ['John', 'van der Berg']
E + where None = <function split_name at 0x7f1489a34ca0>('John van der Berg')
E + where <function split_name at 0x7f1489a34ca0> = name_helper.split_name

tests/test_name_helper.py:11: AssertionError
===== short test summary info =====
FAILED tests/test_name_helper.py::test_middle_names - AssertionError: assert None == [...]
FAILED tests/test_name_helper.py::test_surname_prefixes - AssertionError: assert None ...
[...]
> [2 failed, 3 passed in 0.27s]

```

Image 5: Progress: some of our tests pass now.

The rest of the output indicates that it's the middle names and surname prefix examples that are still tripping up our function, so let's add the code we need to fix those. Another important aspect of TDD is keeping your functions as small as possible so that they are easier to understand, test, and reuse, so let's write a second function to handle the three or more names cases.

Add the new function called `split_name_three_plus()` and add an `else` clause to the existing `split_name` function where you call this new function. The entire file should now look as follows.

```

1 def split_name_three_plus(names):
2     first_names = []
3     last_names = []
4
5     for i, name in enumerate(names):
6         if i == len(names) - 1:
7             last_names.append(name)
8         elif name[0].islower():
9             last_names.extend(names[i:])
10            break
11        else:

```

```

12         first_names.append(name)
13     first_name = " ".join(first_names)
14     last_name = " ".join(last_names)
15     return [first_name, last_name]
16
17 def split_name(name):
18     names = name.split(" ")
19
20     if not name:
21         return ["", ""]
22
23     if len(names) == 1:
24         return ["", name]
25
26     if len(names) == 2:
27         firstname, lastname = name.split(" ")
28         return [firstname, lastname]
29     else:
30         return split_name_three_plus(names)

```

The new function works by always appending names to the `first_names` list until it gets to the last name, or until it encounters a name that starts with a lowercase letter, at which point it adds all of the remaining names to `last_names` list. If you run the tests again, they should all pass now.

```

https://grudges.garethdwyer1.repl.run
platform linux -- Python 3.8.3, pytest-6.0.1, py-1.9.0, pluggy-0.13.1
rootdir: /home/runner/grudges
collected 5 items

tests/test_name_helper.py ..... [100%]

5 passed in 0.04s

```

Image 6: All of the tests pass after adding a new function.

The tests were already helpful in making sure that we understood the problem and that our function

worked for specific examples. If we had made any off-by-one mistakes in our code that deals with three or more names, our tests would have caught them. If we need to refactor or change our code in future, we can also use our tests to make sure that our new code doesn't introduce any regressions (where fixing problems causes code to break on other examples that worked before the fix.)

## Using our function

Let's build a very basic application to use our function. Replace the testing code in `main.py` with the following.

```

1 from utils import name_helper
2
3 name = input("Please enter your full name: ")
4
5 first_name, last_name = name_helper.split_name(name)
6
7 print(f"Your first name is: {first_name}")
8 print(f"Your last name is: {last_name}")

```

If you run this, it will prompt the user for their name and then display their first and last name.

```

from utils import name_helper
name = input("Enter your name: ")
first_name, last_name = name_helper.split_name(name)
print(f"Your first name is: {first_name}")
print(f"Your last name is: {last_name}")

```

Image 7: Using our function in a basic console application.

## Make it your own

We've written a name splitter that can handle some names more complicated than just "John Smith". It's not perfect though: for example, if you put in a name with two consecutive spaces it will crash

our program. You could fork the project and fix this by first writing a test with consecutive spaces and then modifying the code to handle this (and any other edge cases you can think of).

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/namesplitter?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

Because you're using the `main.py` file now, you can also invoke pytest directly from the output console on the right by typing `import pytest; pytest.main()`. Note that updates to your code are only properly applied when you press the Run button though, so make sure to run your code between changes before running the tests.

```
from utils import name_helper
name = input("Enter your name: ")
first_name, last_name = name_helper.split_name(name)
print(f"Your first name is: {first_name}")
print(f"Your last name is: {last_name}")

def test_name():
    assert name_helper.split_name("Gareth Dwyer") == ("Gareth", "Dwyer")
    assert name_helper.split_name("GarethDwyer") == ("GarethDwyer", "")
    assert name_helper.split_name("Gareth Dwyer") == ("Gareth", "Dwyer")
    assert name_helper.split_name("Gareth Dwyer") == ("Gareth", "Dwyer")
    assert name_helper.split_name("Gareth Dwyer") == ("Gareth", "Dwyer")
```

```
https://grudges.garethdwyer1.repl.run
Enter your name: Gareth Dwyer
Traceback (most recent call last):
  File "main.py", line 5, in <module>
    first_name, last_name = name_helper.split_name(name)
  File "/home/runner/grudges/utils/name_helper.py", line 30, in split_name
    return split_name_three_plus(names)
  File "/home/runner/grudges/utils/name_helper.py", line 8, in split_name_three_plus
    elif name[0].islower():
IndexError: string index out of range
> import pytest; pytest.main()
platform linux -- Python 3.8.3, pytest-6.0.1, py-1.9.0, pluggy-0.13.1
rootdir: /home/runner/grudges
collected 5 items

tests/test_name_helper.py ..... [100%]

5 passed in 0.04s
<ExitCode.OK: 0>
```

Image 8: Triggering a new error and invoking pytest from the output pane.

## Where next

You've learned to do TDD in this project. It's a popular style of programming, but it's not for everyone. Even if you decide not to use TDD, having tests is still very useful and it's not uncommon for large projects to have thousands or millions of tests.

Take a look at the [big list of naughty strings<sup>54</sup>](#) for a project that collects inputs that often cause software to break. You could also read [How SQLite Is Tested<sup>55</sup>](#) which explains how SQLite, a popular lightweight database, has 150 thousand lines of code and nearly 100 million(!) lines of tests.

In the next tutorial, we'll show you how to become a Repl.it poweruser by taking advantage of the productivity features it offers.

<sup>54</sup><https://github.com/minimaxir/big-list-of-naughty-strings>

<sup>55</sup><https://www.sqlite.org/testing.html>

# Productivity hacks

After coding for a while, you may find that there are some repetitive things that take up unnecessary time. For example, searching for and updating a variable name can seem laborious. Luckily, Repl.it has some built-in productivity tools that we'll take a look at in this tutorial.

Specifically, you'll see how to:

- Make simultaneous changes in several parts of your file using multiple cursors
- Use keyboard shortcuts to quickly carry out tasks without the delay of reaching for your mouse
- Switch to Vim or Emacs keybindings for full mouseless control.

Similarly to learning to touch type, there is often a steep *learning curve* when you start to use advanced code editing features. They might even substantially slow you down at first, but once you master them you'll soar past the limits of what you could achieve without these aids.

## Using the command palette

The command palette is the easiest way to see the tools available to you as well as their corresponding keybindings. It's a drop-down menu that will appear in the coding pane at the push of a button.

To access the command palette press F1 or Ctrl+Shift+P (cmd+shift+P on MacOS). Note that if you're using Firefox the second option will open an incognito window instead.

You can use the shortcuts directly from the command palette by selecting the code you wish to edit and clicking on the command in the drop-down menu, or use it to refresh your memory on the keybindings associated with the shortcuts you use often.

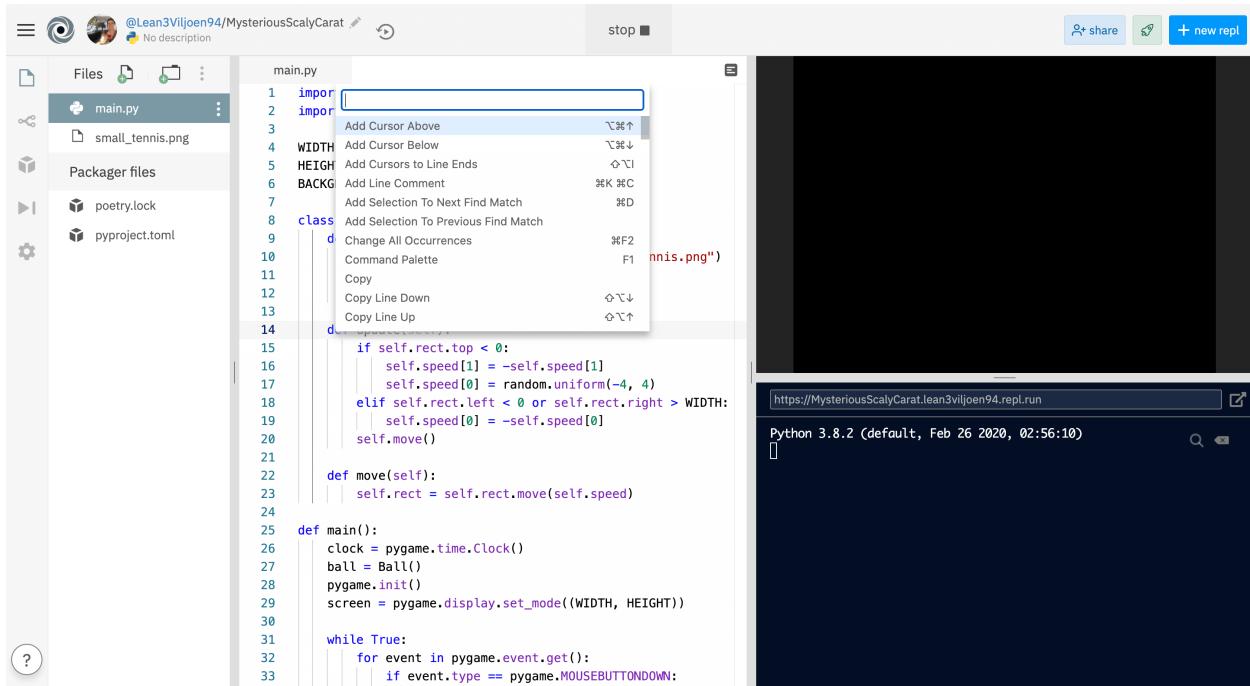


Image 1: *Opening the command palette.*

Let's take a look at how these work by editing the [PyGame juggling project](#)<sup>56</sup> that we covered in a previous tutorial<sup>57</sup>.

Instead of carrying out the suggested operations as you usually would, use Repl.it's productivity features instead.

## Duplicating entire lines

Sometimes you need two very similar lines of code directly after each other. Instead of copying and pasting the line or typing it out again, you can use the *duplicate row* feature, which will replicate the current line either above or below.

For example, our juggling project includes the following code to instantiate the initial three balls.

```
1 ball1 = Ball()
2 ball2 = Ball()
3 ball3 = Ball()
```

Instead of typing out all three lines, you can type out the first one, leave your cursor position on that line, and press Shift+Alt+down (shift+option+down on MacOS) twice. This will create two copies of the line, directly below the original one, and then you can simply change the number in the variable to account for the second two balls.

<sup>56</sup><https://repl.it/@GarethDwyer1/cwr-07-juggling-with-pygame>

<sup>57</sup><http://www.codewithrepl.it/07-building-a-game-with-pygame.html>



Image 2: Copying the current selected line.

## Deleting entire lines

There may be instances where you'd want to delete large chunks of code at a time (it happens to the best of us!).

Pressing `Ctrl+Shift+K` (`cmd+shift+K` on MacOS) deletes the line underneath your cursor (or if you have multiple lines selected it will delete all of them.)

Instead of deleting the entire line, you can also delete from your cursor up to the end of the line or from your cursor to the beginning of the line. The shortcuts for these are

- `Ctrl+Backspace` (`cmd+backspace` on MacOS) to delete backwards
- `Ctrl+K` (same on MacOS) to delete forwards

As an example, below you can see how we might use this to first delete one of the our `elif` blocks by doing two “delete line” operations. We then change our random speed to be constant by using a “delete to end of line” operation from the `=` sign and then typing our constant.



Image 3: Deleting selected lines of code.

## Inserting blank lines

It's also common to need to add a new line of code above or below the current one. Instead of using your mouse or arrow keys to get to the right place and then pressing `Enter`, you can instead use an “insert line” operation.

Press `Ctrl+Shift+Enter` (`cmd+shift+enter` on MacOS) to insert a blank line directly above the current one and move the cursor to the start of it (Repl.it will even maintain the current level of indentatin for you).



Image 4: Inserting lines.

## Indenting and dedenting lines

When writing Python, you probably pay more attention to whitespace (spaces or tabs) than in other languages, which use braces to handle logic. You're probably used to indenting and dedenting using Tab and Shift+Tab, which requires you to first place the cursor at the start of the line.

Instead you can use Ctrl+] (cmd+] on MacOS) to indent and dedent the line no matter where your cursor is. For example, if you need to fix the indentation in the following code, you can

- \* put your cursor on the for line
- \* press Ctrl+]
- \* press down
- \* press Shift+down
- \* press Ctrl+] again twice.

Now your code's indentation will be fixed.



Image 5: *Indenting a line.*

## Moving blocks of code within a file

Sometimes you need to move a block of code up or down in the file. For example, our `update()` function uses our `move()` function, but `move()` is only defined later. For readability, it's good to try ensure that your functions only call functions that have already been defined further up (assuming that someone else is reading the code top down, they will remember the `move()` function's definition before seeing it used).

Instead of cutting and pasting a block, you can shunt it by pressing Alt+up or Alt+down (option+up and option+down on MacOS). As with the others, this works on the line under your cursor or a larger selection.



Image 6: *Moving the current line selection.*

## Adding cursors

Sometimes it's useful to make exactly the same changes in multiple places at once. For example, we might want to rename our `speed` attribute to `velocity`. Put your cursor anywhere on the word that you want to change and press Ctrl+D (cmd+D on MacOS). Repeatedly press Ctrl+D to select matching words individually, each with their own cursor. Now you can apply edits and they will appear at each selection, as below.



Image 7: Adding cursors to multiple instances of the same selection.

If you want multiple cursors on consecutive lines, press `Ctrl+Alt+up` or `Ctrl+Alt+down` (`cmd+option+up` and `cmd+option+down` on MacOs). For example, if we want a square game we could change both width and height to be 1000 simultaneously as follows.



Image 8: Adding cursors to multiple lines.

## Navigating to specific pieces of code

Sometimes, especially in larger projects, you'll call a function or instantiate an object far from where that function or object is defined (either thousands of lines away in the same file, or in a different file altogether).

If you're reading a piece of code that calls a function and you want to quickly see what that function actually does, you can use the **go to definition** keybinding (`F12` or `cmd+F12` on MacOs). This will jump to the definition of the function or class selected. The **peek definition** has a similar functionality, but instead of jumping to the definition, it opens in a separate modal. For example, below, the cursor is on the instantiation of `Ball()` and we can quickly see how this class is defined.



Image 9: Peeking the definition.

The **go to line** operation (`Ctrl+G`) allows you to navigate to a line by giving its line number. This is useful to track down the source of those error messages that tell you what line had an issue, or if you're on a call with someone who says "I'm looking at line 23" and you can quickly jump to the same place.

Finally, you can open a specific file by searching for a part of the name by pressing `Ctrl+P` (`cmd+P` on MacOs), which can be quicker than scrolling through the files pane if you have a lot of files.



Image 10: Opening existing files.

## Vim and Emacs key bindings

Once you get hooked on keyboard shortcuts, you might wonder if you ever need to use your mouse again. Most of the time it only slows you down. Luckily, people thought of this decades ago. Before mice existed, all text editing was done using only a keyboard, and many developers still prefer editors that were created in this setting over more modern ones.

The two main keyboard-focused text editors are called [Vim](#)<sup>58</sup> and [Emacs](#)<sup>59</sup>. They both have steep learning curves (and there's a long-standing tradition that users of either fiercely argue about which is superior), but once you've put in the time to master them you can get rid of your mouse for good.

If you've gotten used to either, you can emulate the experience in Repl.it by switching your keybinds. Go to the "Settings" tab and scroll down to where you can toggle between "default", "emacs" and "vim".

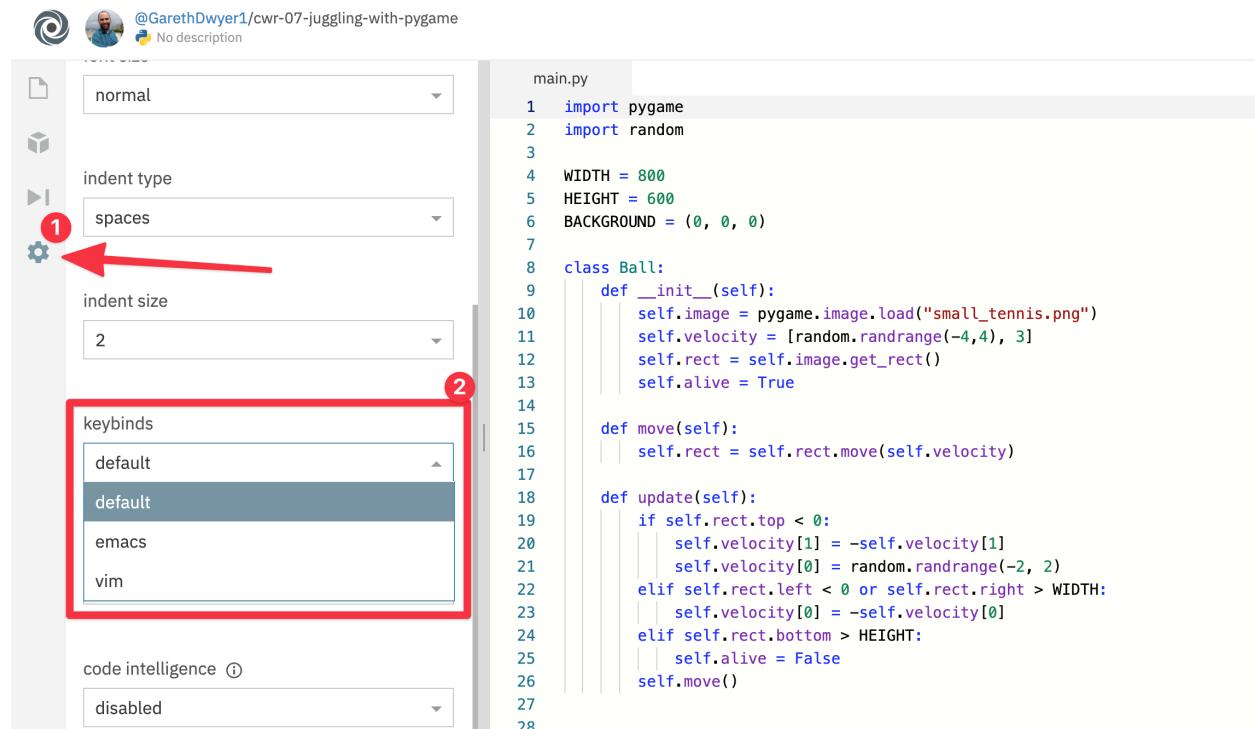


Image 11: Setting your keybinds to vim or emacs.

## Make it your own

If you want to keep hacking on the PyGame project using your new keyboard prowess, you can continue from where we left off below.

<sup>58</sup><https://www.vim.org/>

<sup>59</sup><https://www.gnu.org/software/emacs/>

```
<iframe height="400px" width="100%" src="https://repl.it/@GarethDwyer1/cwr-10-productivity?lite=true" scrolling="no" frameborder="no" allowtransparency="true" allowfullscreen="true" sandbox="allow-forms allow-pointer-lock allow-popups allow-same-origin allow-scripts allow-modals"></iframe>
```

## Where next?

You've reached the end of this collection of tutorials that teach you the ins and outs of Repl.it, and you should be able to build any project that you can imagine now.

If you're stuck for ideas, continue on to Part 3 where we'll walk you through 8 practical projects, focusing more on coding concepts than Repl.it features.

# Beginner web scraping with Python and Repl.it

In this guide, we'll walk through how to grab data from web sites automatically. Most websites are created with a *human* audience in mind - you use a search engine or type a URL into your web browser, and see information displayed on the page. Sometimes, we might want to *automatically* extract and process this data, and this is where web scraping can save us from boring repetitive labour. We can create a custom computer program to visit web sites, extract specific data and process this data in a particular way.

We'll be extracting news data from the [bbc.com<sup>60</sup>](https://bbc.com/news) news website, but you should be able to adapt it to extract information from any website that you want with a bit of trial and error.

There are many reasons you might wish to use web scraping. For example, you might need to:

- extract numbers from a report that is released weekly and published online
- grab the schedule for your favourite sports team as it's released
- find the release dates for upcoming movies in your favourite genre
- be notified automatically when a website changes

There are many other use cases for web scraping. However, you should also note that copyright law and web scraping laws are complex and differ by country. As long as you aren't blatantly copying their content or doing web scraping for commercial gain, people generally don't mind web scraping. However, there have been some legal cases involving [scraping data from LinkedIn<sup>61</sup>](#) and media attention from [scraping data from OKCupid<sup>62</sup>](#). Web scraping can violate the law, go against a particular website's terms of service, or breach ethical guidelines - so take care with where you apply this skill.

With the disclaimer out of the way, let's learn how to scrape!

## Overview and requirements

Specifically, in this tutorial we'll cover:

- What a website really is and how HTML works

---

<sup>60</sup><https://bbc.com/news>

<sup>61</sup><https://techcrunch.com/2016/08/15/linkedin-sues-scrapers/>

<sup>62</sup><https://www.engadget.com/2016/05/13/scientists-release-personal-data-for-70-000-okcupid-profiles/>

- Viewing HTML in your web browser
- Using Python to download web pages
- Using [BeautifulSoup<sup>63</sup>](#) to extract parts of scraped data

We'll be using the online programming environment [Repl.it<sup>64</sup>](#) so you won't need to install any software locally to follow along step by step. If you want to adapt this guide to your own needs, you should create a free account by going to [repl.it<sup>65</sup>](#) and follow their sign up process.

It would help if you have basic familiarity with Python or another high-level programming language, but we'll be explaining each line of code we write in detail so you should be able to keep up, or at least replicate the result, even if you don't.

## Webpages: beauty and the beast

You have no doubt visited web pages using a web browser before. Websites exist in two forms:

1. The one you are used to, where you can see text, images, and other media. Different fonts, sizes, and colours are used to display information in a useful and (usually) aesthetic way.
2. The "source" of the webpage. This is the computer code that tells your web browser (e.g. Mozilla Firefox or Google Chrome) what to display and how to display it.

Websites are created through a combination of three computer languages: HTML, CSS and JavaScript. This in itself is a huge and complicated field with a messy history, but having a basic understanding of how some of it works is necessary to automate web scraping effectively. If you open any website in your browser and right-click somewhere on the page, you'll see a menu which should include an option to "view page source" – to inspect the code form of a website, before your web browser interprets it.

This is shown in the image below: a normal web page on the left, with an open menu (displayed by right-clicking on the page). Clicking "view page source" on this menu produces the result on the right – we can see the code that contains all the data and supporting information that the web browser needs to display the complete page. While the page on the left is easy to read, use, and looks good, the one on the right is a monstrosity. It takes some effort and experience to make any sense of it, but it's possible and necessary if we want to write custom web scrapers.

---

<sup>63</sup><https://www.crummy.com/software/BeautifulSoup/>

<sup>64</sup><https://repl.it>

<sup>65</sup><https://repl.it>

**Why Hollywood writer Ubah Mohamed hated her name**

By Naima Mohamud  
BBC Africa

○ 8 hours ago

f t Share



Getting a TV job in the immediate aftermath of the 9/11 terror attacks in the US with a name like Ubah Mohamed was not easy.

"As a test, I changed my name and I immediately got offered work as a production assistant," says Mohamed, who opted for familiar Western names.

"I had a new alias every week. It was frustrating," she told me on the phone from California.

Growing up in the US, it was difficult being viewed as "different" and choosing an "American name" was a path many of her contemporaries with unfamiliar names took, she says.

**Top Stories**

- Huge surge in Indonesia quake toll ○ 39 minutes ago
- 'Dozens trapped' under collapsed hotel ○ 1 hour ago
- Musk out as Tesla chair over fraud case ○ 7 hours ago

**Features**



Special words that don't exist in English (yet)

Back Forward Reload Page ⌘R

Bookmark Page ⌘D

Save Page as... ⌘S

Find... ⌘F

Print... ⌘P

View Page Source ⌘U

Inspect Element

```

1 <!DOCTYPE html>
2 <html lang="en" id="responsive-news">
3   <head prefix="og: http://ogp.me/ns#>
4     <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
5     <title>Why Hollywood writer Ubah Mohamed hated her name - BBC News</title>
6     <meta name="description" content="The Hollywood writer who hated her name used aliases to get work before breaking into Hollywood.">
7     <link rel="preload" as="style" href="https://static.bbc.co.uk/news/1_258_03020/styleheets/services/news/compat.css" media="(min-width: 600px)">
8     <link rel="preload" as="script" href="https://static.bbc.co.uk/news/1_258_03020/javascript/services/news/compat.js" media="(min-width: 600px)">
9     <script> window.ontouchstart = true; </script>
10    <link rel="preload" as="script" href="https://fbcdn.on.net/fbsr/fig_isCallback=orb.orb_fig">
11    <link href="/static.bbci.co.uk/x" rel="preconnect" crossorigin="origin">
12    <link href="/static.bbci.co.uk/x" rel="dns-prefetch" crossorigin="origin">
13    <link href="/i/chef.bbci.co.uk" rel="preconnect" crossorigin="origin">
14    <link rel="dns-prefetch" href="/api.bbci.co.uk/">
15    <link rel="dns-prefetch" href="/za.bbci.co.uk/">
16    <link rel="dns-prefetch" href="/i/chef.bbci.co.uk/">
17    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
18    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
19    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
20    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
21    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
22    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
23    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
24    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
25    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
26    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
27    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
28    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
29    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
30    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
31    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
32    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
33    <link rel="dns-prefetch" href="https://www.bbci.co.uk/news/world-africa-45362050">
34    <meta name="x-country" content="za">
35    <meta name="x-audience" content="International">
36    <meta name="og:type" content="article">
37    <meta name="og:title" content="The Hollywood writer who hated her name" />
38    <meta property="og:type" content="article">
39    <meta property="og:title" content="Ubah Mohamed used aliases to get work before breaking into Hollywood." />
40    <meta property="og:image" content="https://ichef.bbci.co.uk/images/ic/45362050.jpg" />
41    <meta property="og:site_name" content="BBC News" />
42    <meta property="article:author" content="https://www.facebook.com/bbcnews" />
43    <meta property="og:url" content="https://www.bbci.co.uk/news/world-africa-45362050" />
44    <meta property="og:image" content="https://ichef.bbci.co.uk/images/ic/1024/branded_news/1193P/production/_103499917_d6b0ba3e-0241-4917-1c0d-001a00000000.jpg" />
45    <meta property="fb:app_id" content="87838159186,13923058827668140,74274325875840,18524858168196,15806587783370,13782079598335,193439354048/8930658,260212261139,29446213128,1086413581439034,283348121682953,25983058646,23931389545417,304714573046,310719525611571,647687223,661398,285361880228,51243982152360,238039464549831,176663550714,2059670664989230,10097870664989230,15286229625,1221030879723;67>
46    <meta name="twitter:card" content="summary_large_image">
47    <meta name="twitter:site" content="@bbcNewsId">
48    <meta name="twitter:title" content="The Hollywood writer who hated her name" />
49    <meta name="twitter:description" content="Ubah Mohamed used aliases to get work before breaking into Hollywood." />
50    <meta name="twitter:creator" content="@bbcNewsId">
51    <meta name="twitter:image:src" content="https://ichef.bbci.co.uk/images/ic/1024/branded_news/1193P/production/_103499917_d6b0ba3e-0241-4917-1c0d-001a00000000.jpg" />
52    <meta name="twitter:domain" content="www.bbci.com" />

```

Normal and source view of the same BBC news article.

## Navigating the source code using Find

The first thing to do is to work out how the two pages correspond: which parts of the normally displayed website match up to which parts of the code. You can use "find" (Ctrl + F) in the source code view to find specific pieces of text that are visible in the normal view to help with this. In the web page on the left, we can see that the story starts with the phrase "Getting a TV job". If we search for this phrase in the code view, we can find the corresponding text within the code, on line 805.



```

782 </ui>
783
784
785     </div>
786
787         </div>
788         </div>
789     <div id="topic-tags"><div id="u516982898581773"><noscript></noscript></div></div>
790     </div>
791
792     <div class="story-body__inner" property="articleBody">
793         <figure class="media-landscape no-caption full-width lead">
794             <span class="image-and-copyright-container">
795
796                 
797
798
799
800             <span class="off-screen">Image copyright</span>
801             <span class="story-image-copyright">Ubah Mohamed</span>
802
803         </span>
804
805     </figure><p class="story-body__introduction">Getting a TV job in the immediate aftermath of the 9/11 terror attacks in the US with a
Mohamed was not easy. </p><p>"As a test, I changed my name and I immediately got offered work as a production assistant," says Mohamed, who
Western names. </p><p>"I had a new alias every week. It was frustrating," she told me on the phone from California. </p><p>Growing up in the
difficult being viewed as "different" and choosing an "American name" was a path many of her contemporaries with unfamiliar names took, she
family's move to Memphis from New York when she was aged 13 was when she found her Somali identity most problematic.</p><div id="bbccom_mpu_
class="bbccom_slot mpu-ad" aria-hidden="true">
806     <div class="bbccom_advert">
807         <script type="text/javascript">
808             /**
809             (function() {
810                 if (window.bbcdotcom && bbcdotcom.adverts && bbcdotcom.adverts.slotAsync) {
811                     bbcdotcom.adverts.slotAsync('mpu', [1,2,3]);
812                 }
813             })();
814             /**
815         </script>
816     </div>

```

#### Finding text in the source code of a web page.

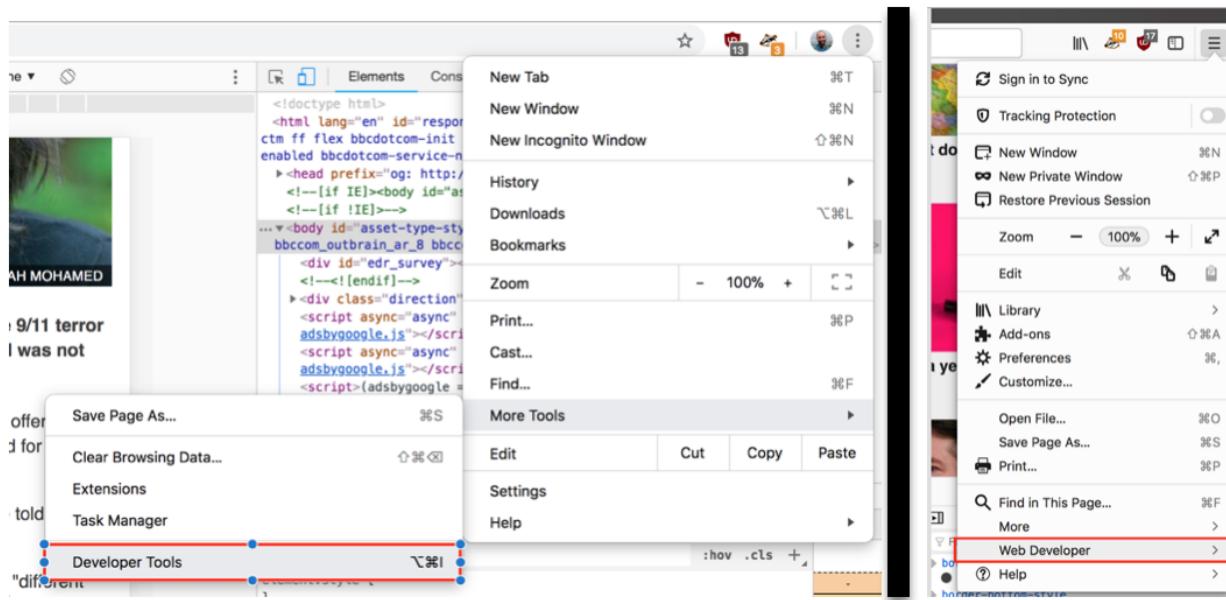
The `<p class="story-body__introduction">` just before the highlighted section is HTML code to specify that a paragraph (`<p>` in HTML) starts here and that this is a special kind of paragraph (an introduction to the story). The paragraph continues until the `</p>` symbol. You don't need to worry about understanding HTML completely, but you should be aware that it contains **both** the text data that makes up the news article and additional data about how to display it.

A large part of web scraping is viewing pages like this to a) identify the data that we are interested in and b) to separate this from the markup and other code that it is mixed with. Even before we start writing our own code, it can still be tricky first to understand other people's.

In most pages, there is a lot of code to define the structure, layout, interactivity, and other functionality of a web page, and relatively little that contains the actual text and images that we usually view. For especially complex pages it can be quite difficult, even with the help of the find function, to locate the code that is responsible for a particular part of the page. For this reason, most web browsers come with so-called "developer tools", which are aimed primarily at programmers to assist in the creation and maintenance of web sites, though these tools are also handy for doing web scraping.

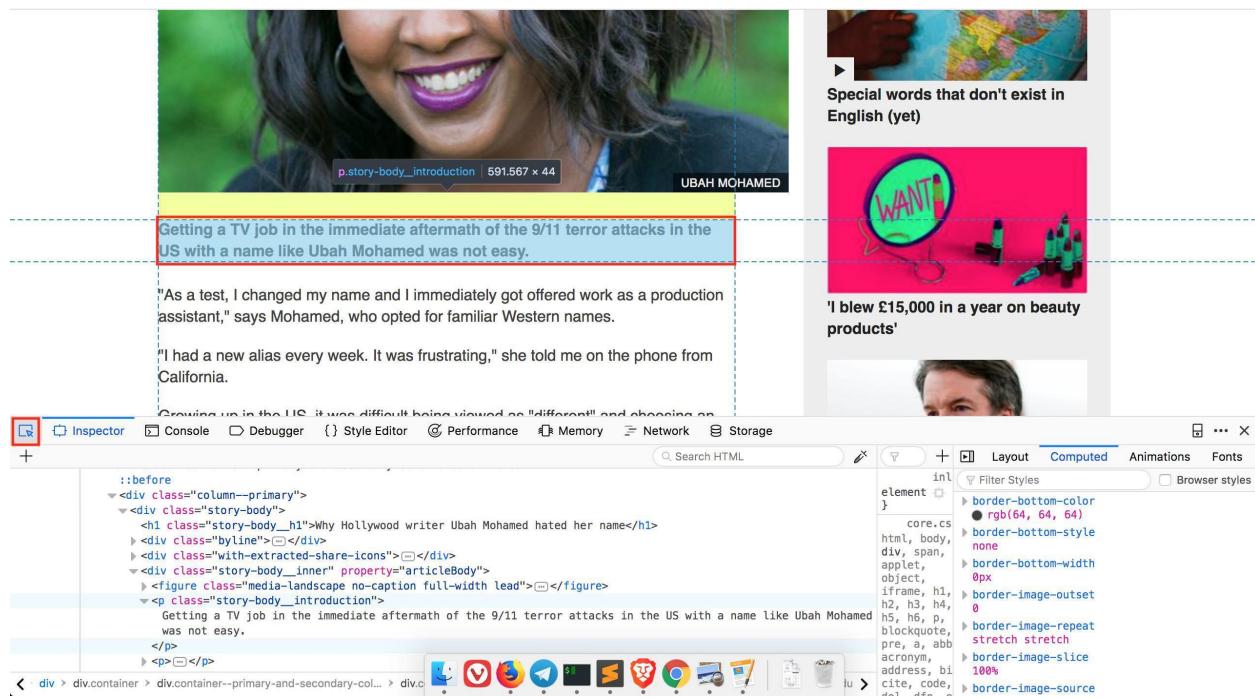
## Navigating the source code using developer tools

You can open the developer tools for your browser from the main menu, with Google Chrome shown on the left and Mozilla Firefox on the right below. If you're using a different web browser, you should be able to find a similar setting.



Opening Developer Tools in Chrome (left) and Firefox (right)

Activating the tool brings up a new panel in your web browser, usually at the bottom or on the right-hand side. The tool contains an “Inspector” panel and a selector tool, which can be chosen by pressing the icon highlighted in red below. Once the selector tool is active, you can click on parts of the web page to view the corresponding source code. In the image below, we selected the same first paragraph in the normal view and we can see the `<p class="story-body__introduction">` code again in the panel below.



### Viewing the code for a specific element using developer tools

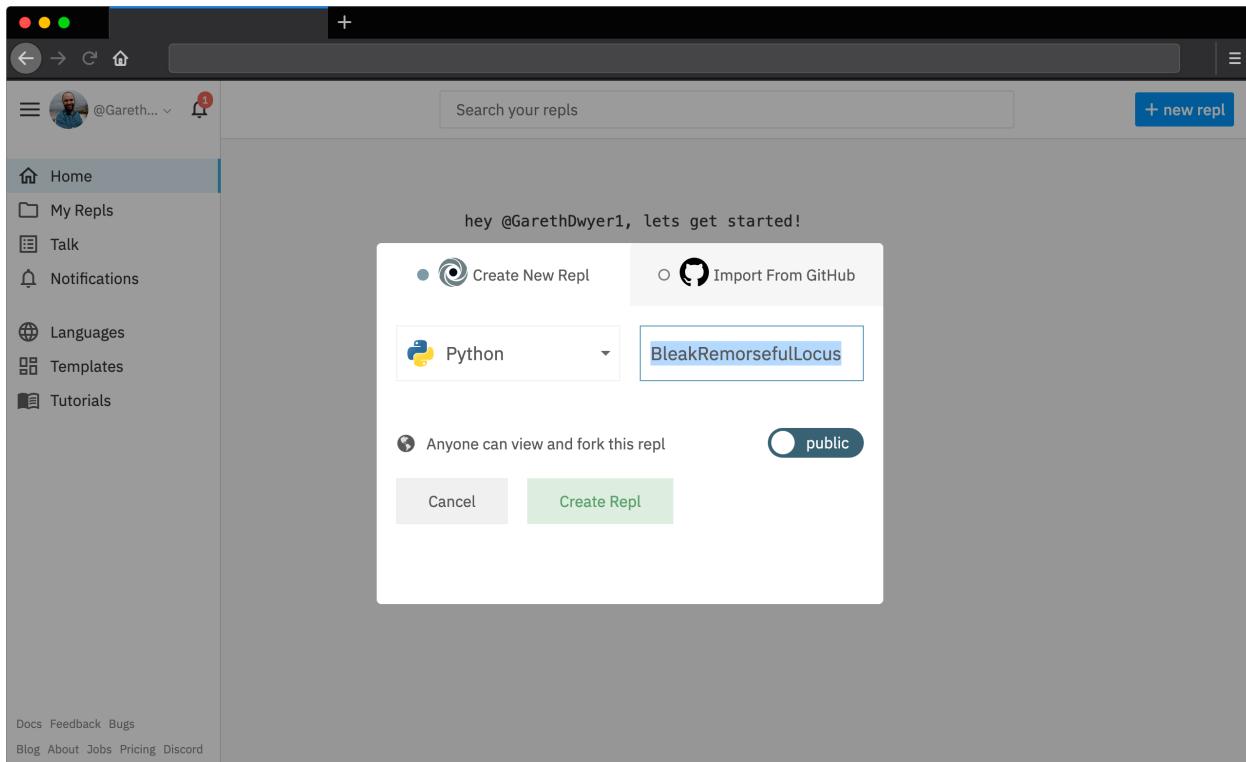
The Developer Tools are significantly more powerful than using the simple find tool, but they are also more complicated. You should choose a method based on your experience and the complexity of the page that you are trying to analyze.

## Downloading a web page with Python

Now that we've seen a bit more of how web pages are built in our browser, we can start retrieving and manipulating them using Python. Since Python is not a web browser, we'll only be able to retrieve and manipulate the HTML source code, rather than viewing the 'normal' representation of a web page.

We'll do this through a Python Repl using the `requests` library. Open [repl.it](https://repl.it)<sup>66</sup> and choose to create a new Python Repl.

<sup>66</sup><https://repl.it>

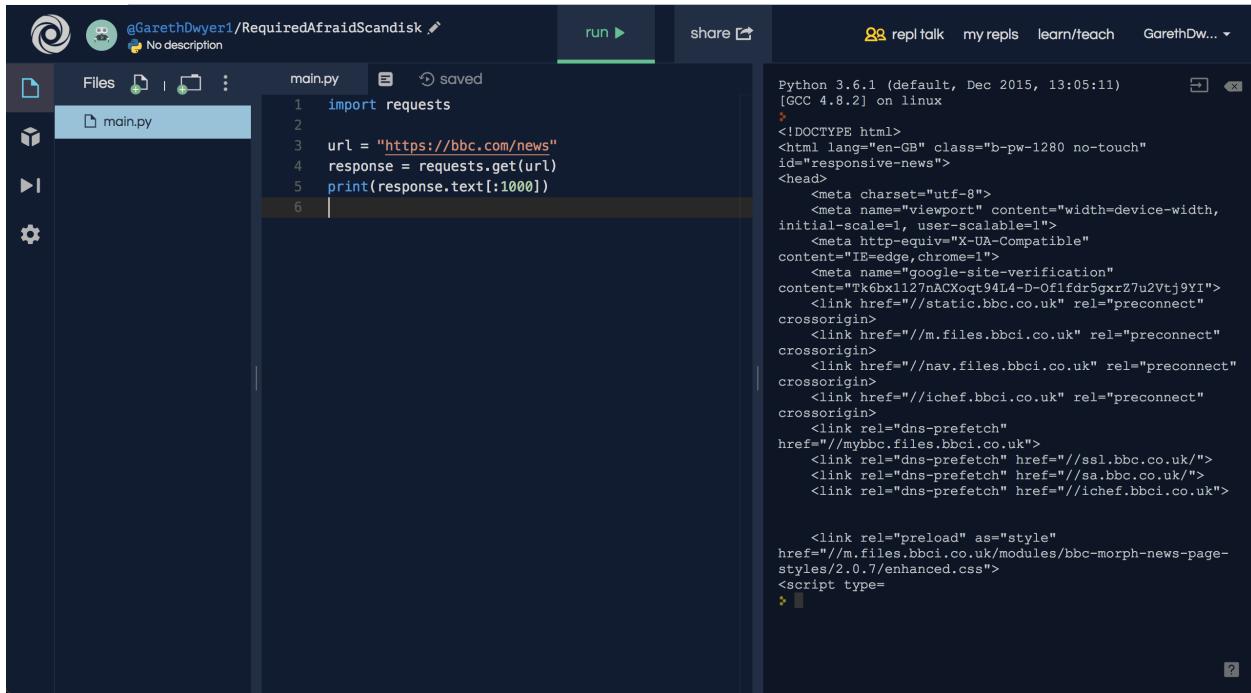


This will take you to a working Python coding environment where you can write and run Python code. To start with, we'll download the content from the BBC News homepage, and print out the first 1000 characters of HTML source code.

You can do this with the following four lines of Python:

```
1 import requests  
2  
3 url = "https://bbc.com/news"  
4 response = requests.get(url)  
5 print(response.text[:1000])
```

Put this code in the `main.py` file that Repl automatically creates for you and press the “Run” button. After a short delay, you should see the output in the output pane - the beginning of HTML source code, similar to what we viewed in our web browser above.



The screenshot shows a Repl.it interface. On the left, there's a sidebar with icons for Files, GitHub, and Settings. The main area has tabs for 'main.py' and 'saved'. The 'main.py' tab contains the following code:

```

1 import requests
2
3 url = "https://bbc.com/news"
4 response = requests.get(url)
5 print(response.text[:1000])
6

```

To the right of the code editor is a terminal window showing the output of the script. The output is the first 1000 characters of the BBC News homepage's HTML source code.

Downloading a single page using Python

Let's pull apart each of these lines.

- In line 1, we import the Python `requests` library, which is a library that allows us to make web requests.
- In line 3, we define a variable containing the URL of the main BBC news site. You can visit this URL in your web browser to see the BBC News home page.
- In line 4, we pass the URL we defined to the `requests.get` function, which will visit the web page that the URL points to and fetch the HTML source code. We load this into a new variable called `response`.
- In line 5, we access the `text` attribute of our `response` object, which contains all of the HTML source code. We take only the first 1000 characters of this, and pass them to the `print` function, which simply dumps the resulting text to our output pane.

We have now automatically retrieved a web page and we can display parts of the content. We are unlikely to be interested in the full source code dump of a web page (unless we are storing it for archival reasons), so let's extract some interesting parts of the page, instead of only the first 1000 characters.

## Using BeautifulSoup to extract all URLs

The world wide web is built from pages that link to each other using hyperlinks, links, or URLs. (These terms are all used more-or-less interchangeably).

Let's assume for now that we want to find all the news articles on the BBC News homepage, and get their URLs. If we look at the main page below, we'll see there are a bunch of stories on the home page. By mousing over any of the headlines with the "inspect" tool, we can see that each has a unique URL which takes us to that news story. For example, mousing over the main "US and Canada agree new trade deal" story in the image below is a link to <https://www.bbc.com/news/business-45702609>.

If we inspect that element using the browser's developer tools, we can see it is a `<a>` element, which is HTML for a link, with an `<href>` component that points to the URL. Note that the `href` section goes only to the last part of the URL, omitting the `https://www.bbc.com` part. Because we are already on BBC, the site can use *relative URLs* instead of *absolute URLs*. This means that when you click on the link, your browser will figure out that the URL isn't complete and prepend it with `https://www.bbc.com`. If you look around the source code of the main BBC page, you'll find both relative and absolute URLs, which already makes scraping all of the URLs on the page more difficult.

The screenshot shows the BBC News homepage with several news items. One story is highlighted with a red box:

**US and Canada agree new trade deal**

The existing deal had governed more than a trillion dollars in trade between the US, Canada and Mexico.

1h | Business

Below the story, the URL <https://www.bbc.com/news/business-45702609> is highlighted with a red box in the browser's developer tools. The developer tools also show the HTML structure of the page, including the `<a>` element with the `href` attribute.

### Viewing headline links using Developer Tools.

We could try to use Python's built-in text search functions like `find()` or regular expressions to extract all of the URLs from the BBC page, yet it is not actually possible to do this reliably. HTML is a complex language which allows web developers to do many unusual things. For an amusing take on why we should avoid a "naive" method of looking for links, see [this very famous<sup>67</sup>](#) StackOverflow question and the first answer.

Luckily, there is a powerful and simple-to-use HTML parsing library called [BeautifulSoup<sup>68</sup>](#), which will help us extract all the links from a given piece of HTML. We can use it by modifying the code

<sup>67</sup><https://stackoverflow.com/questions/1732348/regex-match-open-tags-except-xhtml-self-contained-tags>

<sup>68</sup><https://www.crummy.com/software/BeautifulSoup/>

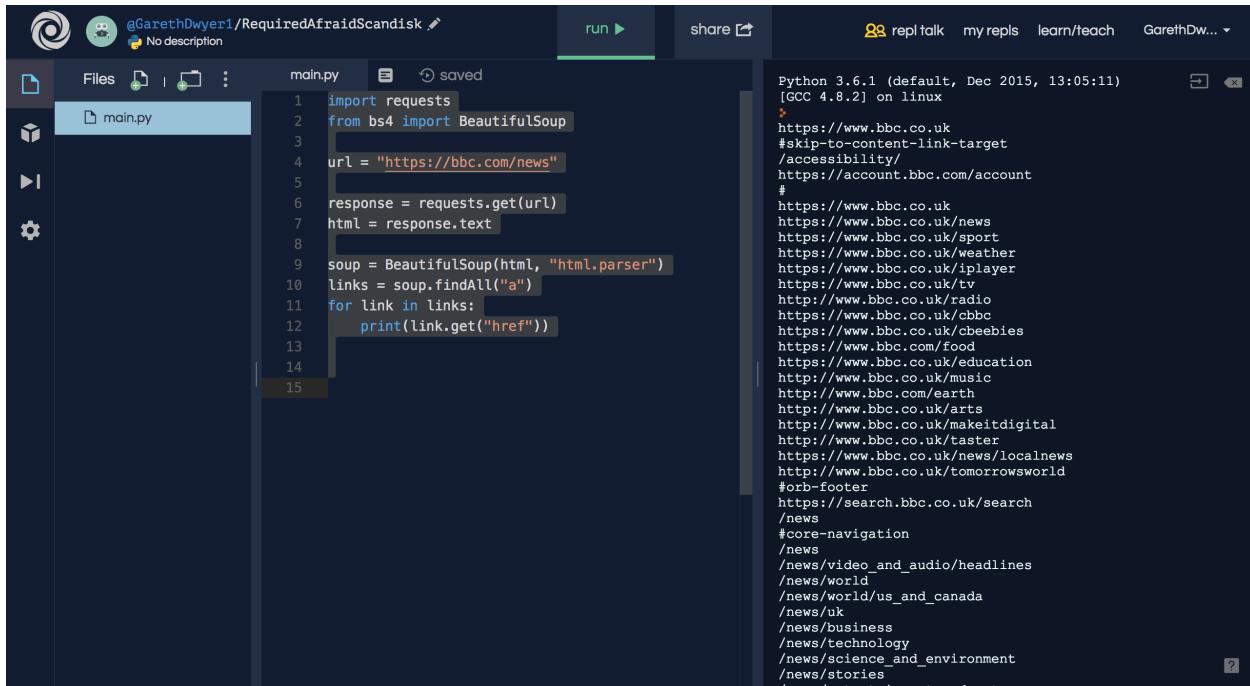
in our Repl to look as follows.

```

1 import requests
2 from bs4 import BeautifulSoup
3
4 url = "https://bbc.com/news"
5
6 response = requests.get(url)
7 html = response.text
8
9 soup = BeautifulSoup(html, "html.parser")
10 links = soup.findAll("a")
11 for link in links:
12     print(link.get("href"))

```

If you run this code, you'll see that it outputs dozens of URLs, one per line. You'll probably notice that the code now takes quite a bit longer to run than before – BeautifulSoup is not built into Python, it is a third-party module. This means that before running the code, Repl has to go and fetch this library and install it for you. Subsequent runs will be faster.



The screenshot shows a Repl.it interface with a dark theme. On the left, there's a sidebar with icons for files, terminal, and settings. The main area has tabs for 'Files' (with 'main.py' selected), 'Terminal', and 'Output'. The 'main.py' tab contains the provided Python code. The 'Terminal' tab shows the command 'Python 3.6.1 (default, Dec 2015, 13:05:11) [GCC 4.8.2] on linux' followed by a list of URLs extracted from the BBC News page. The 'Output' tab is empty.

```

Python 3.6.1 (default, Dec 2015, 13:05:11) [GCC 4.8.2] on linux
>
https://www.bbc.co.uk
#skip-to-content-link-target
/Accessibility/
https://account.bbc.com/account
#
https://www.bbc.co.uk
https://www.bbc.co.uk/news
https://www.bbc.co.uk/sport
https://www.bbc.co.uk/weather
https://www.bbc.co.uk/iplayer
https://www.bbc.co.uk/tv
http://www.bbc.co.uk/radio
https://www.bbc.co.uk/cbbc
https://www.bbc.co.uk/cbeebies
https://www.bbc.com/food
https://www.bbc.co.uk/education
http://www.bbc.co.uk/music
http://www.bbc.co.uk/earth
http://www.bbc.co.uk/arts
http://www.bbc.co.uk/makeitdigital
http://www.bbc.co.uk/taster
https://www.bbc.co.uk/news/localnews
http://www.bbc.co.uk/tomorrowsworld
#orb-footer
https://search.bbc.co.uk/search
/news
#core-navigation
/news
/news/video_and_audio/headlines
/news/world
/news/world/us_and_canada
/news/uk
/news/business
/news/technology
/news/science_and_environment
/news/stories

```

Extracting all links from BBC News.

The code is similar to what we had before with a few additions.

- In line 2, we import the BeautifulSoup library, which is used for parsing and processing HTML.

- One line 9, we transform our HTML into “soup”. This is BeautifulSoup’s representation of a web page, which contains a bunch of useful programmatic features to search and modify the data in the page. We use the “html.parser” option to parse HTML which is included by default – BeautifulSoup also allows you specify a custom HTML parser here. For example, you could install and specify a faster parser which can be useful if you need to process a lot of HTML data.
- In line 10, we find all the `a` elements in our HTML and extract them to a list. Remember, when we were looking at the URLs using our web browser (Image 7), we noted that the `<a>` element in HTML was used to define links, with the `href` attribute being used to specify where the link should go to. This line finds all of the HTML `<a>` elements.
- In line 11, we loop through all of the links we have, and in line 12 we print out the `href` section.

These last two lines show why BeautifulSoup is useful. To try and find and extract these elements without it would be remarkably difficult, but now we can do it in two lines of readable code!

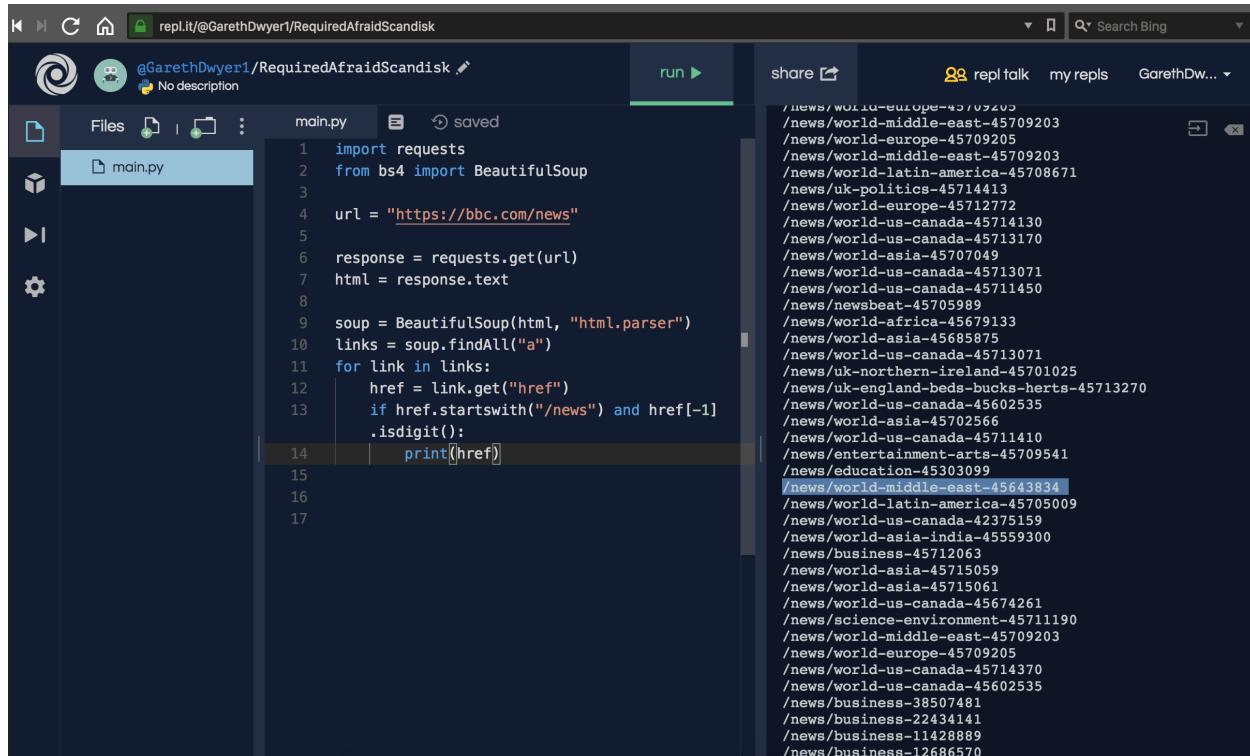
If we look at the URLs in the output pane, we’ll see quite a mixed bag of results. We have absolute URLs (starting with “http”) and relative ones (starting with “/”). Most of them go to general pages rather than specific news articles. We need to find a pattern in the links we’re interested in (that go to news articles) so that we can extract only those.

Again, trial and error is the best way to do this. If we go to the BBC News home page and use developer tools to inspect the links that go to news articles, we’ll find that they all have a similar pattern. They are relative URLs which start with “/news” and end with a long number, e.g. `/news/newsbeat-45705989`

We can make a small change to our code to only output URLs that match this pattern. Replace the last two lines of our Python code with the following four lines:

```
1 for link in links:  
2     href = link.get("href")  
3     if href.startswith("/news") and href[-1].isdigit():  
4         print(href)
```

Here we still loop through all of the links that BeautifulSoup found for us, but now we extract the `href` to its own variable immediately after. We then inspect this variable to make sure that it matches our conditions (starts with “/news” and ends with a digit), and only if it does, then we print it out.



```

import requests
from bs4 import BeautifulSoup

url = "https://bbc.com/news"
response = requests.get(url)
html = response.text

soup = BeautifulSoup(html, "html.parser")
links = soup.findAll("a")
for link in links:
    href = link.get("href")
    if href.startswith("/news") and href[-1].isdigit():
        print(href)

```

The screenshot shows a Repl.it interface with a Python file named 'main.py' containing the code above. The code uses the requests library to get the HTML of the BBC News homepage and BeautifulSoup to find all anchor tags. It then prints the href attribute of each link that starts with '/news' and ends with a digit, which corresponds to internal BBC URLs for news articles. The output window on the right shows the results of running the script, listing numerous BBC news URLs.

Printing only links to news articles from BBC.

## Fetching all of the articles from the homepage

Now that we have the link to every article on the BBC News homepage, we can fetch the data for each one of these individual articles. As a toy project, let's extract the proper nouns (people, places, etc) from each article and print out the most common ones to get a sense on what things are being talked about today.

Adapt your code to look as follows:

```

1 import requests
2 import string
3
4 from collections import Counter
5
6 from bs4 import BeautifulSoup
7
8
9 url = "https://bbc.com/news"
10
11

```

```
12 response = requests.get(url)
13 html = response.text
14 soup = BeautifulSoup(html, "html.parser")
15 links = soup.findAll("a")
16
17 news_urls = []
18 for link in links:
19     href = link.get("href")
20     if href.startswith("/news") and href[-1].isdigit():
21         news_url = "https://bbc.com" + href
22         news_urls.append(news_url)
23
24
25 all_nouns = []
26 for url in news_urls[:10]:
27     print("Fetching {}".format(url))
28     response = requests.get(url)
29     html = response.text
30     soup = BeautifulSoup(html, "html.parser")
31
32     words = soup.text.split()
33     nouns = [word for word in words if word.isalpha() and word[0] in string.ascii_uppercase]
34     all_nouns += nouns
35
36
37 print(Counter(all_nouns).most_common(100))
```

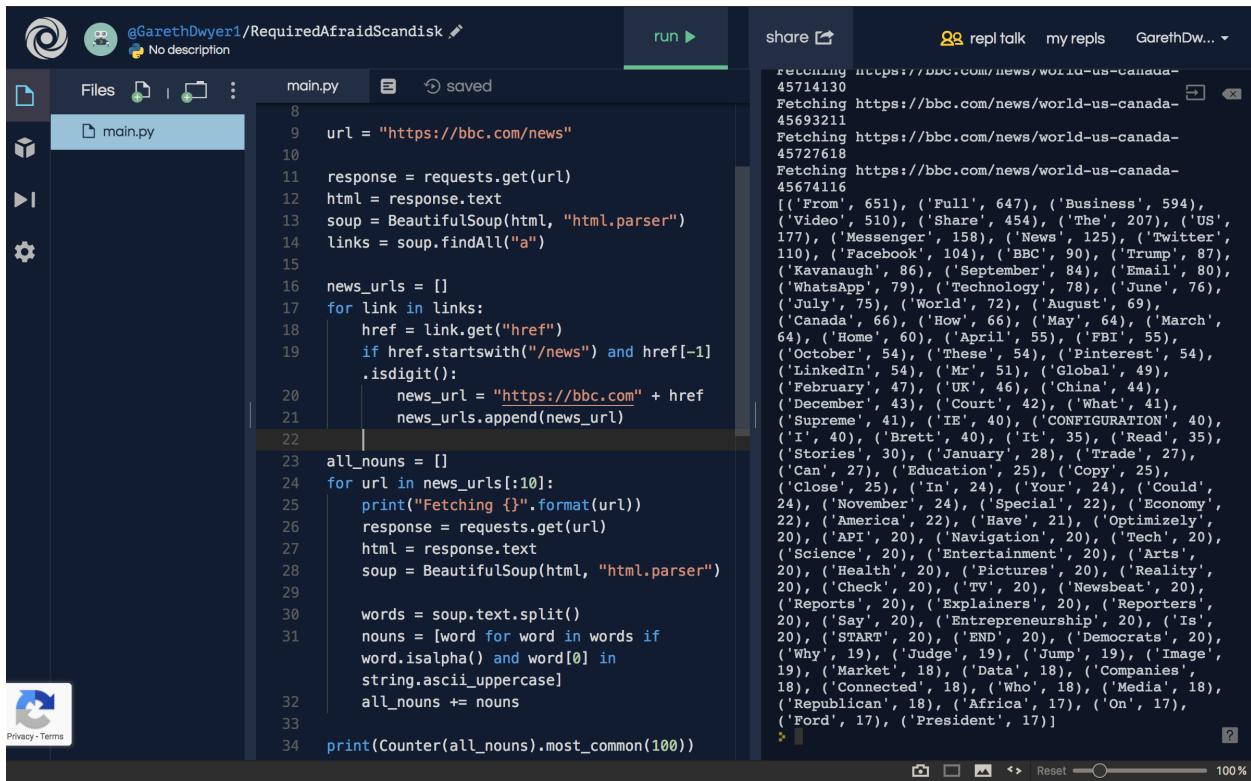
This code is quite a bit more complicated than what we previously wrote, so don't worry if you don't understand all of it. The main changes are:

- At the top, we add two new imports in addition to the `requests` library. The first new module is one for `string`, which is a standard Python module that contains some useful word and letter shortcuts. We'll use it to identify all the capital letters in our alphabet. The second module is a `Counter`, which is part of the built-in `collections` module. This will let us find the most common nouns in a list, once we have built a list of all the nouns.
- We've added `news_urls = []` at the top of the first `for` loop. Instead of printing out each URL once we've identified it as a "news URL", we add it to this list so we can download each page later. Inside the `for` loop two lines down, we combine the root domain ("http://bbc.com") with each `href` attribute and then add the complete URL to our `news_urls` list.
- We then go into another `for` loop, where we loop through the first 10 news URLs (if you have more time, you can remove the `[:10]` part to iterate through all the news pages, but for efficiency, we'll just demonstrate with the first 10).

- We print out the URL that we're fetching (as it takes a second or so to download each page, it's nice to display some feedback so we can see that the program is working).
- We then fetch the page and turn it into soup, as we did before.
- With `words = soup.text.split()` we extract all the text from the page and split this resulting big body of text into individual words. The Python `split()` function splits on white space, which is a crude way to extract words from a piece of text, but it will serve our purpose for now.
- The next line loops through all the words in that given article and keeps only the ones that are made up of numeric characters and which start with a capital letter (`string.ascii_uppercase` is just the uppercase alphabet). This is also an extremely crude way of extracting nouns, and we will get a lot of words (like those at the start of sentences) which are not actually proper nouns, but again it's a good enough approximation for now.
- We then add all the words that look like nouns to our `all_nouns` list and move on to the next article to do the same.
- Finally, once we've downloaded all the pages, we print out the 100 most common nouns along with a count of how often they appeared using Python's convenient `Counter` object.

You should see output similar to that in the image below (though your words will be different, as the news changes every few hours). We have the most common “nouns” followed by a count of how often that noun appeared in all 10 of the articles we looked at.

We can see that our crude extraction and parsing methods are far from perfect – words like “Twitter” and “Facebook” appear in most articles because of the social media links at the bottom of each article, so their presence doesn’t mean that Facebook and Twitter themselves are in the news today. Similarly, words like “From” aren’t nouns, and other words like “BBC” and “Business” are also included because they appear on each page, outside of the main article text.



The screenshot shows a Repl.it interface with a Python script named `main.py`. The code uses the `requests` library to fetch BBC news URLs and `BeautifulSoup` to parse the HTML. It extracts href attributes from links and filters them to only include news URLs that start with "news" and end with a digit. These URLs are then used to fetch the full article content, which is split into words and filtered to keep only nouns (words that are alphabetic and start with an uppercase letter). Finally, it prints the 100 most common nouns found across the first 10 news articles.

```

@GarethDwyer1/RequiredAfraidScandisk
No description

Files main.py saved
main.py
8 url = "https://bbc.com/news"
10
11 response = requests.get(url)
12 html = response.text
13 soup = BeautifulSoup(html, "html.parser")
14 links = soup.findAll("a")
15
16 news_urls = []
17 for link in links:
18     href = link.get("href")
19     if href.startswith("news") and href[-1]
20         .isdigit():
21             news_url = "https://bbc.com" + href
22             news_urls.append(news_url)
23
24 all_nouns = []
25 for url in news_urls[:10]:
26     print("Fetching {}".format(url))
27     response = requests.get(url)
28     html = response.text
29     soup = BeautifulSoup(html, "html.parser")
30
31     words = soup.text.split()
32     nouns = [word for word in words if
33             word.isalpha() and word[0] in
34             string.ascii_uppercase]
            all_nouns += nouns
35
36 print(Counter(all_nouns).most_common(100))

```

share ➔ repl talk my repls GarethDw... ▾

Fetching https://bbc.com/news/world-us-canada-45714130  
 Fetching https://bbc.com/news/world-us-canada-45693211  
 Fetching https://bbc.com/news/world-us-canada-45727618  
 Fetching https://bbc.com/news/world-us-canada-45674116  
 [ ('From', 651), ('Full', 647), ('Business', 594),  
 ('Video', 510), ('Share', 454), ('The', 207), ('US', 177), ('Messenger', 158), ('News', 125), ('Twitter', 110), ('Facebook', 104), ('BBC', 90), ('Trump', 87), ('Kavanaugh', 86), ('September', 84), ('Email', 80), ('WhatsApp', 79), ('Technology', 78), ('June', 76), ('July', 75), ('World', 72), ('August', 69), ('Canada', 66), ('How', 66), ('May', 64), ('March', 64), ('Home', 60), ('April', 55), ('FBI', 55), ('October', 54), ('These', 54), ('Pinterest', 54), ('LinkedIn', 54), ('Mr', 51), ('Global', 49), ('February', 47), ('UK', 46), ('China', 44), ('December', 43), ('Court', 42), ('What', 41), ('Supreme', 41), ('IE', 40), ('CONFIGURATION', 40), ('I', 40), ('Brett', 40), ('It', 35), ('Read', 35), ('Stories', 30), ('January', 28), ('Trade', 27), ('Can', 27), ('Education', 25), ('Copy', 25), ('Close', 25), ('In', 24), ('Your', 24), ('Could', 24), ('November', 24), ('Special', 22), ('Economy', 22), ('America', 22), ('Have', 21), ('Optimizely', 20), ('API', 20), ('Navigation', 20), ('Tech', 20), ('Science', 20), ('Entertainment', 20), ('Arts', 20), ('Health', 20), ('Pictures', 20), ('Reality', 20), ('Check', 20), ('TV', 20), ('Newsbeat', 20), ('Reports', 20), ('Explainers', 20), ('Reporters', 20), ('Say', 20), ('Entrepreneurship', 20), ('Is', 20), ('START', 20), ('END', 20), ('Democrats', 20), ('Why', 19), ('Judge', 19), ('Jump', 19), ('Image', 19), ('Market', 18), ('Data', 18), ('Companies', 18), ('Connected', 18), ('Who', 18), ('Media', 18), ('Republican', 18), ('Africa', 17), ('On', 17), ('Ford', 17), ('President', 17)]

The final output of our program, showing the words that appear most often in BBC articles.

## Where next?

We've completed the basics of web scraping and have looked at how the web works, how to extract information from web pages, and how to do some very basic text extraction. You will probably want to do something other than extract words from BBC! You can fork this Repl from <https://repl.it/@GarethDwyer1/beginnerwebscraping> and modify it to change which site it scrapes and what content it extracts. You can also join the [Repl Discord Server](#)<sup>69</sup> to chat with other developers who are working on similar projects and who will happily exchange ideas with you or help if you get stuck.

We have walked through a very flexible method of web scraping, but it's the "quick and dirty" way. If BBC updates their website and some of our assumptions (e.g. that news URLs will end with a number) break, our web scraper will also break.

Once you've done a bit of web scraping, you'll notice that the same patterns and problems come up again and again. Because of this, there are many frameworks and other tools that solve these common problems (finding all the URLs on the page, extracting text from the other code, dealing with changing web sites, etc), and for any big web scraping project, you'll definitely want to use these instead of starting from scratch.

<sup>69</sup>[https://discord.com/login?redirect\\_to=%2Fchannels%2F40me](https://discord.com/login?redirect_to=%2Fchannels%2F40me)

Some of the best Python web scraping tools are:

- **Scrapy<sup>70</sup>**: A framework used by people who want to scrape millions or even billions of web pages. Scrapy lets you build “spiders” – programmatic robots that move around the web at high speed, gathering data based on rules that you specify.
- **Newspaper<sup>71</sup>**: we touched on how it was difficult to separate the main text of an online news article from all the other content on the page (headers, footers, adverts, etc). This problem is an incredibly difficult one to solve. Newspaper uses a combination of manually specified rules and some clever algorithms to remove the “boilerplate” or non-core text from each article.
- **Selenium<sup>72</sup>**: we scraped some basic content without using a web browser, and this works fine for images and text. Many parts of the modern web are dynamic though – e.g. they only load when you scroll down a page far enough or click on a button to reveal more content. These dynamic sites are challenging to scrape, but Selenium allows you to fire up a real web browser and control it just as a human would (but automatically), and this allows you to access this kind of dynamic content.

There is no shortage of other tools, and a lot can be done simply by using them in combination with each other. Web scraping is a vast world that we’ve only just touched on, but we’ll explore some more web scraping use cases in the next chapter, in particular, building news word clouds.

---

<sup>70</sup><https://scrapy.org/>

<sup>71</sup><https://github.com/codelucas/newspaper>

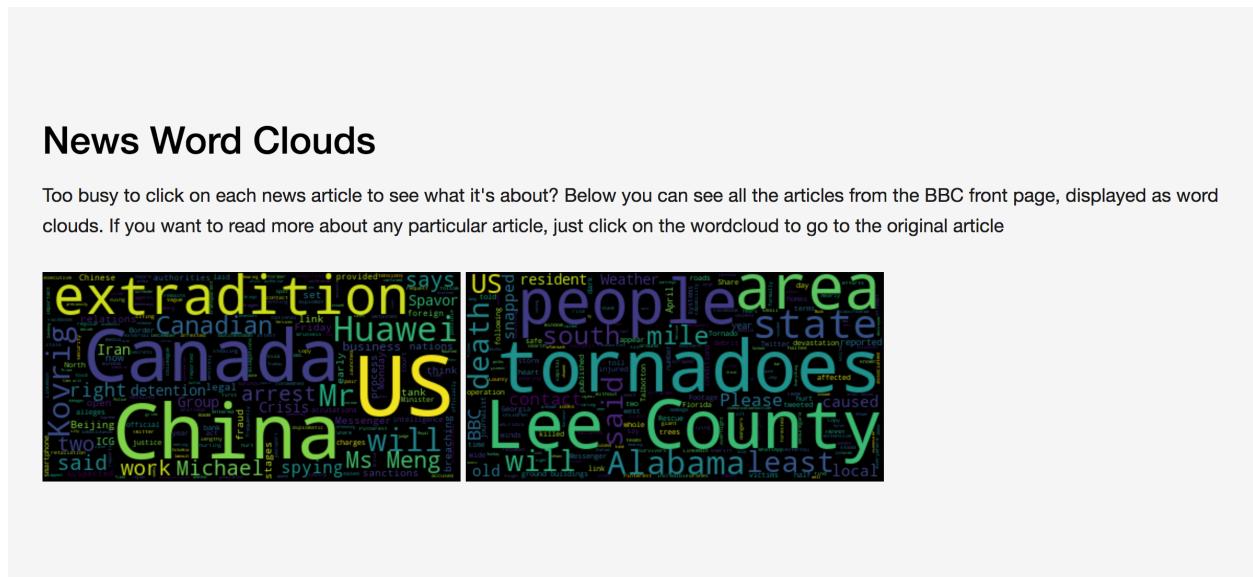
<sup>72</sup><https://www.seleniumhq.org/>

# Building news word clouds using Python and Repl.it

Word clouds, which are images showing scattered words in different sizes, are a popular way to visualise large amounts of text. Words that appear more frequently in the given text are larger, and less common words are smaller or not shown at all.

In this tutorial, we'll build a web application using Python and Flask that transforms the latest news stories into word clouds and displays them to our visitors.

At the end of this tutorial, our users will see a page similar to the one shown below, but containing the latest news headlines from BBC news. We'll learn some tricks about web scraping, RSS feeds, and building image files directly in memory along the way.



## **Overview and requirements**

We'll be building a simple web application step-by-step and explaining each line of code in detail. To follow, you should have some basic knowledge of programming and web concepts, such as what if statements are and how to use URLs. We'll be using Python for this tutorial, but we won't assume that you're a Python expert.

Specifically, we'll:

- Look at RSS feeds and how to use them in Python
- Show how to set up a basic Flask web application
- Use BeautifulSoup to extract text from online news articles
- Use WordCloud to transform the text into images
- Import Bootstrap and add some basic CSS styling

We'll be using the online programming environment [Repl.it<sup>73</sup>](#) so you won't need to install any software locally to follow along step by step. If you want to adapt this guide to your own needs, you should create a free account by going to [repl.it<sup>74</sup>](#) and follow their sign up process.

## Web scraping

We previously looked at basic web scraping in an [introduction to web scraping<sup>75</sup>](#). If you're completely new to the idea of automatically retrieving content from the internet, have a look at that tutorial first.

In this tutorial, instead of scraping the links to news articles directly from the BBC homepage, we'll be using [RSS feeds<sup>76</sup>](#) - an old but popular standardised format that publications use to let readers know when new content is available.

## Taking a look at RSS Feeds

RSS feeds are published as XML documents. Every time BBC (and other places) publishes a new article to their home page, they also update an XML machine-readable document at <http://feeds.bbci.co.uk/news/world/rss.xml>. This is a fairly simple feed consisting of a `<channel>` element, which has some metadata and then a list of `<item>` elements, each of which represents a new article. The articles are arranged chronologically, with the newest ones at the top, so it's easy to retrieve new content.

---

<sup>73</sup><https://repl.it>

<sup>74</sup><https://repl.it>

<sup>75</sup><https://www.codementor.io/garethdwyer/beginner-web-scraping-with-python-and-repl-it-nzr27jvnq>

<sup>76</sup><https://en.wikipedia.org/wiki/RSS>

[Not Secure | view-source:feeds.bbci.co.uk/news/world/rss.xml](https://feeds.bbci.co.uk/news/world/rss.xml)

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <rss xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:atom="http://www.w3.org/2005/Atom"
3 version="2.0" xmlns:media="http://search.yahoo.com/mrss/">
4   <channel>
5     <title><![CDATA[BBC News - World]]></title>
6     <description><![CDATA[BBC News - World]]></description>
7     <link>https://www.bbc.co.uk/news/</link>
8     <image>
9       <url>https://news.bbciimg.co.uk/nol/shared/img/bbc_news_120x60.gif</url>
10      <title>BBC News - World</title>
11      <link>https://www.bbc.co.uk/news/</link>
12    </image>
13    <generator>RSS for Node</generator>
14    <lastBuildDate>Wed, 06 Mar 2019 15:49:06 GMT</lastBuildDate>
15    <copyright><![CDATA[Copyright: (C) British Broadcasting Corporation, see http://news.bbc.co.uk/2/hi/help/rss/4498287.stm for terms and conditions of reuse.]]></copyright>
16    <language><![CDATA[en-gb]]></language>
17    <ttl>15</ttl>
18  <item>
19    <title><![CDATA[Trump dealt blow as US trade deficit jumps]]></title>
20    <description><![CDATA[The gap between US imports and exports hits a 10-year high despite the president's reduction plan.]]></description>
21    <link>https://www.bbc.co.uk/news/business-47472282</link>
22    <guid isPermaLink="true">https://www.bbc.co.uk/news/business-47472282</guid>
23    <pubDate>Wed, 06 Mar 2019 15:30:26 GMT</pubDate>
24    <media:thumbnail width="976" height="549" url="http://c.files.bbci.co.uk/3B93/production/_105915251_trump.jpg"/>
25  </item>
26  <item>
27    <title><![CDATA[Putin: Russia foiled work of almost 600 spies]]></title>
28    <description><![CDATA[The Russian president accuses foreign intelligence services of beefing up activities in Russia.]]></description>
29    <link>https://www.bbc.co.uk/news/world-europe-47467823</link>
30    <guid isPermaLink="true">https://www.bbc.co.uk/news/world-europe-47467823</guid>
31    <pubDate>Wed, 06 Mar 2019 15:40:54 GMT</pubDate>
32    <media:thumbnail width="976" height="549" url="http://c.files.bbci.co.uk/17233/production/_105917749_gettyimages-1128827252.jpg"/>
33  </item>
34  <item>
35    <title><![CDATA[South Africa 'sunflower murder' farmers jailed]]></title>
36    <description><![CDATA[Two white farmers accused a black teenager of stealing sunflowers worth $5 and threw him from a van.]]></description>
37    <link>https://www.bbc.co.uk/news/world-africa-47469522</link>
38    <guid isPermaLink="true">https://www.bbc.co.uk/news/world-africa-47469522</guid>
39    <pubDate>Wed, 06 Mar 2019 12:55:13 GMT</pubDate>
40    <media:thumbnail width="976" height="549" url="http://c.files.bbci.co.uk/8607/production/_105911343_mediaitem105911342.jpg"/>
41  </item>
42  <item>
```

If you click on the link above, you won't see the XML directly. Instead, it has some associated styling information so that most web browsers will display something that's a bit more human friendly. For example, opening the page in Google Chrome shows the page below. In order to view the raw XML directly, you can right-click on the page and click "view source".

[Not Secure | feeds.bbci.co.uk/news/world/rss.xml](https://feeds.bbci.co.uk/news/world/rss.xml)

**BBC NEWS**

**What is this page?**  
This is an RSS feed from the BBC News - World website. RSS feeds allow you to stay up to date with the latest news and features you want from BBC News - World.  
To subscribe to it, you will need a News Reader or other similar device. If you would like to use this feed to display BBC News - World content on your site, [please go here](#).  
? Help, I don't know what a news reader is and still don't know what this is about.

**RSS Feed For:** BBC News - World  
Below is the latest content available from this feed. [This isn't the feed I want.](#)

**Trump dealt blow as US trade deficit jumps**  
The gap between US imports and exports hits a 10-year high despite the president's reduction plan.

**Putin: Russia foiled work of almost 600 spies**  
The Russian president accuses foreign intelligence services of beefing up activities in Russia.

**South Africa 'sunflower murder' farmers jailed**  
Two white farmers accused a black teenager of stealing sunflowers worth \$5 and threw him from a van.

**Will Smith 'casting as Richard Williams' sparks colourism debate**  
Smith is reported to be playing the darker-skinned father of Serena and Venus Williams in a biopic.

**UN rights chief Bachelet warns of threat from 'gross inequality'**  
UN human rights chief Michelle Bachelet calls on nations to do more, and condemns xenophobia.

**Kenya airport strike strands thousands of passengers**  
A strike condemned as "criminal" by a government minister causes major disruption to flights.

**North Korea rebuilding Sohae rocket launch site, say observers**  
The North had begun to dismantle Sohae but satellite images indicate the launch pad is being restored.

**Goldman Sachs relaxes dress code for all employees**  
The firm urged staff to use "good judgement" in their fashion choices as it relaxed its dress standards.

**France Chiloé stabbing: Guards protest after jail 'terror attack'**  
An inmate stabbed two guards at a high-security jail, in an attack described as a "terror incident".

**Subscribe to this feed**

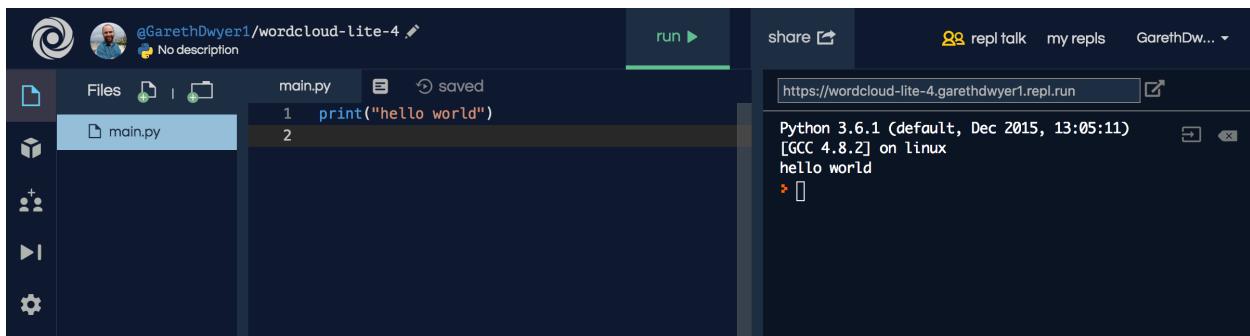
You can subscribe to this RSS feed in a number of ways, including the following:

- Drag the orange RSS button into your News Reader
- Drag the URL of the RSS feed into your News Reader
- Cut and paste the URL of the RSS feed into your News Reader

RSS feeds are used internally by software such as the news reader [Feedly<sup>77</sup>](#) and various email clients. We'll be consuming these RSS feeds with a Python library to retrieve the latest articles from BBC.

## Setting up our online environment (Repl.it)

In this tutorial, we'll be building our web application using [Repl.it<sup>78</sup>](#), which will allow us to have a consistent code editor, environment, and deployment framework in a single click. Head over there and create an account. Choose to create a Python Repl, and you should see an editor where you can write and run Python code, similar to the image below. You can write Python code in the middle pane, run it by pressing the green "run" button, and see the output in the right pane. In the left pane, you can see a list of files, with `main.py` added there by default.



## Pulling data from our feed and extracting URLs

In the [previous webscraping tutorial<sup>79</sup>](#) we used [BeautifulSoup<sup>80</sup>](#) to look for hyperlinks in a page and extract them. Now that we are using RSS, we can simply parse the feed as described above to find these same URLs. We will be using the Python [feedparser<sup>81</sup>](#) library to do this.

Let's start by simply printing out the URLs for all of the latest articles from BBC. Switch back to the `main.py` file in the Repl.it IDE and add the following code.

<sup>77</sup>[feedly.com](#)

<sup>78</sup>[repl.it](#)

<sup>79</sup><https://www.codementor.io/garethdwyer/beginner-web-scraping-with-python-and-repl-it-nzr27jvnq>

<sup>80</sup><https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

<sup>81</sup><https://pythonhosted.org/feedparser/>

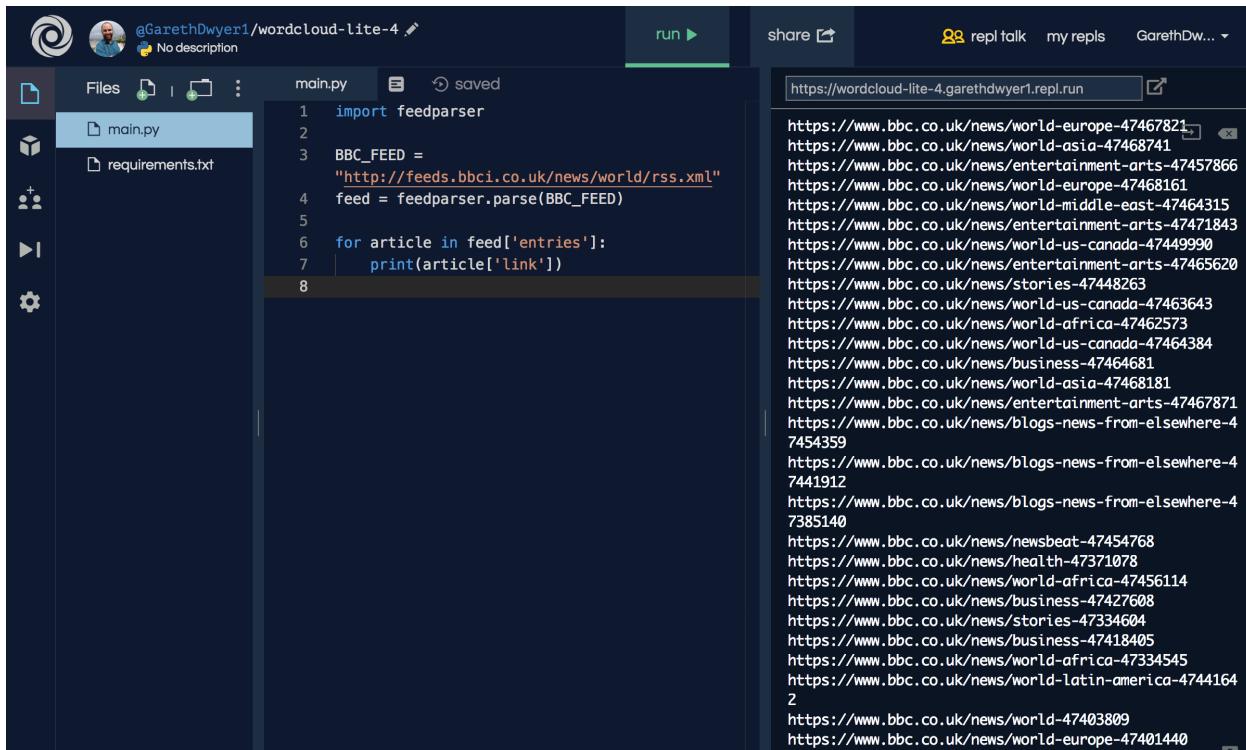
```

1 import feedparser
2
3 BBC_FEED = "http://feeds.bbci.co.uk/news/world/rss.xml"
4 feed = feedparser.parse(BBC_FEED)
5
6 for article in feed['entries']:
7     print(article['link'])

```

Feedparser does most of the heavy lifting for us, so we don't have to get too close to the slightly cumbersome XML format. In the code above, we parse the feed into a nice Python representation (line 4), loop through all of the entries (the `<item>` entries from the XML we looked at earlier), and print out the `link` elements.

If you run this code, you should see a few dozen URLs output on the right pane, as in the image below.



The screenshot shows a Repl.it workspace. On the left, there's a file tree with `main.py` and `requirements.txt`. The `main.py` file contains the code provided in the text block. On the right, the code is being run, and the output pane shows a list of URLs. The first URL is `https://wordcloud-lite-4.garethdwyer1.repl.run`, followed by many BBC news URLs such as `https://www.bbc.co.uk/news/world-europe-47467821`, `https://www.bbc.co.uk/news/world-asia-47468741`, and many others.

```

https://wordcloud-lite-4.garethdwyer1.repl.run
https://www.bbc.co.uk/news/world-europe-47467821
https://www.bbc.co.uk/news/world-asia-47468741
https://www.bbc.co.uk/news/entertainment-arts-47457866
https://www.bbc.co.uk/news/world-europe-47468161
https://www.bbc.co.uk/news/world-middle-east-47464315
https://www.bbc.co.uk/news/entertainment-arts-47471843
https://www.bbc.co.uk/news/world-us-canada-47449990
https://www.bbc.co.uk/news/entertainment-arts-47465620
https://www.bbc.co.uk/news/stories-47448263
https://www.bbc.co.uk/news/world-us-canada-47463643
https://www.bbc.co.uk/news/world-africa-47462573
https://www.bbc.co.uk/news/world-us-canada-47464384
https://www.bbc.co.uk/news/business-47464681
https://www.bbc.co.uk/news/world-asia-47468181
https://www.bbc.co.uk/news/world-europe-47467871
https://www.bbc.co.uk/news/entertainment-arts-47467871
https://www.bbc.co.uk/news/blogs-news-from-elsewhere-47454359
https://www.bbc.co.uk/news/blogs-news-from-elsewhere-47441912
https://www.bbc.co.uk/news/blogs-news-from-elsewhere-47385140
https://www.bbc.co.uk/news/newsbeat-47454768
https://www.bbc.co.uk/news/health-47371078
https://www.bbc.co.uk/news/world-africa-47456114
https://www.bbc.co.uk/news/business-47427608
https://www.bbc.co.uk/news/stories-47334604
https://www.bbc.co.uk/news/business-47418405
https://www.bbc.co.uk/news/world-africa-47334545
https://www.bbc.co.uk/news/world-latin-america-4744162
https://www.bbc.co.uk/news/world-47403809
https://www.bbc.co.uk/news/world-europe-47401440

```

## Setting up a web application with Flask

We don't just want to print this data out in the Repl console. Instead, our application should return information to anyone who uses a web browser to visit our application. We'll, therefore, install the lightweight web framework [Flask<sup>82</sup>](#) and use this to serve web content to our visitors.

<sup>82</sup><http://flask.pocoo.org/>

In the `main.py` file, we need to modify our code to look as follows:

```
1 import feedparser
2 from flask import Flask
3
4 app = Flask(__name__)
5
6 BBC_FEED = "http://feeds.bbci.co.uk/news/world/rss.xml"
7
8 @app.route("/")
9 def home():
10     feed = feedparser.parse(BBC_FEED)
11     urls = []
12
13     for article in feed['entries']:
14         urls.append(article['link'])
15
16     return str(urls)
17
18
19 if __name__ == '__main__':
20     app.run('0.0.0.0')
```

Here we still parse the feed and extract all of the latest article URLs, but instead of printing them out, we add them to a list (`urls`), and return them from a function. The interesting parts of this code are:

- Line 2: we import Flask
- Line 4: we initialise Flask to turn our project into a web application
- Line 8: we use a decorator to define the homepage of our application (an empty route, or `/`).
- Lines 19-20: We run Flask's built-in webserver to serve our content.

Press “run” again, and you should see a new window appear in the top right pane. Here we can see a basic web page (viewable already to anyone in the world by sharing the URL you see above it), and we see the same output that we previously printed to the console.

The screenshot shows a Repl.it workspace for a project named 'wordcloud-lite-4'. The left sidebar lists files: 'main.py' (selected), 'requirements.txt', and a 'saved' file. The main code editor contains the following Python code:

```

1 import feedparser
2 from flask import Flask
3
4 app = Flask(__name__)
5
6 BBC_FEED =
7     "http://feeds.bbci.co.uk/news/world/rss.xml"
8
9 @app.route("/")
10 def home():
11     feed = feedparser.parse(BBC_FEED)
12     urls = []
13
14     for article in feed['entries']:
15         urls.append(article['link'])
16
17     return str(urls)
18
19 if __name__ == '__main__':
20     app.run('0.0.0.0')

```

The right side shows the browser output with a long list of URLs from BBC news articles. Below the browser is a terminal window displaying the command-line output of the application's execution.

```

https://wordcloud-lite-4--garethdwyer1.repl.co
[https://www.bbc.co.uk/news/business-47472282,
 https://www.bbc.co.uk/news/world-europe-47467823,
 https://www.bbc.co.uk/news/world-africa-47469522,
 https://www.bbc.co.uk/news/world-us-canada-47461006,
 https://www.bbc.co.uk/news/world-us-canada-47468011,
 https://www.bbc.co.uk/news/entertainment-arts-47472507,
 https://www.bbc.co.uk/news/world-47468749,
 https://www.bbc.co.uk/news/world-africa-47467271,
 https://www.bbc.co.uk/news/world-asia-47465082,
 https://www.bbc.co.uk/news/world-europe-47467821,
 https://www.bbc.co.uk/news/world-asia-47468741,
 https://www.bbc.co.uk/news/entertainment-arts-47457866,
 https://www.bbc.co.uk/news/world-europe-47468161,
 https://www.bbc.co.uk/news/world-middle-east-47464315,
 https://www.bbc.co.uk/news/entertainment-arts-47471843,
 https://www.bbc.co.uk/news/world-us-canada-47449990,
 https://www.bbc.co.uk/news/world-us-canada-47465600]

Repl.it: package installation success
* Serving Flask app "main" (lazy loading)
* Environment: production
  WARNING: Do not use the development server in a production environment.
    Use a production WSGI server instead.
* Debug mode: off
* Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)
172.18.0.1 - - [06/Mar/2019 16:38:11] "GET / HTTP/1.1" 200 -

```

## Downloading articles and extracting the text

The URLs aren't that useful to us, as we eventually want to display a summary of the *content* of each URL. The actual text of each article isn't included in the RSS feed that we have (some RSS feeds contain the full text of each article), so we'll need to do some more work to download each article. First, we'll add the third-party libraries `requests` and `BeautifulSoup` as dependencies, again just using the “magic import”. We'll be using these to download the content of each article from the URL and strip out extra CSS and JavaScript to leave us with plain text.

Now we're ready to download the content from each article and serve that up to the user. Modify the code in `main.py` to look as follows.

```

1 import feedparser
2 import requests
3
4 from flask import Flask
5 from bs4 import BeautifulSoup
6
7 app = Flask(__name__)
8
9 BBC_FEED = "http://feeds.bbci.co.uk/news/world/rss.xml"
10 LIMIT = 2

```

```
11
12 def parse_article(article_url):
13     print("Downloading {}".format(article_url))
14     r = requests.get(article_url)
15     soup = BeautifulSoup(r.text, "html.parser")
16     ps = soup.find_all('p')
17     text = "\n".join(p.get_text() for p in ps)
18     return text
19
20 @app.route("/")
21 def home():
22     feed = feedparser.parse(BBC_FEED)
23     article_texts = []
24
25     for article in feed['entries'][:LIMIT]:
26         text = parse_article(article['link'])
27         article_texts.append(text)
28     return str(article_texts)
29
30 if __name__ == '__main__':
31     app.run('0.0.0.0')
```

Let's take a closer look at what has changed.

- We import our new libraries on **lines 2 and 5**.
- We create a new global variable `LIMIT` on **line 10** to limit how many articles we want to download.
- **Lines 12-18** define a new function that takes a URL, downloads the article, and extracts the text. It does this using a crude algorithm that assumes anything inside HTML `<p>` (paragraph) tags is interesting content.
- We modify **lines 23, 25, 26, and 27** so that we use the new `parse_article` function to get the actual content of the URLs that we found in the RSS feed and return that to the user instead of returning the URL directly. Note that we limit this to two articles by truncating our list to `LIMIT` for now, as the downloads take a while and Repl's resources on free accounts are limited.

If you run the code now, you should see output similar to that shown in the image below (you may need to hit refresh in the right pane). You can see text from the first article in the top-right pane now, and the text for the second article is further down the page. You'll notice that our text extraction algorithm isn't perfect and there's still some extra text about "Share this" at the top that isn't actually part of the article, but this is good enough for us to create word clouds from later.

The screenshot shows a Repl.it workspace. On the left, there's a file tree with 'main.py' and 'requirements.txt'. The main area displays the code for 'main.py':

```

1 import feedparser
2 import requests
3
4 from flask import Flask
5 from bs4 import BeautifulSoup
6
7 app = Flask(__name__)
8
9 BBC_FEED =
10 "http://feeds.bbci.co.uk/news/world/rss.xml"
11 LIMIT = 2
12
13 def parse_article(article_url):
14     print("Downloading {}".format(article_url))
15     r = requests.get(article_url)
16     soup = BeautifulSoup(r.text,
17                           "html.parser")
18     ps = soup.find_all('p')
19     text = "\n".join(p.get_text() for p in ps)
20
21     return text
22
23 @app.route("/")
24 def home():
25     feed = feedparser.parse(BBC_FEED)
26     article_texts = []
27
28     for article in feed['entries'][:LIMIT]:
29         text = parse_article(article['link'])
30         article_texts.append(text)
31
32     return str(article_texts)

```

To the right, there's a terminal window showing the output of running the application. It includes a warning about using a development server in production:

```

[GCC 4.8.2] on linux
* Serving Flask app "main" (lazy loading)
* Environment: production
  WARNING: Do not use the development server in a production environment.
  Use a production WSGI server instead.
* Debug mode: off
* Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)

```

Below the warning, it shows the application fetching news articles from BBC feeds:

```

Downloading https://www.bbc.co.uk/news/business-47472282
Downloading https://www.bbc.co.uk/news/world-europe-47467823
172.18.0.1 - - [06/Mar/2019 16:54:43] "GET / HTTP/1.1" 200 -

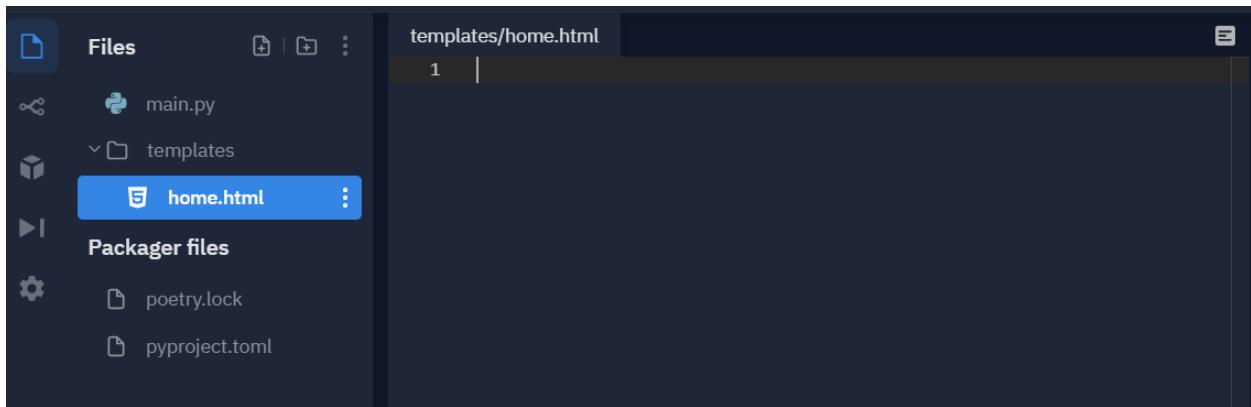
```

## Returning HTML instead of plain text to the user

Although Flask allows us to return Python `str` objects directly to our visitors, the raw result is ugly compared to how people are used to seeing web pages. To take advantage of HTML formatting and CSS styling, it's better to define HTML *templates*, and use Flask's template engine, `jinja`, to inject dynamic content into these. Before we get to creating image files from our text content, let's set up a basic Flask template.

To use Flask's templates, we need to set up a specific file structure. Press the “new folder” button (next to the “new file” button, on the left pane), and name the resulting new folder `templates`. This is a special name recognised by Flask, so make sure you get the spelling exactly correct.

Select the new folder and press the “new file” button to create a new file inside our `templates` folder. Call the file `home.html`. Note below how the `home.html` file is indented one level, showing that it is inside the folder. If yours is not, drag and drop it into the `templates` folder so that Flask can find it.



In the `home.html` file, add the following code, which is a mix between standard HTML and Jinja's templating syntax to mix dynamic content into the HTML.

```

1 <html>
2   <body>
3     <h1>News Word Clouds</h1>
4     <p>Too busy to click on each news article to see what it's about? Below you \
5 can see all the articles from the BBC front page, displayed as word clouds. If you w\
6 ant to read more about any particular article, just click on the wordcloud to go to \
7 the original article</p>
8     {% for article in articles %}
9       <p>{{article}}</p>
10      {% endfor %}
11    </body>
12 </html>
```

Jinja uses the special characters `{%` and `{} (in opening and closing pairs)` to show where dynamic content (e.g. variables calculated in our Python code) should be added and to define control structures. Here we loop through a list of `articles` and display each one in a set of `<p>` tags.

We'll also need to tweak our Python code a bit to account for the template. In the `main.py` file, make the following changes.

- Add a new import near the top of the file, below the existing Flask import

```
1 from flask import render_template
```

- Update the last line of the `home()` function to make a call to `render_template` instead of returning a `str` directly as follows.

```

1 @app.route("/")
2 def home():
3     feed = feedparser.parse(BBC_FEED)
4     article_texts = []
5
6     for article in feed['entries'][:LIMIT]:
7         text = parse_article(article['link'])
8         article_texts.append(text)
9     return render_template('home.html', articles=article_texts)

```

The `render_template` call tells Flask to prepare some HTML to return to the user by combining data from our Python code and the content in our `home.html` template. Here we pass `article_texts` to the renderer as `articles`, which matches the `articles` variable we loop through in `home.html`.

If everything went well, you should see different output now, which contains our header from the HTML and static first paragraph, followed by two paragraphs showing the same article content that we pulled before. If you don't see the updated webpage, you may need to hit refresh in the right pane again.

The screenshot shows the Repl.it interface. On the left, the code editor displays `main.py` with the Python script provided above. On the right, the browser window shows the resulting web page titled "News Word Clouds". The page content includes a header and two paragraphs of BBC news text. Below the browser window, a terminal window shows the command-line output of the application's requests to the BBC news API.

```

* Debug mode: off
* Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)
downloading https://www.bbc.co.uk/news/business-4747228
2
downloading https://www.bbc.co.uk/news/world-europe-47467823
172.18.0.1 - - [06/Mar/2019 17:15:39] "GET / HTTP/1.1" 200 -
downloading https://www.bbc.co.uk/news/business-4747228
2
downloading https://www.bbc.co.uk/news/world-europe-47467823
172.18.0.1 - - [06/Mar/2019 17:15:40] "GET / HTTP/1.1" 200 -
172.18.0.1 - - [06/Mar/2019 17:15:41] "GET /favicon.ico HTTP/1.1" 404 -

```

Now it's time to move on to generating the actual word clouds.

## Generating word clouds from text in Python

Once again, there's a nifty Python library that can help us. This one will take in text and return word clouds as images and is called `wordcloud`.

Images are usually served as files living on your server or from an image host like [imgur<sup>83</sup>](#). Because we'll be creating small, short-lived images dynamically from the text, we'll simply keep them in memory instead of saving them anywhere permanently. In order to do this, we'll have to mess around a bit with the Python `io` and `base64` libraries, alongside our newly installed `wordcloud` library.

To import all the new libraries, we'll be using to process images, modify the top of our `main.py` to look as follows.

```

1 import base64
2 import feedparser
3 import io
4 import requests
5
6 from bs4 import BeautifulSoup
7 from wordcloud import WordCloud
8 from flask import Flask
9 from flask import render_template

```

We'll be converting the text from each article into a separate word cloud, so it'll be useful to have another helper function that can take text as input and produce the word cloud as output. We can use `base64`<sup>84</sup> to represent the images, which can then be displayed directly in our visitors' web browsers.

Add the following function to the `main.py` file.

```

1 def get_wordcloud(text):
2     pil_img = WordCloud().generate(text=text).to_image()
3     img = io.BytesIO()
4     pil_img.save(img, "PNG")
5     img.seek(0)
6     img_b64 = base64.b64encode(img.getvalue()).decode()
7     return img_b64

```

This is probably the hardest part of our project in terms of readability. Normally, we'd generate the word cloud using the `wordcloud` library and then save the resulting image to a file. However, because we don't want to use our file system here, we'll create a `BytesIO` Python object in memory instead

---

<sup>83</sup>[imgur.com](#)

<sup>84</sup><https://en.wikipedia.org/wiki/Base64>

and save the image directly to that. We'll convert the resulting bytes to base64 in order to finally return them as part of our HTML response and show the image to our visitors.

In order to use this function, we'll have to make some small tweaks to the rest of our code.

For our template, in the `home.html` file, change the for loop to read as follows.

```
1  {% for article in articles %}  
2        
3  {% endfor %}
```

Now instead of displaying our article in `<p>` tags, we'll put it inside an `<img/>` tag so that it can be displayed as an image. We also specify that it is formatted as a png and encoded as base64.

The last thing we need to do is modify our `home()` function to call the new `get_wordcloud()` function and to build and render an array of images instead of an array of text. Change the `home()` function to look as follows.

```
1  @app.route("/")  
2  def home():  
3      feed = feedparser.parse(BBC_FEED)  
4      clouds = []  
5  
6      for article in feed['entries'][:LIMIT]:  
7          text = parse_article(article['link'])  
8          cloud = get_wordcloud(text)  
9          clouds.append(cloud)  
10     return render_template('home.html', articles=clouds)
```

We made changes on lines 4, 8, 9, and 10, to change to a `clouds` array, populate that with images from our `get_wordcloud()` function, and return that in our `render_template` call.

If you restart the Repl and refresh the page, you should see something similar to the following. We can see the same content from the articles, however, we can now see the important keywords without having to read the entire article.

The screenshot shows the Repl.it interface. On the left, the file structure includes `main.py`, `templates` (containing `home.html`), and `requirements.txt`. The `main.py` file contains Python code for a Flask application that fetches BBC news articles and generates word clouds from their content. The `home.html` template is shown below:

```

<html>
<head>
    <title>News Word Clouds</title>
</head>
<body>
    <h1>News Word Clouds</h1>
    <p>Too busy to click on each news article to see what it's about? Below you can see all the articles from the BBC front page, displayed as word clouds. If you want to read more about any particular article, just click on the wordcloud to go to the original article</p>
    <div>
        <img alt="Word cloud visualization of BBC news articles" data-bbox="628 238 861 308"/>
    </div>
</body>
</html>

```

The right side of the interface shows the browser output with the generated word cloud and the Python application logs.

For a larger view, you can pop out the website in a new browser tab using the button in the top right of the Repl editor (indicated in red above).

The last thing we need to do is add some styling to make the page look a bit prettier and link the images to the original articles.

## Adding some finishing touches

Our text looks a bit stark, and our images touch each other, which makes it hard to see that they are separate images. We'll fix that up by adding a few lines of CSS and importing the Bootstrap framework.

### Adding CSS

Edit the `home.html` file to look as follows

```

1 <html>
2   <head>
3     <title>News in WordClouds | Home</title>
4     <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/3.4.1/\n
5       css/bootstrap.min.css" integrity="sha384-HSMxcRTRxN+N+Bdg0JdbxYKrThec0KuH5zCYot1SAcP1 \
6       +c8xmyTe9GYg119a69psu" crossorigin="anonymous">
7
8     <style type="text/css">
9       body {padding: 20px;}
10      img{padding: 5px;}
11    </style>
12  </head>
13
14  <body>
15    <h1>News Word Clouds</h1>
16    <p>Too busy to click on each news article to see what it's about? Below you can see all the articles from the BBC front page, displayed as word clouds. If you want to read more about any particular article, just click on the wordcloud to go to the original article</p>
17    {% for article in articles %}
18      <a href="{{article.url}}>
19    </a>
20    {% endfor %}
21  </body>
22 </html>

```

On **line 3** we add a title, which is displayed in the browser tab. On **line 4**, we import **Bootstrap**<sup>85</sup>, which has some nice CSS defaults right out the box (it's probably a bit heavy-weight for our project as we have so little content and won't use most of Bootstrap's features, but it's nice to have if you're planning on extending the project.)

On **lines 6-8**, we add padding to the main body to stop the text going to close to the edges of the screen, and also add padding to our images to stop them touching each other.

On **line 16**, we use an `<a>` tag to add a link to our image. We also change the Jinja templates to `{{article.url}}` and `{{article.image}}` so that we can have images that link back to the original news article.

Now we need to tweak our backend code again to pass through the URL and image for each article, as the template currently doesn't have access to the URL.

---

<sup>85</sup><https://getbootstrap.com/>

## Passing through the URLs

To easily keep track of pairs of URLs and images, we'll add a basic Python helper class called `Article`. In the `main.py` file, add the following code before the function definitions.

```
1 class Article:  
2     def __init__(self, url, image):  
3         self.url = url  
4         self.image = image
```

This is a simple class with two attributes: `url` and `image`. We'll store the original URL from the RSS feed in `url` and the final base64 wordcloud in `image`.

To use this class, modify the `home()` function to look as follows.

```
1 @app.route("/")  
2 def home():  
3     feed = feedparser.parse(BBC_FEED)  
4     articles = []  
5  
6     for article in feed['entries'][:LIMIT]:  
7         text = parse_article(article['link'])  
8         cloud = get_wordcloud(text)  
9         articles.append(Article(article['link'], cloud))  
10    return render_template('home.html', articles=articles)
```

We changed the name of our `clouds` list to `articles`, and populated it by initialising `Article` objects in the for loop and appending them to this list. We then pass across `articles=articles` instead of `articles=clouds` in the return statement so that the template can access our list of `Article` objects, which each contain the image and the URL of each article.

If you refresh the page again and expand the window using the pop out button, you'll be able to click any of the images to go to the original article, allowing readers to view a brief summary of the day's news or to read more details about any stories that catch their eye.

## Where next?

We've included several features in our web application, and looked at how to use RSS feeds and process and serve images directly in Python, but there are a lot more features we could add. For example:

- Our application only shows two stories at a time as the download time is slow. We could instead look at implementing a threaded solution to downloading web pages so that we could process several articles in parallel. Alternatively (or in addition), we could also download the articles on a schedule and cache the resulting images so that we don't have to do the resource heavy downloading and parsing each time a visitor visits our site.
- Our web application only shows articles from a single source (BBC), and only from today. We could add some more functionality to show articles from different sources and different time frames. We could also consider allowing the viewer to choose which category of articles to view (news, sport, politics, finance, etc) by using different RSS feeds as sources.
- Our design and layout is very basic. We could make our site look better and be more responsive by adding more CSS. We could lay out the images in a grid of rows and columns to make it look better on smaller screens such as mobile phones.

If you'd like to keep working on the web application, simply head over to [the Repl<sup>86</sup>](#) and fork it to continue your own version.

In the next chapter, we'll be looking at how to build our own Discord Chatbot.

---

<sup>86</sup><https://repl.it/@GarethDwyer1/news-to-wordcloud>

# Building a Discord Bot with Python and Repl.it

In this tutorial, we'll use [Repl.it](#)<sup>87</sup> and Python to build a Discord Chatbot. If you're reading this tutorial, you probably have at least heard of Discord and likely have an existing account. If not, Discord is a VoIP and Chat application that is designed to replace Skype for gamers. The bot we create in this tutorial will be able to join a Discord server and respond to messages sent by people.

If you prefer JavaScript, the next chapter is the same tutorial using NodeJS instead of Python.

You'll find it easier to follow along if you have some Python knowledge and have used Discord or a similar app such as Skype or Telegram before. We won't be covering the very basics of Python, but we will explain each line of code in detail, so if you have any experience with programming, you should be able to follow along.

## Overview and requirements

We'll be doing all of our coding through the Repl.it web IDE and hosting our bot with Repl.it as well, so you won't need to install any additional software on your machine. For this tutorial you will need to create a [Discord](#)<sup>88</sup> account (if you already have one, you can skip this). There are instructions for how to do this in the next section.

In this tutorial, we will be covering:

- Creating an application and a bot user in your Discord account
- Creating a server on Discord
- Adding our bot to our Discord server

Let's get through these admin steps first and then we can get to the fun part of coding our bot.

## Creating a bot in Discord and getting a token

You can sign up for a free account over at [the Discord register page](#)<sup>89</sup>, and can download one of their desktop or mobile applications from [the Discord homepage](#)<sup>90</sup>. You can also use Discord in the browser.

Once you have an account, you'll want to create a Discord application. Visit [the Discord developer's page](#)<sup>91</sup> and press the "New application" button, as in the image below.

---

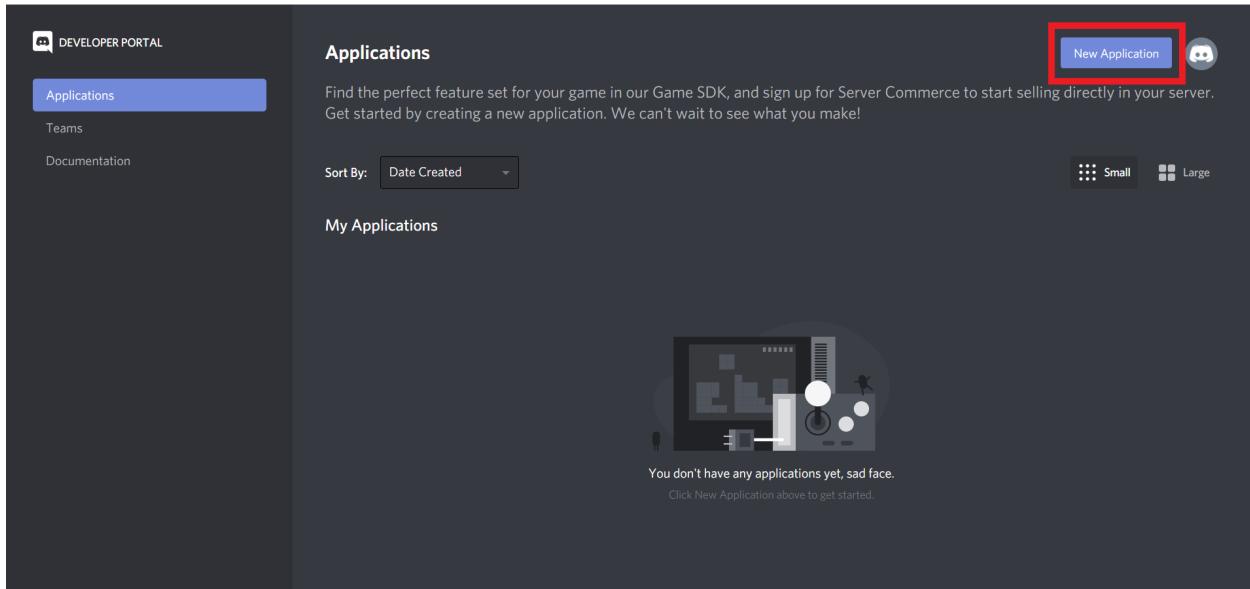
<sup>87</sup><https://repl.it>

<sup>88</sup><https://discordapp.com/>

<sup>89</sup><https://discordapp.com/register>

<sup>90</sup><https://discordapp.com/>

<sup>91</sup><https://discordapp.com/developers/applications/>



### Creating a new Discord application

Fill out a name for your bot and select “Create”.

The first thing to do on the next page is to note your Client ID, which you’ll need to add the bot to the server. You can come back later and get it from this page, or copy it somewhere where you can easily find it later.

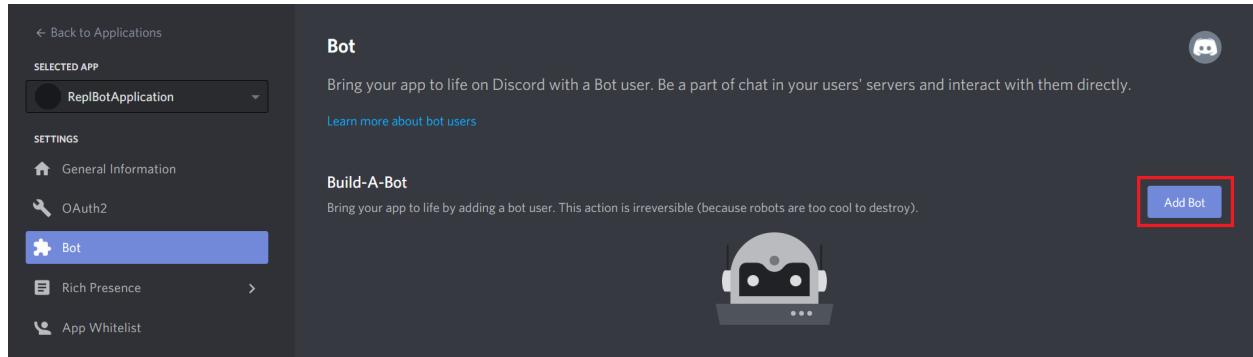
The screenshot shows the 'General Information' page for the 'ReplBotApplication'. The left sidebar shows 'General Information' (selected), 'OAuth2', 'Bot', 'Rich Presence', and 'App Whitelist'. The main area has fields for 'NAME' (ReplBotApplication), 'CLIENT ID' (746269162917331028), 'DESCRIPTION (MAXIMUM 400 CHARACTERS)' (A simple Discord bot), and 'URL EXAMPLE' (ReplBotApplication). A red box highlights the 'CLIENT ID' field. At the bottom, there's a note 'Careful — you have unsaved changes!' and a 'Save Changes' button, which is also highlighted with a red box.

### Record your Client ID

You can also rename the application and provide a description for your bot at this point and press “Save Changes”.

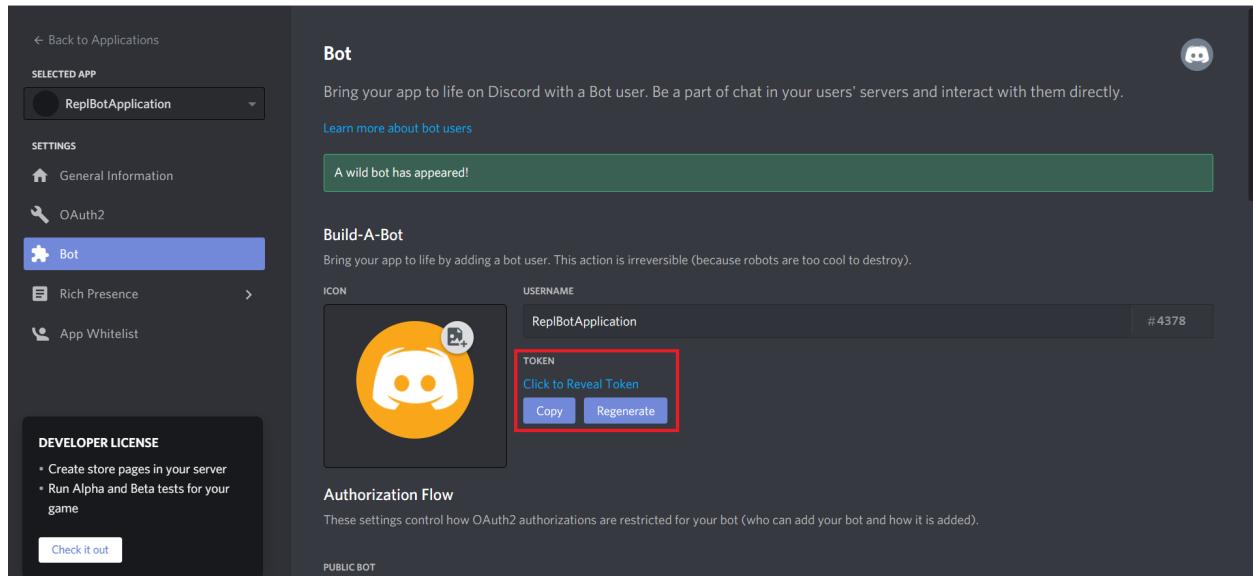
You have now created a Discord application. The next step is to add a bot to this application, so head

over to the “Bot” tab using the menu on the left and press the “Add Bot” button, as indicated below. Click “Yes, do it” when Discord asks if you’re sure about bringing a new bot to life.



Adding a bot to our Discord Application

The last thing we’ll need from our bot is a Token. Anyone who has the bot’s token can prove that they own the bot, so you’ll need to be careful not to share this with anyone. You can get the token by pressing “Click to Reveal Token”, or copy it to your clipboard without seeing it by pressing “Copy”.

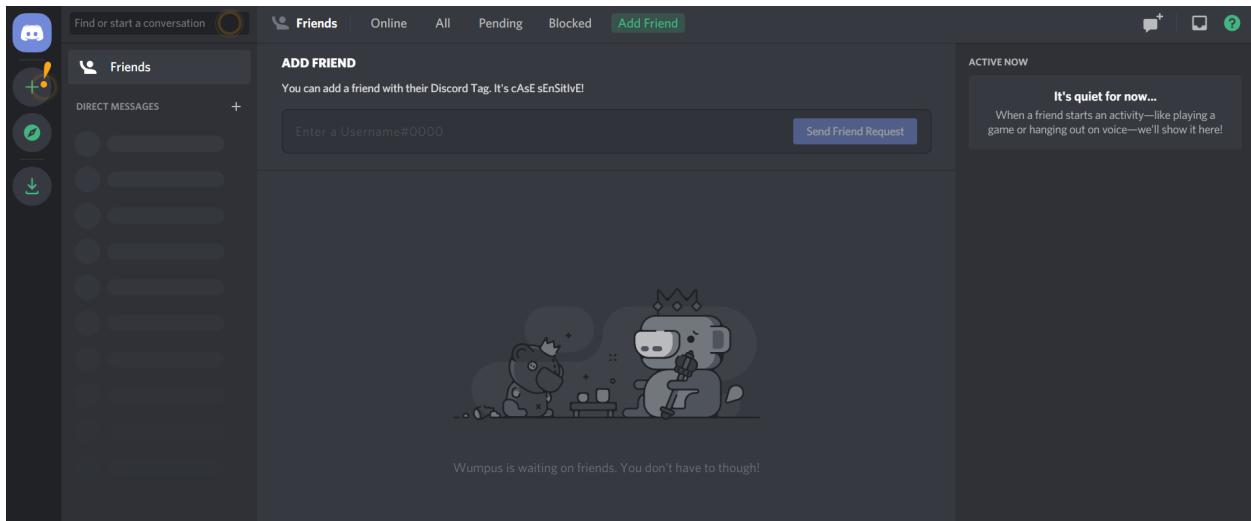


Generating a token for our Discord bot

Take note of your token or copy it to your clipboard, as we’ll need to add it to our code soon.

## Creating a Discord server

If you don’t have a Discord server to add your bot to, you can create one by either opening the desktop Discord application that you downloaded earlier or returning to the Discord home page in your browser. Press the “+” icon indicated by the exclamation mark, as shown below, to create a server.



Creating a Discord server

Press “Create a server” in the screen that follows, and then give your server a name. Once the server is up and running, you can chat with yourself, or invite some friends to chat with you. Soon we’ll invite our bot to chat with us as well.

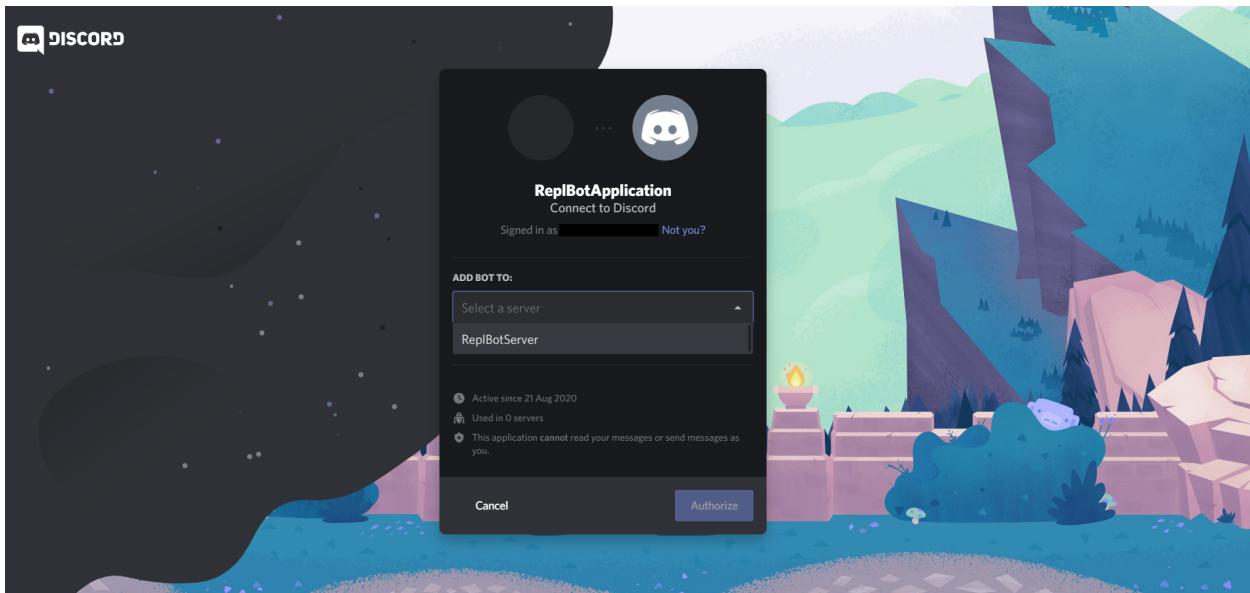
## Adding your Discord bot to your Discord server

Our Discord bot is still just a shell at this stage as we haven’t written any code to allow him to do anything, but let’s go ahead and add him to our Discord server anyway. To add a bot to your server, you’ll need the Client ID from the “General Information” page that we looked at before when we created our ReplBotApplication (ie. the client ID, not the secret bot Token).

Create a URL that looks as follows, but using your Client ID instead of mine at the end:

[https://discordapp.com/api/oauth2/authorize?scope=bot&client\\_id=746269162917331028](https://discordapp.com/api/oauth2/authorize?scope=bot&client_id=746269162917331028)

Visit the URL that you created in your web browser and you’ll see a page similar to the following where you can choose which server to add your bot to.



Authorizing our bot to join our server

Select the server we created in the step before this and hit the “authorize” button. After completing the captcha, you should get an in-app Discord notification telling you that your bot has joined your server.

Now we can get to the fun part of building a brain for our bot!

## Creating a Repl and installing our Discord dependencies

The first thing we need to do is create a Python Repl to write the code for our Discord bot. Over at [repl.it](https://repl.it)<sup>92</sup>, create a new Repl, choosing “Python” as your language.

We don’t need to reinvent the wheel, as there is already a great Python wrapper for the Discord bot API over on [GitHub](https://github.com/Rapptz/discord.py)<sup>93</sup>, which makes it a lot faster to get set up with a basic Python discord bot. To use library, we can simply write `import discord` at the top of `main.py`. Repl.it will handle installing this dependency when you press the “run” button.

Our bot is nearly ready to go – but we still need to plug in our secret token. This will authorize our code to control our bot.

## Setting up authorization for our bot

By default, Repl.it code is public. This is great as it encourages collaboration and learning, but we need to be careful not to share our secret bot token (which gives anyone who has access to it full

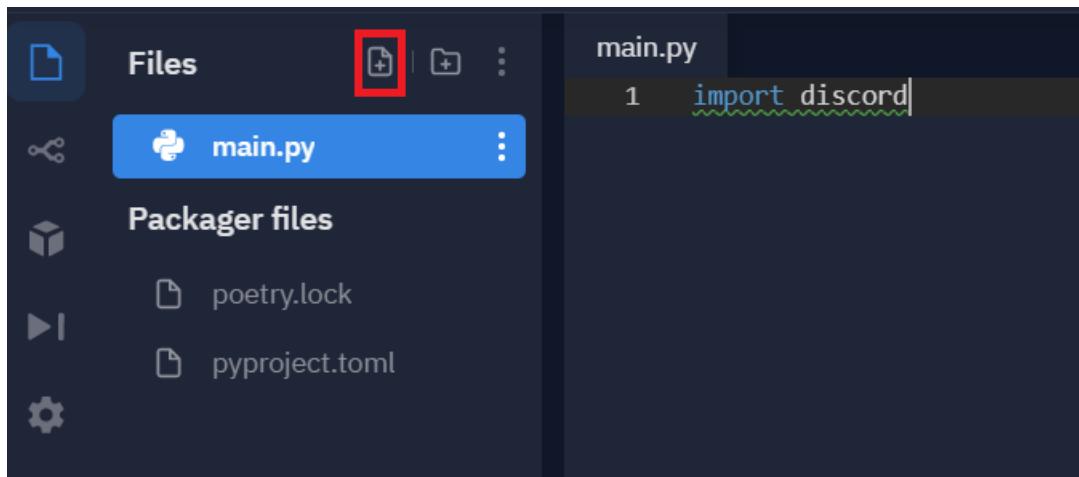
<sup>92</sup><https://repl.it>

<sup>93</sup><https://github.com/Rapptz/discord.py>

control of our bot).

To get around the problem of needing to give our *code* access to the token while allowing others to access our code but *not* our token, we'll be using [environment variables<sup>94</sup>](#). On a normal machine, we'd set these directly on our operating system, but using Repl.it we don't have access to this. Repl.it allows us to set secret environment variables through a special `.env` file.

First, we need to create a new file called exactly `.env`. Select "Add file" in the left pane, as shown in the image below, and name this file `.env`. It is important not to leave out the `.` at the beginning.



Create a new file called `.env`

Open this new file and add a variable to define your bot's secret token (note that this is the second token that we got while setting up the bot – different from the Client ID that we used to add our bot to our server). It should look something like:

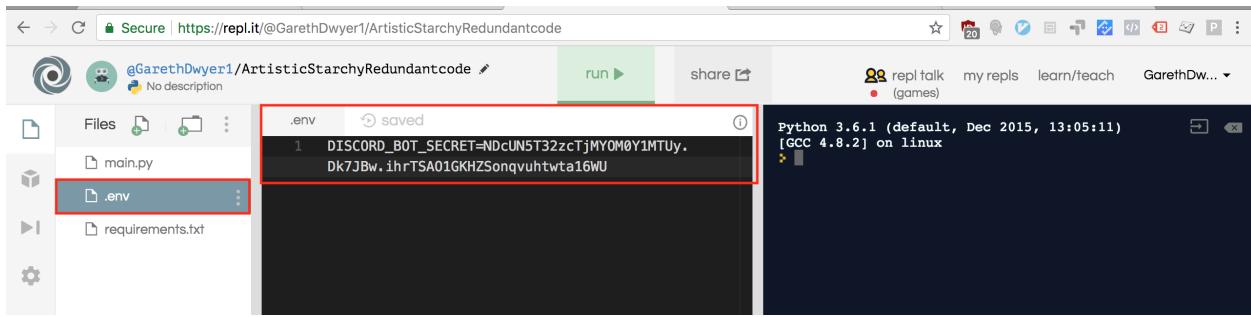
```
1 DISCORD_BOT_SECRET=NDcUN5T32zcTjMYOM0Y1MTUy.Dk7JBw.ihrTSAO1GKHZSonqvuhTwta16WU
```

You'll need to:

- Replace the token (after the `=` sign) with the token that Discord gave you when creating your own bot.
- Be careful about **spacing**. Unlike in Python, if you put a space on either side of the `=` in your `.env` file, these spaces will be part of the variable name or the value, so make sure you don't have any spaces around the `=` or at the end of the line.
- Run the code again. Sometimes you'll need to refresh the whole page to make sure that your environment variables are successfully loaded.

---

<sup>94</sup><https://www.digitalocean.com/community/tutorials/how-to-read-and-set-environmental-and-shell-variables-on-a-linux-vps>



Creating our .env file

Let's make a Discord bot that repeats everything we say but in reverse. We can do this in only a few lines of code. In your `main.py` file, add the following:

```
1 import discord
2 import os
3
4 client = discord.Client()
5
6 @client.event
7 async def on_ready():
8     print("I'm in")
9     print(client.user)
10
11 @client.event
12 async def on_message(message):
13     if message.author != client.user:
14         await client.send_message(message.channel, message.content[::-1])
15
16 token = os.environ.get("DISCORD_BOT_SECRET")
17 client.run(token)
```

Let's tear this apart line by line to see what it does.

- Lines 1-2 import the discord library that we installed earlier and the built-in operating system library, which we'll need to access our bot's secret token.
- In line 4, we create a Discord Client. This is a Python object that we'll use to send various commands to Discord's servers.
- In line 6, we say we are defining an event for our client. This line is a Python decorator, which will take the function directly below it and modify it in some way. The Discord bot is going to run *asynchronously*, which might be a bit confusing if you're used to running standard Python. We won't go into asynchronous Python in depth here, but if you're interested in what this is and why it's used, there's a good guide over at [FreeCodeCamp<sup>95</sup>](#). In short, instead of running

<sup>95</sup><https://medium.freecodecamp.org/a-guide-to-asynchronous-programming-in-python-with-asyncio-232e2afa44f6>

the code in our file from top to bottom, we'll be running pieces of code in response to specific events.

- In **lines 7-9** we define what kind of event we want to respond to, and what the response should be. In this case, we're saying that in response to the `on_ready` event (when our bot joins a server successfully), we should output some information server-side (i.e. this will be displayed in our Repl's output, but not sent as a message through to Discord). We'll print a simple `I'm in` message to see that the bot is there and print our bot's user id (if you're running multiple bots, this will make it easier to work out who's doing what).
- **Lines 11-14** are similar, but instead of responding to an `on_ready` event, we tell our bot how to handle new messages. **Line 13** says we only want to respond to messages that aren't from us (otherwise our bot will keep responding to himself – you can remove this line to see why that's a problem), and **line 15** says we'll send a new message to the same channel where we received a message (`message.channel`) and the content we'll send will be the same message that we received, but backwards (`message.content[::-1]` - `::-1` is a slightly odd but useful Python idiom to reverse a string or list).

The last two lines get our secret token from the environment variables that we set up earlier and then tell our bot to start up.

Press the big green “Run” button again and you should see your bot reporting a successful channel join in the Repl output.

The screenshot shows a Repl.it workspace for a project named "discord-bot". The left sidebar shows files: main.py, .env, and requirements.txt. The main area displays the code for main.py:

```

1 import discord
2 import os
3
4 client = discord.Client()
5
6 @client.event
7 async def on_ready():
8     print("I'm in")
9     print(client.user)
10
11 @client.event
12 async def on_message(message):
13     if message.author != client.user:
14         await client.send_message(
15             message.channel, message.content[::-1])
16
17 token = os.environ.get("DISCORD_BOT_SECRET")
18 client.run(token)

```

A green "restart" button is visible above the code editor. To the right, the browser window shows the Repl.it URL https://discord-bot--garethdwyer1.repl.co. The page displays the output of the bot's execution:

```

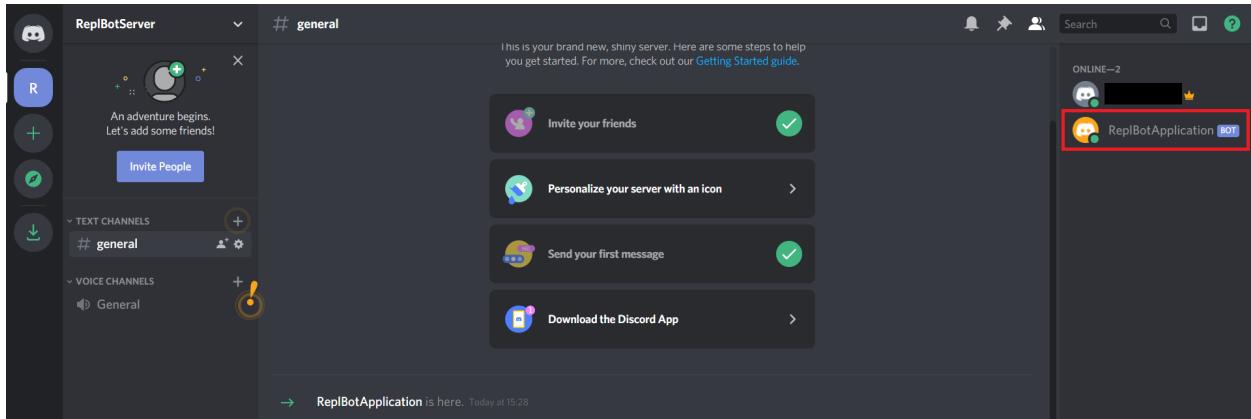
Downloading https://files.pythonhosted.org/packages/97/3c/0.16.12.tar.gz (414kB) 97b47e
Collecting aiohttp (from -r requirements.txt (line 2))
Downloaded https://files.pythonhosted.org/packages/54/f9/3.3.2-cp36-cp36-manylinux1_x86_64.whl (876kB) 47199c
Collecting websockets (from -r requirements.txt (line 3))
Downloaded https://files.pythonhosted.org/packages/5c/fe/6.0-cp36-cp36-manylinux1_x86_64.whl (88kB) eaef97
Collecting yarl<2.0,>=1.0 (from aiohttp->r requirements.txt (line 2))
Downloaded https://files.pythonhosted.org/packages/61/67/1.2.6-cp36-cp36-manylinux1_x86_64.whl 71b367
/home/rumer/.cache/pip/wheel/057007b3732d618e19e08a739751dd6bf998cfed277728f8b2127ad4eb7
Successfully built discord.py idna-ssl
Installing collected packages: multidict, idna, yarl, chardet, attrs, idna-ssl, async-timeout, aiohttp, websockets, discord.py
Successfully installed aiohttp-3.3.2 async-timeout-3.0.0 attrs-18.1.0 chardet-3.0.4 discord.py-0.16.12 idna-2.7 idna-ssl-1.1.0 multidict-4.3.1 websockets-6.0 yarl-1.2.6
You are using pip version 9.0.1, however version 18.0 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.

Repl.it: package installation success
I'm in
My first bot#4787

```

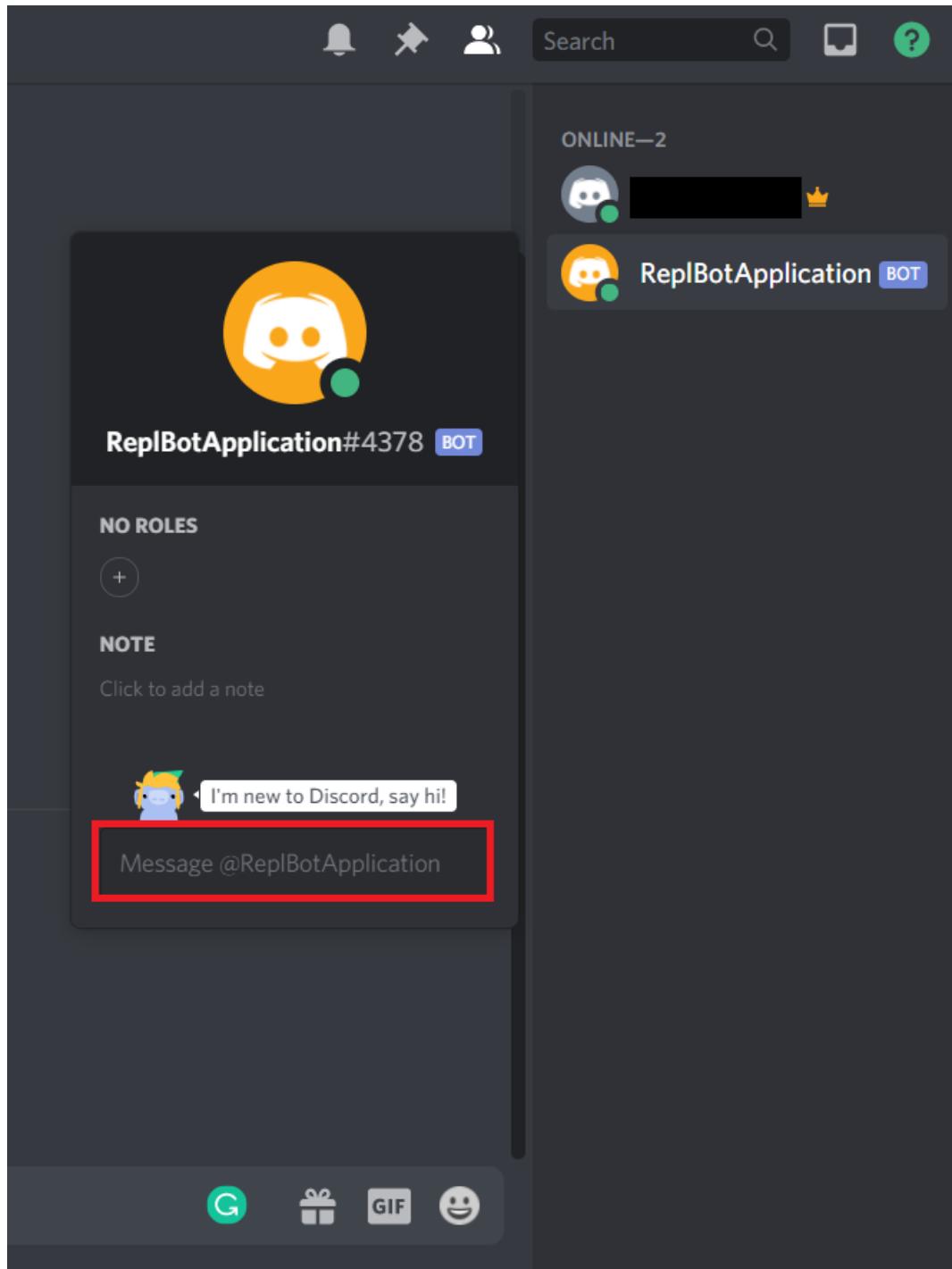
Seeing our bot join our server

Open Discord, and from within the server we created earlier, select your ReplBotApplication from the pane on the right-hand side of the screen.



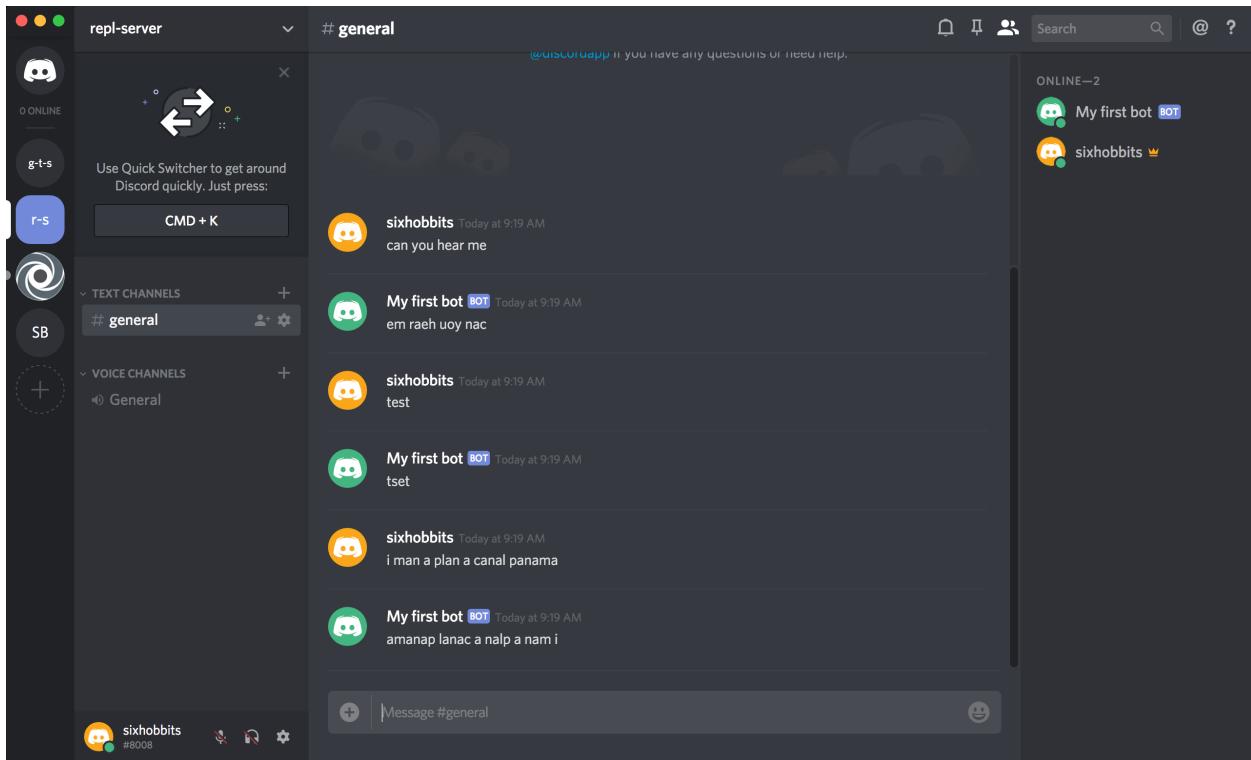
The Repl bot is active

Once you have selected this, you will be able to send a message (by typing into the box highlighted below) and see your bot respond!



Send a message to your bot

The bot responds each time, reversing the text we enter.



Our bot can talk!

## Keeping our bot alive

Your bot can now respond to messages, but only for as long as your Repl is running. If you close your browser tab or shut down your computer, your bot will stop and no longer respond to messages on Discord.

Repl.it will keep your code running after you close the browser tab only if you are running a web server. Because we are using the Python discord.py library, our bot doesn't require an explicit web server, but we can create a server and run it in a separate thread just to keep our Repl alive. We'll do this using the [Flask<sup>96</sup>](#) framework.

Create a new file in your project called `keep_alive.py` and add the following code:

<sup>96</sup><http://flask.pocoo.org/>

```
1 from flask import Flask
2 from threading import Thread
3
4 app = Flask('')
5
6 @app.route('/')
7 def home():
8     return "I'm alive"
9
10 def run():
11     app.run(host='0.0.0.0', port=8080)
12
13 def keep_alive():
14     t = Thread(target=run)
15     t.start()
```

We won't go over this in detail as it's not central to our bot, but here we start a web server that will return "I'm alive" if anyone visits it, and we'll provide a method to start this in a new thread (leaving the main thread for our Repl bot).

In our `main.py` file, we need to add an import for this server at the top. Add the following line near the top of `main.py`.

```
1 from keep_alive import keep_alive
```

In `main.py` we need to start up the web server just before you start up the bot. Add these three lines to `main.py`, just before the line with `token = os.environ.get("DISCORD_BOT_SECRET")`:

```
1 keep_alive()
2 token = os.environ.get("DISCORD_BOT_SECRET")
3 client.run(token)
```

After doing this and hitting the green "Run" button again, you should see some changes to your Repl. For one, you'll see a new pane in the top right which shows the web output from your server. We can see that visiting our Repl now returns a basic web page showing the "I'm alive" string that we told our web server to return by default. In the bottom-right pane, you can also see some additional output from Flask starting up and running continuously, listening for requests.

```

import discord
import os

from keep_alive import keep_alive

client = discord.Client()

@client.event
async def on_ready():
    print("I'm in")
    print(client.user)

@client.event
async def on_message(message):
    if message.author != client.user:
        await client.send_message(message.channel, message.content[::-1])

keep_alive()
token = os.environ.get("DISCORD_BOT_SECRET")
client.run(token)

```

I'm alive

```

* Tip: There are .env files present. Do pip install python-dotenv to use them.
* Serving Flask app "" (lazy loading)
* Environment: production
WARNING: Do not use the development server in a
production environment.
Use a production WSGI server instead.
* Debug mode: off
* Running on http://0.0.0.0:8080/ (Press CTRL+C to
quit)
172.18.0.1 - - [17/Aug/2018 07:02:16] "GET / HTTP/1.1"
200 -
I'm in
My first bot#4787
172.18.0.1 - - [17/Aug/2018 07:07:55] "GET / HTTP/1.1"
200 -
172.18.0.1 - - [17/Aug/2018 07:07:55] "GET /favicon.ico
HTTP/1.1" 404 -

```

Output from our Flask server

Now your bot will stay alive even after closing your browser or shutting down your development machine. Repl will still clean up your server and kill your bot after about one hour of inactivity, so if you don't use your bot for a while, you'll have to log into Repl and start the bot up again. Alternatively, you can set up a third-party (free!) service like [Uptime Robot](#)<sup>97</sup>. Uptime Robot pings your site every 5 minutes to make sure it's still working – usually to notify you of unexpected downtime, but in this case, the constant pings have the side effect of keeping our Repl alive as it will never go more than an hour without receiving any activity.

## Forking and extending our basic bot

This is not a very useful bot as is, but the possibilities are only limited by your creativity now! You can have your bot receive input from a user, process the input, and respond in any way you choose. In fact, with the basic input and output that we've demonstrated, we have most of the components of any modern computer, all of which are based on the [Von Neumann architecture](#)<sup>98</sup> (we could easily add the missing memory by having our bot write to a file, or with a bit more effort link in a [SQLite database](#)<sup>99</sup> for persistent storage).

If you followed along with this tutorial, you'll have your own basic Repl bot to play around with and extend. If you were simply reading, you can this bot at <https://repl.it/@GarethDwyer1/discord-bot>

<sup>97</sup><https://uptimerobot.com/>

<sup>98</sup>[https://en.wikipedia.org/wiki/Von\\_Neumann\\_architecture](https://en.wikipedia.org/wiki/Von_Neumann_architecture)

<sup>99</sup><https://www.sqlite.org/index.html>

bot<sup>100</sup> and extend it how you want (you'll need to add your own token and recreate the .env file still). Happy hacking!

If you're stuck for ideas, why not link up your Discord bot to the Twitch API<sup>101</sup> to get notified when your favourite streamers are online, or build a text adventure<sup>102</sup>.

In the next chapter, we'll build exactly the same bot again but using NodeJS instead of Python. Even if you prefer Python, it's often a good idea to build the same project in two languages so that you can better appreciate the differences and similarities.

---

<sup>100</sup><https://repl.it/@GarethDwyer1/discord-bot>

<sup>101</sup><https://dev.twitch.tv/>

<sup>102</sup>[https://en.wikipedia.org/wiki/Interactive\\_fiction](https://en.wikipedia.org/wiki/Interactive_fiction)

# Building a Discord bot with Node.js and Repl.it

In this tutorial, we'll use [Repl.it<sup>103</sup>](#) and Node.js to build a Discord Chatbot. If you're reading this tutorial, you probably have at least heard of Discord and likely have an existing account. If not, Discord is a VoIP and Chat application that is designed to replace Skype for gamers. The bot we create in this tutorial will be able to join a Discord server and respond to messages sent by people.

If you don't like JavaScript, there's also a Python version of this tutorial in the previous chapter.

You'll find it easier to follow along if you have some JavaScript knowledge and have used Discord or a similar app such as Skype or Telegram before. We won't be covering the very basics of JavaScript, but we will explain each line of code in detail, so if you have any experience with programming, you should be able to follow along.

## Overview and requirements

We'll be doing all of our coding through the Repl.it web IDE and hosting our bot with Repl.it as well, so you won't need to install any additional software on your machine. For this tutorial you will need to create a [Discord<sup>104</sup>](#) account (if you already have one, you can skip this). There are instructions for how to do this in the next section.

In this tutorial, we will be covering:

- Creating an application and a bot user in your Discord account
- Creating a server on Discord
- Adding our bot to our Discord server

Let's get through these admin steps first and then we can get to the fun part of coding our bot.

## Creating a bot in Discord and getting a token

You can sign up for a free account over at [the Discord register page<sup>105</sup>](#), and can download one of their desktop or mobile applications from [the Discord homepage<sup>106</sup>](#). You can also use Discord in the browser.

Once you have an account, you'll want to create a Discord application. Visit [the Discord developer's page<sup>107</sup>](#) and press the "New application" button, as in the image below.

---

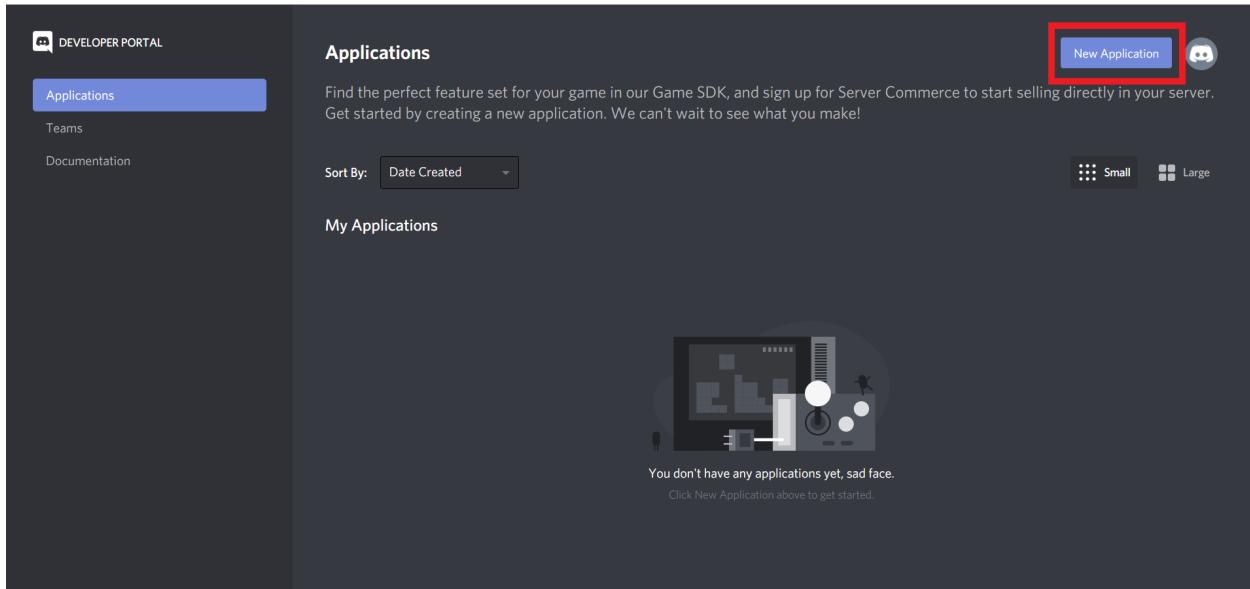
<sup>103</sup><https://repl.it>

<sup>104</sup><https://discordapp.com/>

<sup>105</sup><https://discordapp.com/register>

<sup>106</sup><https://discordapp.com/>

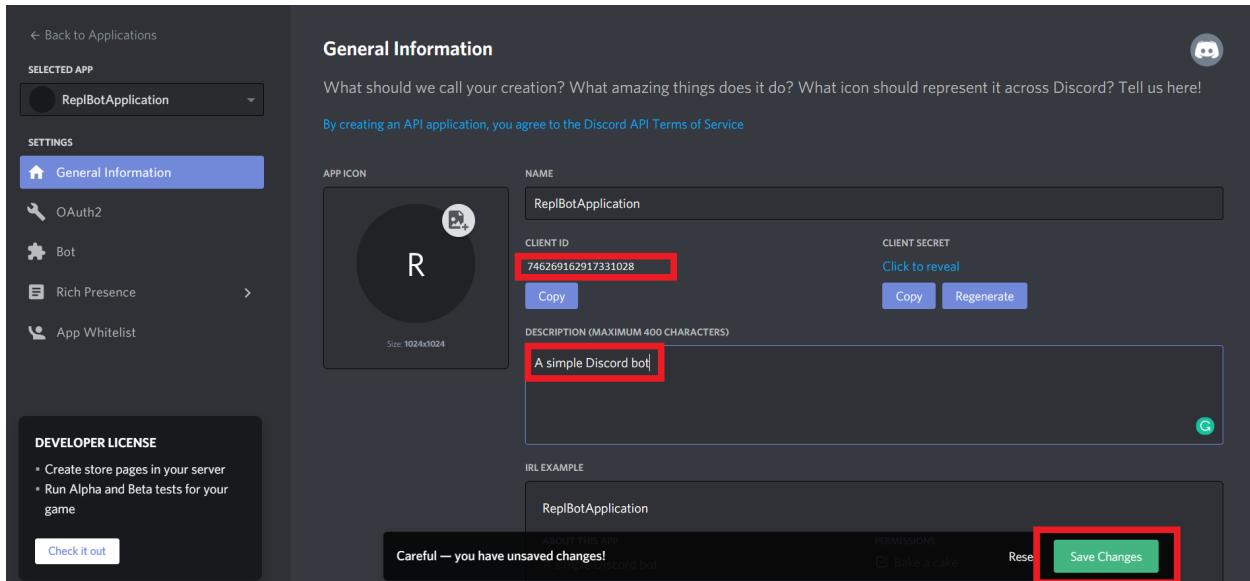
<sup>107</sup><https://discordapp.com/developers/applications/>



### Creating a new Discord application

Fill out a name for your bot and select “Create”.

The first thing to do on the next page is to note your Client ID, which you’ll need to add the bot to the server. You can come back later and get it from this page, or copy it somewhere where you can easily find it later.

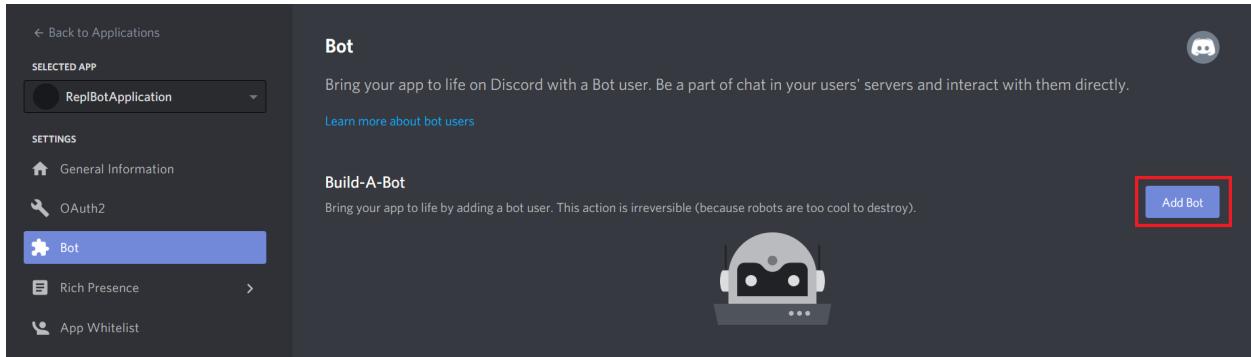


### Record your Client ID

You can also rename the application and provide a description for your bot at this point and press “Save Changes”.

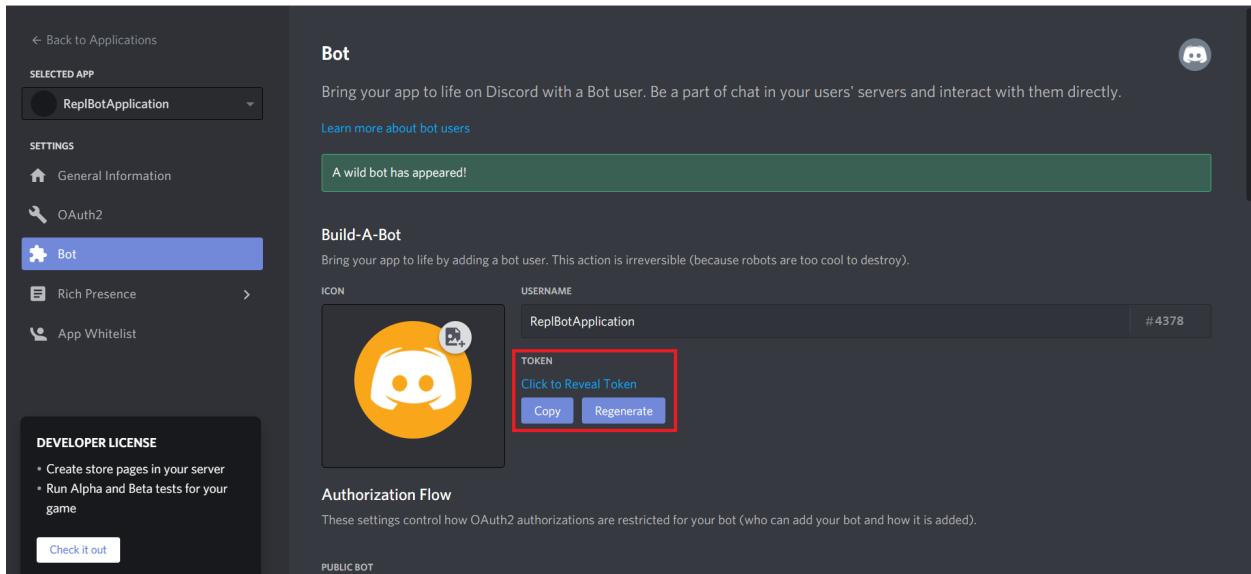
You have now created a Discord application. The next step is to add a bot to this application, so head

over to the “Bot” tab using the menu on the left and press the “Add Bot” button, as indicated below. Click “Yes, do it” when Discord asks if you’re sure about bringing a new bot to life.



Adding a bot to our Discord Application

The last thing we’ll need from our bot is a Token. Anyone who has the bot’s token can prove that they own the bot, so you’ll need to be careful not to share this with anyone. You can get the token by pressing “Click to Reveal Token”, or copy it to your clipboard without seeing it by pressing “Copy”.

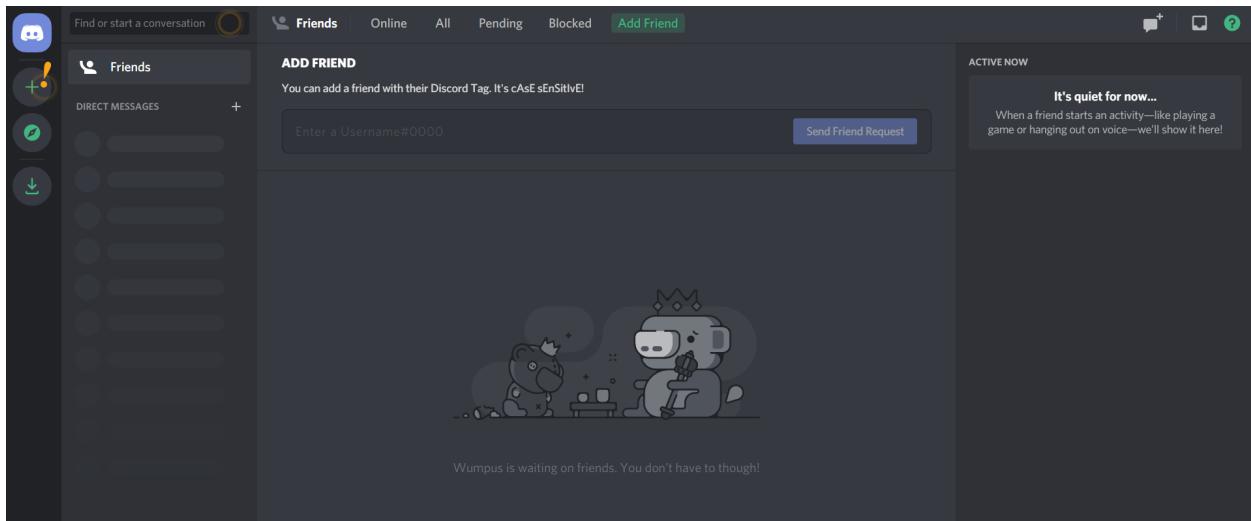


Generating a token for our Discord bot

Take note of your token or copy it to your clipboard, as we’ll need to add it to our code soon.

## Creating a Discord server

If you don’t have a Discord server to add your bot to, you can create one by either opening the desktop Discord application that you downloaded earlier or returning to the Discord home page in your browser. Press the “+” icon indicated by the exclamation mark, as shown below, to create a server.



Creating a Discord server

Press “Create a server” in the screen that follows, and then give your server a name. Once the server is up and running, you can chat with yourself, or invite some friends to chat with you. Soon we’ll invite our bot to chat with us as well.

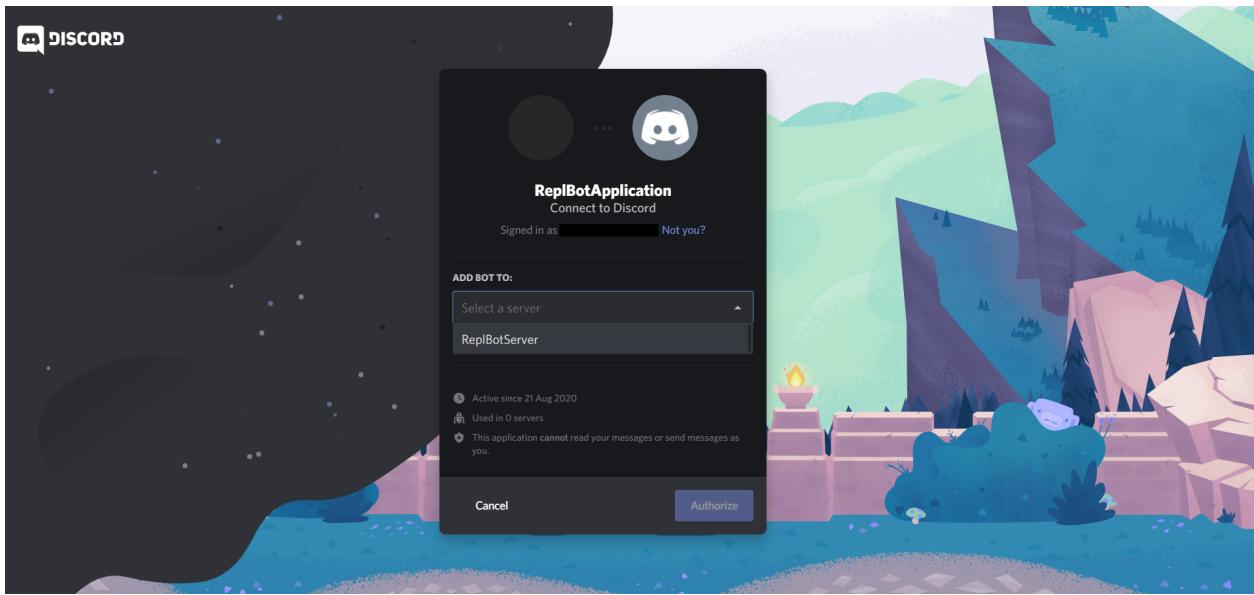
## Adding your Discord bot to your Discord server

Our Discord bot is still just a shell at this stage as we haven’t written any code to allow him to do anything, but let’s go ahead and add him to our Discord server anyway. To add a bot to your server, you’ll need the Client ID from the “General Information” page that we looked at before when we created our ReplBotApplication (ie. the client ID, not the secret bot Token).

Create a URL that looks as follows, but using your Client ID instead of mine at the end:

[https://discordapp.com/api/oauth2/authorize?scope=bot&client\\_id=746269162917331028](https://discordapp.com/api/oauth2/authorize?scope=bot&client_id=746269162917331028)

Visit the URL that you created in your web browser and you’ll see a page similar to the following where you can choose which server to add your bot to.



Authorizing our bot to join our server

Select the server we created in the step before this and hit the “authorize” button. After completing the captcha, you should get an in-app Discord notification telling you that your bot has joined your server.

Now we can get to the fun part of building a brain for our bot!

## Creating a Repl and installing our Discord dependencies

The first thing we need to do is create a Node.js Repl to write the code for our Discord bot. Over at [repl.it<sup>108</sup>](https://repl.it), create a new Repl, choosing “Node.js” as your language.

We don’t need to reinvent the wheel as there is already a great Node wrapper for the Discord bot API called [discord.js<sup>109</sup>](https://discord.js.org/). Normally we would install this third-party library through [npm<sup>110</sup>](https://www.npmjs.com/), but because we’re using Repl.it, we can skip the installation. Our Repl will automatically pull in all dependencies.

In the default `index.js` file that is included with your new Repl, add the following line of code.

```
1 const Discord = require('discord.js');
```

Press the “Run” button and you should see Repl.it installing the Discord library in the output pane on the right, as in the image below.

---

<sup>108</sup><https://repl.it>

<sup>109</sup><https://discord.js.org/>

<sup>110</sup><https://www.npmjs.com/>

```
node v9.7.1 linux/amd64
>
Repl.it: Installing fresh packages
Repl.it: discord.js
+ discord.js@11.4.2
added 8 packages in 2.499s

Repl.it: package installation success
=> undefined
>
```

Installing Discord.js in our Repl

Our bot is nearly ready to go – but we still need to plug in our secret token. This will authorize our code to control our bot.

## Setting up authorization for our bot

By default, Repl code is public. This is great as it encourages collaboration and learning, but we need to be careful not to share our secret bot token (which gives anyone who has access to it full control of our bot).

To get around the problem of needing to give our *code* access to the token while allowing others to access our code but *not* our token, we'll be using [environment variables](#)<sup>111</sup>. On a normal machine, we'd set these directly on our operating system, but using Repl.it we don't have access to this. Repl.it allows us to set secrets in environment variables through a special `.env` file.

First, we need to create a new file called exactly `.env`. Select “Add file” and name this file `.env`. It is important not to leave out the `.` at the beginning. Open this new file and add a variable to define your bot's secret token (note that this is the second token that we got while setting up the bot – different from the Client ID that we used to add our bot to our server). It should look something like:

```
1 DISCORD_BOT_SECRET=NDcUN5T32zcTjMYOM0Y1MTUy.Dk7JBw.ihrTSAO1GKHZSonqvuhTwta16WU
```

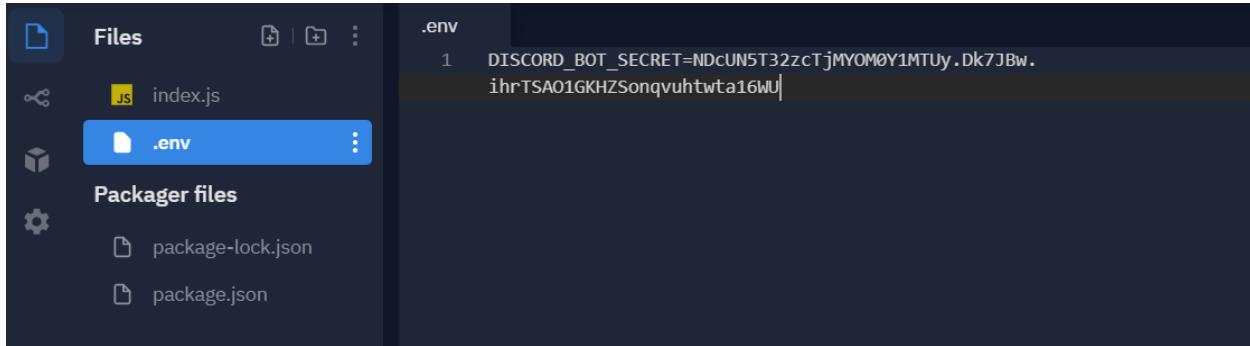
You'll need to:

- Replace the token below (after the `=` sign) with the token that Discord gave you when creating your own bot.
- Be careful about **spacing**. If you put a space on either side of the `=` in your `.env` file, these spaces will be part of the variable name or the value, so make sure you don't have any spaces around the `=` or at the end of the line.

<sup>111</sup><https://www.digitalocean.com/community/tutorials/how-to-read-and-set-environmental-and-shell-variables-on-a-linux-vps>

- Run the code again. Sometimes you'll need to refresh the whole page to make sure that your environment variables are successfully loaded.

In the image below you we've highlighted the “Add file” button, the new file (.env) and how to define the secret token for our bot's use.



Creating our .env file

Let's make a Discord bot that repeats everything we say but in reverse. We can do this in only a few lines of code. In your `index.js` file, add the following:

```

1 const Discord = require('discord.js');
2 const client = new Discord.Client();
3 const token = process.env.DISCORD_BOT_SECRET;
4
5 client.on('ready', () => {
6   console.log("I'm in");
7   console.log(client.user.username);
8 });
9
10 client.on('message', msg => {
11   if (msg.author.id != client.user.id) {
12     msg.channel.send(msg.content.split('')).reverse().join(''));
13   }
14 });
15
16 client.login(token);

```

Let's tear this apart line by line to see what it does.

- Line 1 is what we had earlier. This line both tells Repl.it to install the third party library and brings it into this file so that we can use it.
- In line 2, we create a Discord Client. We'll use this client to send commands to the Discord server to control our bot and send it commands.

- In **line 3** we retrieve our secret token from the environment variables (which Repl.it sets from our `.env` file).
- In **line 5**, we define an event for our client, which defines how our bot should react to the “ready” event. The Discord bot is going to run *asynchronously*, which might be a bit confusing if you’re used to running standard synchronous code. We won’t go into asynchronous coding in depth here, but if you’re interested in what this is and why it’s used, there’s a good guide over at [RisingStack<sup>112</sup>](#). In short, instead of running the code in our file from top to bottom, we’ll be running pieces of code in response to specific events.
- In **lines 6-8** we define how our bot should respond to the “ready” event, which is fired when our bot successfully joins a server. We instruct our bot to output some information server side (i.e. this will be displayed in our Repl’s output, but not sent as a message through to Discord). We’ll print a simple `I'm in` message to see that the bot is there and print our bot’s username (if you’re running multiple bots, this will make it easier to work out who’s doing what).
- **Lines 10-14** are similar, but instead of responding to an “ready” event, we tell our bot how to handle new messages. **Line 11** says we only want to respond to messages that aren’t from us (otherwise our bot will keep responding to himself – you can remove this line to see why that’s a problem), and **line 12** says we’ll send a new message to the same channel where we received a message (`msg.channel`) and the content we’ll send will be the same message that we received, but backwards. To reverse a string, we split it into its individual characters, reverse the resulting array, and then join it all back into a string again.

The last line fires up our bot and uses the token we loaded earlier to log into Discord.

Press the big green “Run” button again and you should see your bot reporting a successful channel join in the Repl output.

---

<sup>112</sup><https://blog.risingstack.com/node-hero-async-programming-in-node-js/>

```

const Discord = require('discord.js');
const client = new Discord.Client();
const token = process.env.DISCORD_BOT_SECRET;

client.on('ready', () => {
    console.log("I'm in");
    console.log(client.user.username);
});

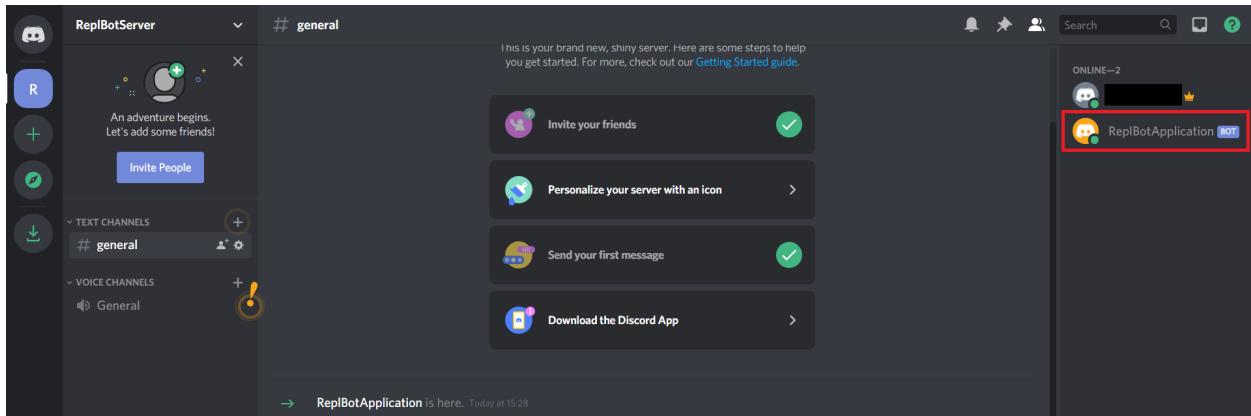
client.on('message', msg => {
    if (msg.author.id != client.user.id) {
        var rev = msg.content.split('').reverse()
            .join('');
        msg.channel.send(rev);
    }
});

client.login(token);

```

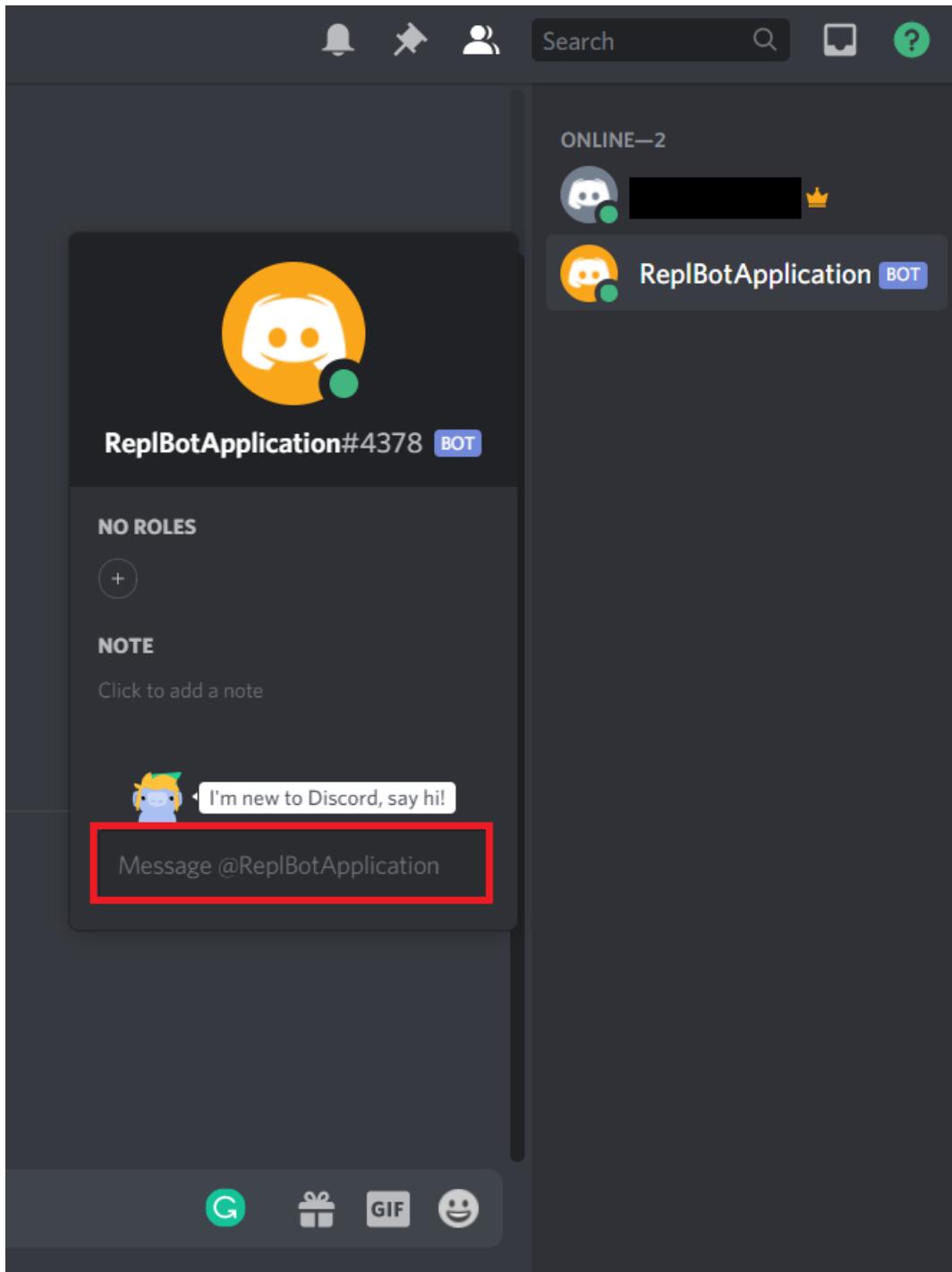
Repl output showing channel join

Open Discord, and from within the server we created earlier, select your ReplBotApplication from the pane on the right-hand side of the screen.



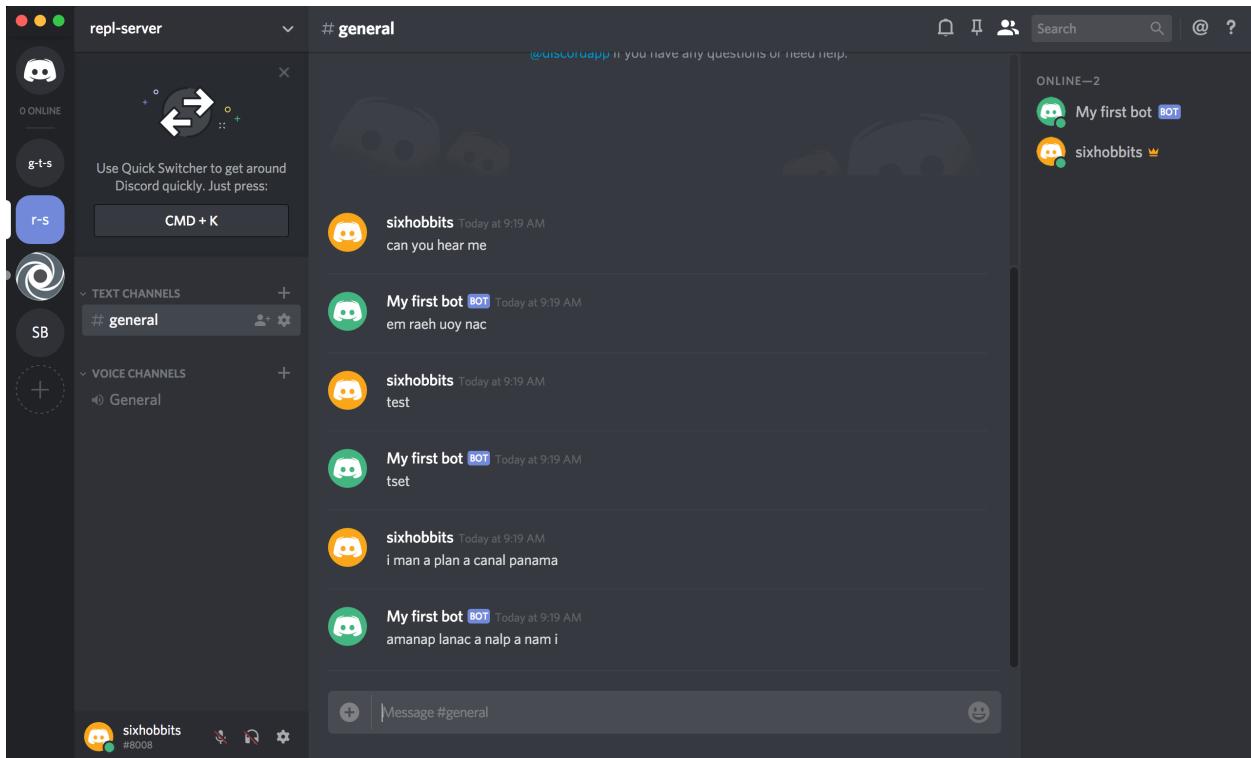
The Repl bot is active

Once you have selected this, you will be able to send a message (by typing into the box highlighted below) and see your bot respond!



Send a message to your bot

The bot responds each time, reversing the text we enter.



Our bot can talk!

## Keeping our bot alive

Your bot can now respond to messages, but only for as long as your Repl is running. If you close your browser tab or shut down your computer, your bot will stop and no longer respond to messages on Discord.

Repl will keep your code running after you close the browser tab only if you are running a web server. Our bot doesn't require an explicit web server to run, but we can create a server and run it in the background just to keep our Repl alive.

Create a new file in your project called `keep_alive.js` and add the following code:

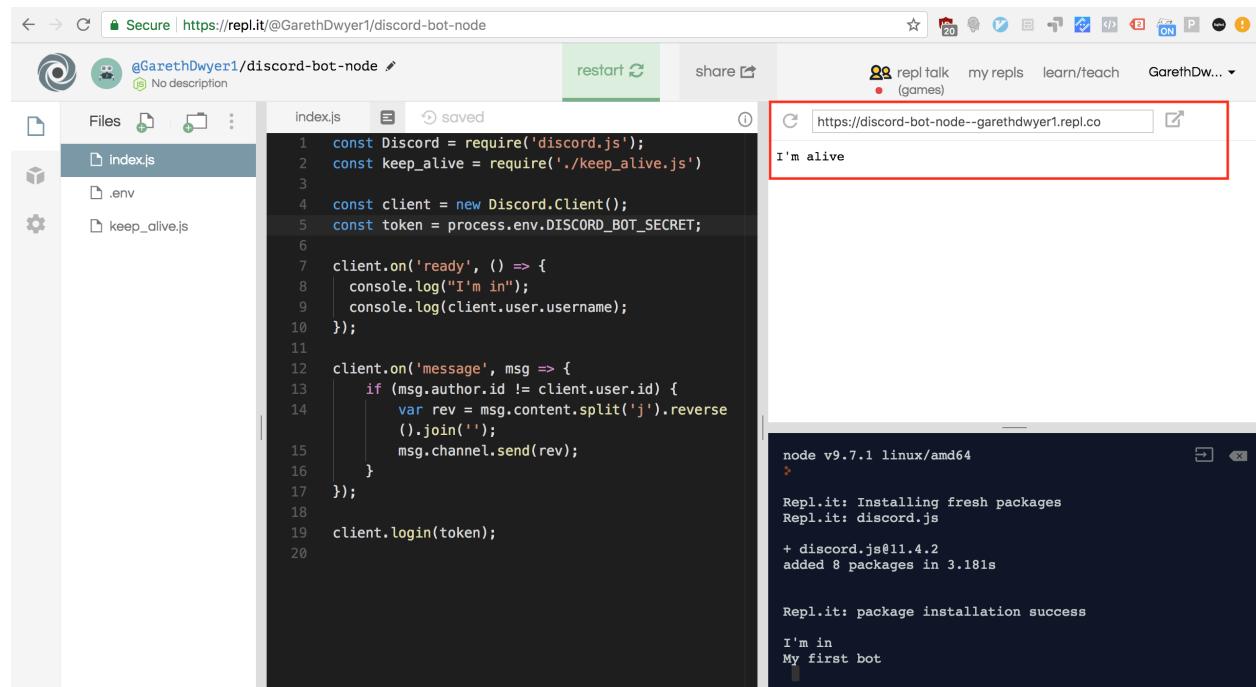
```
1 var http = require('http');
2
3 http.createServer(function (req, res) {
4   res.write("I'm alive");
5   res.end();
6 }).listen(8080);
```

We won't go over this in detail as it's not central to our bot, but here we start a web server that will return "I'm alive" if anyone visits it.

In our `index.js` file, we need to add a require statement for this server at the top. Add the following line near the top of `index.js`.

```
1 const keep_alive = require('./keep_alive.js')
```

After doing this and hitting the green "Run" button again, you should see some changes to your Repl. For one, you'll see a new pane in the top right which shows the web output from your server. We can see that visiting our Repl now returns a basic web page showing the "I'm alive" string that we told our web server to return by default.



Running a Node server in the background

Now your bot will stay alive even after closing your browser or shutting down your development machine. Repl will still clean up your server and kill your bot after about one hour of inactivity, so if you don't use your bot for a while, you'll have to log into Repl and start the bot up again. Alternatively, you can set up a third-party (free!) service like [Uptime Robot](#)<sup>113</sup>. Uptime Robot pings your site every 5 minutes to make sure it's still working – usually to notify you of unexpected downtime, but in this case the constant pings have the side effect of keeping our Repl alive as it will never go more than an hour without receiving any activity. Note that you need to select the HTTP option instead of the Ping option when setting up Uptime Robot, as Repl.it requires regular HTTP requests to keep your chatbot alive.

<sup>113</sup><https://uptimerobot.com/>

## Forking and extending our basic bot

This is not a very useful bot as is, but the possibilities are only limited by your creativity now! You can have your bot receive input from a user, process the input, and respond in any way you choose. In fact, with the basic input and output that we've demonstrated, we have most of the components of any modern computer, all of which are based on the [Von Neumann architecture<sup>114</sup>](#) (we could easily add the missing memory by having our bot write to a file, or with a bit more effort link in a [SQLite database<sup>115</sup>](#) for persistent storage).

If you followed along this tutorial, you'll have your own basic Repl bot to play around with and extend. If you were simply reading, you can easily fork my bot at [https://repl.it/@GarethDwyer1/discord-bot-node<sup>116</sup>](https://repl.it/@GarethDwyer1/discord-bot-node) and extend it how you want (you'll need to add your own token and recreate the .env file still). Happy hacking!

If you're stuck for ideas, why not link up your Discord bot to the [Twitch API<sup>117</sup>](#) to get notified when your favourite streamers are online, or build a [text adventure<sup>118</sup>](#).

In the next chapter, we'll be looking at building our own basic web application, using Django. This tutorial will also introduce you to HTML, JavaScript, and jQuery and will assist you in getting to the point where you can begin to build your own custom web applications.

---

<sup>114</sup>[https://en.wikipedia.org/wiki/Von\\_Neumann\\_architecture](https://en.wikipedia.org/wiki/Von_Neumann_architecture)

<sup>115</sup><https://www.sqlite.org/index.html>

<sup>116</sup><https://repl.it/@GarethDwyer1/discord-bot-node>

<sup>117</sup><https://dev.twitch.tv/>

<sup>118</sup>[https://en.wikipedia.org/wiki/Interactive\\_fiction](https://en.wikipedia.org/wiki/Interactive_fiction)

# Creating and hosting a basic web application with Django and Repl.it

In this tutorial, we'll be using Django to create an online service that shows visitors their current weather and location. We'll develop the service and host it using [Repl.it](#)<sup>119</sup>.

To work through this tutorial, you should ideally have basic knowledge of Python and some knowledge of web application development. However, we'll explain all of our reasoning and each line of code thoroughly, so if you have any programming experience you should be able to follow along as a complete Python or web app beginner too. We'll also be making use of some HTML, JavaScript, and jQuery, so if you have been exposed to these before you'll be able to work through more quickly. If you haven't, this will be a great place to start.

To display the weather at the user's current location, we'll have to tie together a few pieces. The main components of our system are:

- A [Django](#)<sup>120</sup> application, to show the user a webpage with dynamic data
- [Ipify](#)<sup>121</sup> to get our visitors' IP address so that we can guess their location
- [ip-api](#)<sup>122</sup> to look up our visitors' city and country using their IP address
- [Open Weather Map](#)<sup>123</sup> to get the current weather at our visitors' location.

The main goals of this tutorial are to:

- Show how to set up and host a Django application using Repl.it.
- Show how to join existing APIs together to create a new service.

By using this tutorial as a starting point, you can easily create your own bespoke web applications. Instead of showing weather data to your visitors, you could, for example, pull and combine data from any of the hundreds of APIs found at [this list of public APIs](#)<sup>124</sup>.

---

<sup>119</sup><https://repl.it>

<sup>120</sup><https://www.djangoproject.com/>

<sup>121</sup><https://www.ipify.org/>

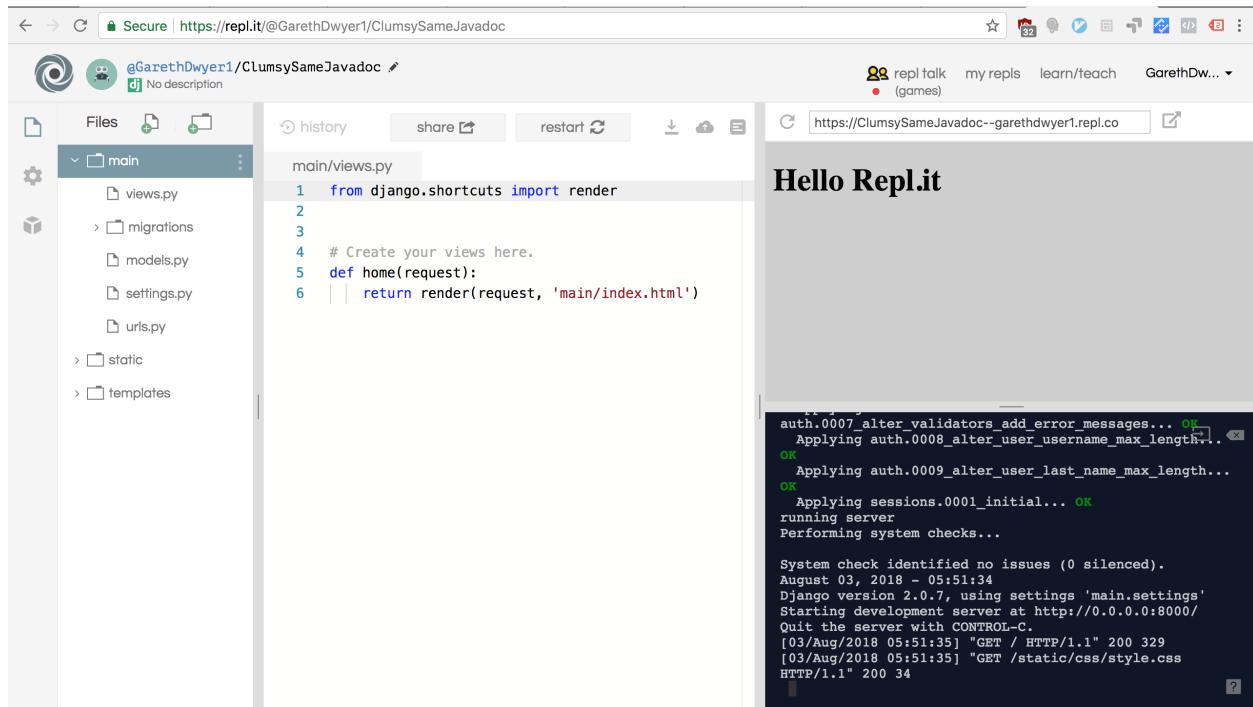
<sup>122</sup><http://ip-api.com/>

<sup>123</sup><https://openweathermap.org>

<sup>124</sup><https://github.com/toddmotto/public-apis>

## Setting up

Create a new Repl and select “Django Template” from the list of available languages and frameworks. Repl.it will go ahead and set up a full Django project for you with some sensible defaults. The project will include a single page saying “Hello Repl.it”. It should look very similar to the page below. (TO DO - update image)



The repl.it IDE on a new Django project

In the left-most panel, you can see all the files that make up your Django project.

We won't explain what all these different components are for and how they tie together in this tutorial. If you want to understand Django better, you should definitely go through their [official tutorial](#)<sup>125</sup>. Normally I recommend Flask over Django<sup>126</sup> for beginners and for projects this simple, but Repl.it makes the Django setup easy by providing this initial set up. We'll be modifying only a few of these files to get the results that we want.

In the middle Repl.it pane, you can see your code. In the top-right pane, you can see your web application, as a visitor would see it. In the bottom-right pane, you can see any errors or other console output from your Repl.it webserver.

<sup>125</sup><https://docs.djangoproject.com/en/2.0/intro/tutorial01/>

<sup>126</sup><https://www.codementor.io/garethdwyer/flask-vs-django-why-flask-might-be-better-4xs7mdf8v>

## Changing our static content

Viewing the default website isn't that interesting. Let's make a small change to make the website our own. The first thing we'll change is the static "front end" of the website – the text that is displayed to all visitors.

The file that we'll need to make changes to is the main template at `templates/main/index.html`. You can access it by selecting it from the left file bar. It should look like this:

```

1  {% extends "base.html" %}

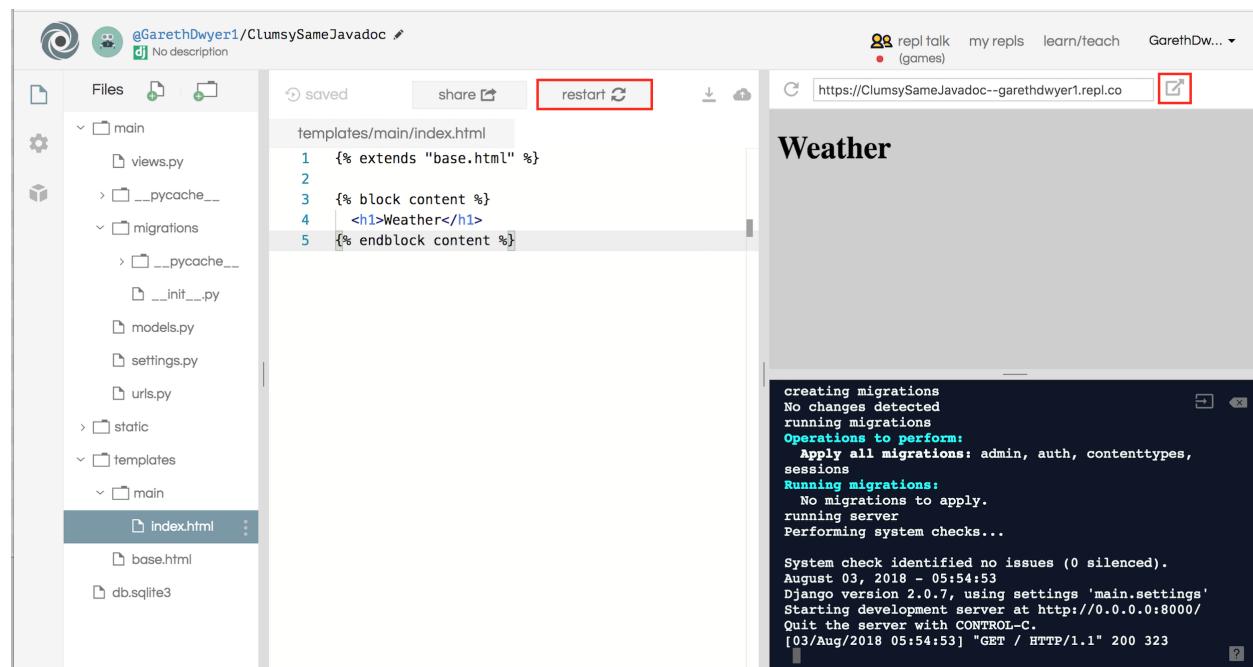
2

3  {% block content %}
4      <h1>Hello Repl.it</h1>
5  {% endblock content %}

```

This is a file written in [Django's template language](#)<sup>127</sup>, which often looks very much like HTML (and is usually found in files with a `.html` extension), but which contains some extra functionality to help us load dynamic data from the back end of our web application into the front end for our users to see.

Change the code where it says "Hello Repl.it" to read "Weather", and press the restart button as indicated below to see the result change from "Hello Repl.it" to "Weather".



Repl.it Restart and pop-out buttons

<sup>127</sup><https://docs.djangoproject.com/en/2.0/topics/templates/>

You can also press the pop-out button to the right of the URL bar to open only the resulting web page that we're building, as a visitor would see it. You can share the URL with anyone and they'll be able to see your Weather website already!

Changing the static text that our visitors see is a good start, but our web application still doesn't *do* anything. We'll change that in the next step by using JavaScript to get our user's IP Address.

## Calling IPIFY from JavaScript

An IP address is like a phone number. When you visit "google.com", your computer actually looks up the name google.com to get a resulting IP address that is linked to one of Google's servers. While people find it easier to remember names like "google.com", computers work better with numbers. Instead of typing "google.com" into your browser toolbar, you could type the IP address 216.58.223.46, with the same results. Every device connecting to the internet, whether to serve content (like google.com) or to consume it (like you, reading this tutorial) has an IP address.

IP addresses are interesting to us because it is possible to guess a user's location based on their IP address. (In reality, this is an [imprecise and highly complicated<sup>128</sup>](#) process, but for our purposes it will be more than adequate). We will use the web service [ipify.org<sup>129</sup>](#) to retrieve our visitors' IP addresses.

In the Repl.it files tab, navigate to `templates/base.html`, which should look as follows.

```

1  {% load staticfiles %}
2  <!DOCTYPE html>
3
4  <html lang="en">
5  <head>
6      <meta charset="UTF-8">
7      <title>Hello Django</title>
8      <meta charset="UTF-8"/>
9      <meta name="viewport" content="width=device-width, initial-scale=1"/>
10     <link rel="stylesheet" href="{% static "css/style.css" %}">
11 </head>
12 <body>
13     {% block content %}{% endblock content %}
14 </body>
15 </html>
```

The "head" section of this template is between lines 5 and 11 – the opening and closing `<head>` tags. We'll add our scripts directly below the `<link ...>` on line 10 and above the closing `</head>` tag on line 11. Modify this part of code to add the following lines:

---

<sup>128</sup><https://dyn.com/blog/finding-yourself-the-challenges-of-accurate-ip-geolocation/>

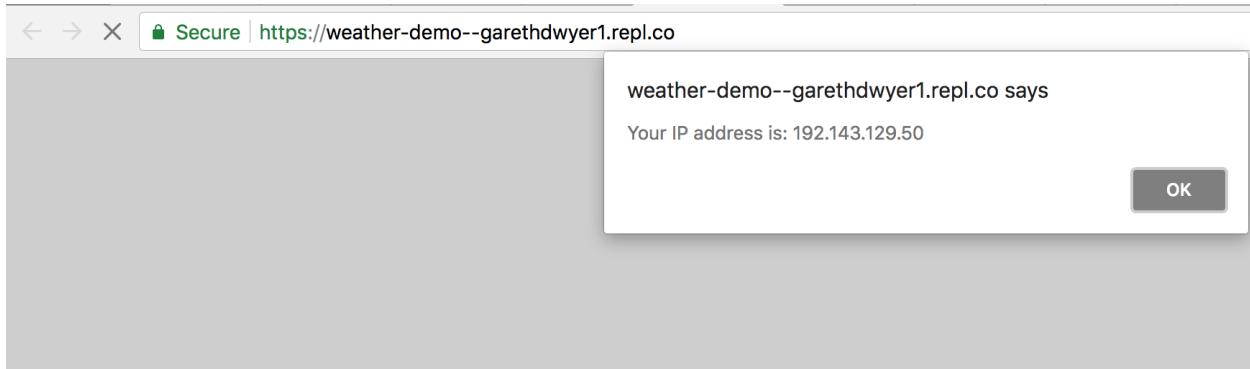
<sup>129</sup><https://www.ipify.org/>

```

1 <script>
2   function use_ip(json) {
3     alert("Your IP address is: " + json.ip);
4   }
5 </script>
6
7 <script src="https://api.ipify.org?format=jsonp&callback=use_ip"></script>
```

These are two snippets of JavaScript. The first (lines 1-5) is a function that when called will display a pop-up box (an “alert”) in our visitor’s browser showing their IP address from the json object that we pass in. The second (line 7) loads an external script from ipify’s API and asks it to pass data (including our visitor’s IP address) along to the use\_ip function that we provide.

If you open your web app again and refresh the page, you should see this script in action (if you’re running an adblocker, it might block the ipify scripts, so try disabling that temporarily if you have any issues).



This code doesn’t do anything with the IP address except display it to the user, but it is enough to see that the first component of our system (getting our user’s IP Address) is working. This also introduces the first *dynamic* functionality to our app – before, any visitor would see exactly the same thing, but now we can show each visitor something related specifically to them (no two people have the same IP address).

Now instead of simply showing this IP address to our visitor, we’ll modify the code to rather pass it along to our Repl webserver (the “backend” of our application), so that we can use it to fetch location information through a different service.

## Adding a new route and view, and passing data

Currently our Django application only has one route, the default (/) route which is loaded as our home page. We’ll add another route at /get\_weather\_from\_ip where we can pass an IP address to our application to detect the location and get a current weather report.

To do this, we’ll need to modify the files at `main/views.py` and `main/urls.py`.

Edit `urls.py` to look as follows (add line 8, but you shouldn't need to change anything else).

```

1 from django.conf.urls import url
2 from django.contrib import admin
3 from main import views
4
5 urlpatterns = [
6     url(r'^admin/', admin.site.urls),
7     url(r'^$', views.home, name='home'),
8     url(r'^get_weather_from_ip/$', views.get_weather_from_ip, name="get_weather_from_ip\
9     "),
10 ]

```

We've added a definition for the `get_weather_from_ip` route, telling our app that if anyone visits [http://weather-demo--garethdwyer1.repl.co/get\\_weather\\_from\\_ip](http://weather-demo--garethdwyer1.repl.co/get_weather_from_ip)<sup>130</sup> then we should trigger a function in our `views.py` file that is also called `get_weather_from_ip`. Let's write that function now.

In your `views.py` file, add a `get_weather_from_ip()` function beneath the existing `home()` one, and add an import for `JsonResponse` on line 2. Your whole `views.py` file should now look like this:

```

1 from django.shortcuts import render
2 from django.http import JsonResponse
3
4
5 # Create your views here.
6 def home(request):
7     return render(request, 'main/index.html')
8
9 def get_weather_from_ip(request):
10    print(request.GET.get("ip_address"))
11    data = {"weather_data": 20}
12    return JsonResponse(data)

```

By default, Django passes a `request` argument to all views. This is an object that contains information about our user and the connection, and any additional arguments passed in the URL. As our application isn't connected to any weather services yet, we'll just make up a temperature (20) and pass that back to our user as JSON.

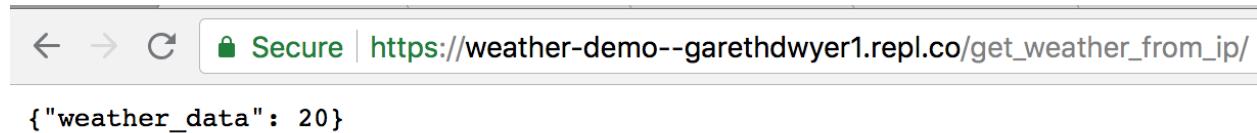
In line 10, we print out the IP address that we will pass along to this route from the GET arguments (we'll look at how to use this later). We then create the fake data (which we'll later replace with real data) and return a JSON response (the data in a format that a computer can read more easily, with no formatting). We return JSON instead of HTML because our system is going to use this route

---

<sup>130</sup>[http://weather-demo--garethdwyer1.repl.co/get\\_weather\\_from\\_ip](http://weather-demo--garethdwyer1.repl.co/get_weather_from_ip)

internally to pass data between the front and back ends of our application, but we don't expect our users to use this directly.

To test this part of our system, open your web application in a new tab and add `/get_weather_from_ip?ip_address=123` to the URL. Here, we're asking our system to fetch weather data for the IP address 123 (not a real IP address). In your browser, you'll see the fake weather data displayed in a format that can easily be programmatically parsed.



A screenshot of a web browser window. The address bar shows a secure connection with a padlock icon and the URL `https://weather-demo--garethdwyer1.repl.co/get_weather_from_ip/`. The main content area displays a single line of JSON code: `{"weather_data": 20}`.

#### Viewing the fake JSON data

In our Repl's output, we can see that the backend of our application has found the "IP address" and printed it out, between some other outputs telling us which routes are being visited and which port our server is running on:



A screenshot of a terminal window showing the output of a Django application. The output includes:

- System check identified no issues (0 silenced).
- July 27, 2018 - 07:16:10
- Django version 2.0.7, using settings 'main.settings'
- Starting development server at `http://0.0.0.0:8000/`
- Quit the server with CONTROL-C.
- Log entries:
  - [27/Jul/2018 07:16:11] "GET / HTTP/1.1" 200 785
  - [27/Jul/2018 07:20:38] "GET / HTTP/1.1" 200 785
  - [27/Jul/2018 07:20:39] "GET /static/css/style.css HTTP/1.1" 200 34
  - 123**
  - [27/Jul/2018 07:20:43] "GET /get\_weather\_from\_ip/?ip\_address=123 HTTP/1.1" 200 23

Django print output of fake IP

The steps that remain now are to:

- pass the user's real IP address to our new route in the background when the user visits our main page
- add more backend logic to fetch the user's location from the IP address
- add logic to fetch the user's weather from their location

- display this data to the user.

Let's start by using [Ajax<sup>131</sup>](#) to pass the user's IP address that we collected before to our new route, without our user having to explicitly visit the `get_weather_from_ip` endpoint or refresh their page.

## Calling a Django route using Ajax and jQuery

We'll use [Ajax through jQuery<sup>132</sup>](#) to do a "partial page refresh" – that is, to update part of the page the user is seeing by sending new data from our backend code without the user needing to reload the page.

To do this, we need to include jQuery as a library.<sup>133</sup>

Note: usually you wouldn't add JavaScript directly to your `base.html` template, but to keep things simpler and to avoid creating too many files, we'll be diverging from some good practices. See [the Django documentation<sup>133</sup>](#) for some guidance on how to structure JavaScript properly in larger projects.

In your `templates/base.html` file, add the following script above the line where we previously defined the `use_ip()` function.

```
1 <script
2   src="https://code.jquery.com/jquery-3.3.1.min.js"
3   integrity="sha256-FgpCb/KJQ1LNf0u91ta32o/NMZxltwRo8QtmkMRdAu8="
4   crossorigin="anonymous"></script>
```

This loads the entire jQuery library from a [CDN<sup>134</sup>](#), allowing us to complete certain tasks using fewer lines of JavaScript.

Now, modify the `use_ip()` script that we wrote before to call our backend route using Ajax. The new `use_ip()` function should be as follows:

---

<sup>131</sup>[https://en.wikipedia.org/wiki/Ajax\\_\(programming\)](https://en.wikipedia.org/wiki/Ajax_(programming))

<sup>132</sup><http://api.jquery.com/jquery.ajax/>

<sup>133</sup><https://docs.djangoproject.com/en/2.0/howto/static-files/>

<sup>134</sup><https://www.cloudflare.com/learning/cdn/what-is-a-cdn/>

```

1  function use_ip(json) {
2    $.ajax({
3      url: {% url 'get_weather_from_ip' %},
4      data: {"ip": json.ip},
5      dataType: 'json',
6      success: function (data) {
7        document.getElementById("weatherdata").innerHTML = data.weather_data
8      }
9    });
10  }

```

Our new `use_ip()` function makes an [asynchronous<sup>135</sup>](#) call to our `get_weather_from_ip` route, sending along the IP address that we previously displayed in a pop-up box. If the call is successful, we call a new function (in the `success:` section) with the returned data. This new function (line 7) looks for an HTML element with the ID of `weatherdata` and replaces the contents with the `weather_data` attribute of the response that we received from `get_weather_from_ip` (which at the moment is still hardcoded to be “20”).

To see the results, we’ll need to add an HTML element as a placeholder with the id `weatherdata`. Do this in the `templates/main/index.html` file as follows.

```

1  {% extends "base.html" %}
2
3  {% block content %}
4    <h1>Weather</h1>
5    <p id=weatherdata></p>
6  {% endblock %}

```

This adds an empty HTML *paragraph* element which our JavaScript can populate once it has the required data.

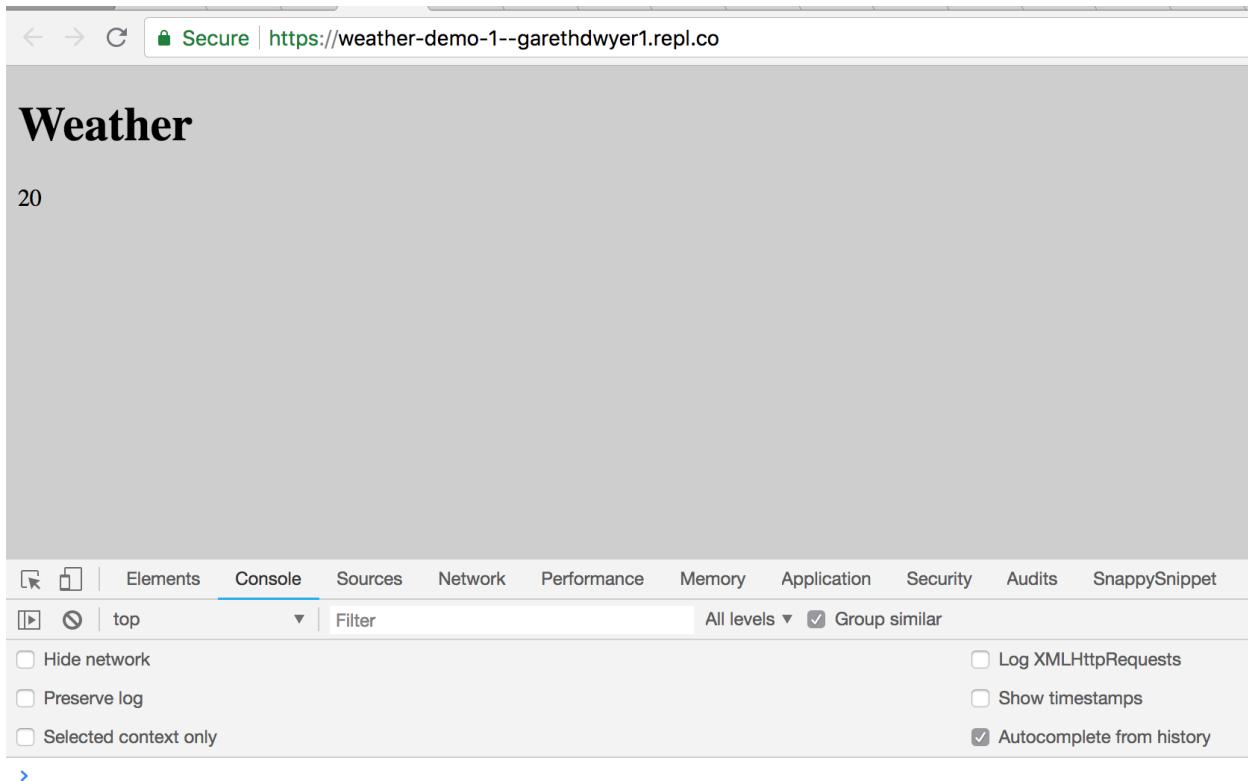
Now reload the app and you should see our fake 20 being displayed to the user. If you don’t see what you expect, open up your browser’s developer tools for [Chrome<sup>136</sup>](#) and [Firefox<sup>137</sup>](#)) and have a look at the Console section for any JavaScript errors. A clean console (with no errors) is shown below.

---

<sup>135</sup><http://api.jquery.com/jquery.ajax/>

<sup>136</sup><https://developers.google.com/web/tools/chrome-devtools/>

<sup>137</sup><https://developer.mozilla.org/en/docs/Tools>



Now it's time to change out our mock data for real data by calling two services backend – the first to get the user's location from their IP address and the second to fetch the weather for that location.

## Using ip-api.com for geolocation

The service at [ip-api.com](http://ip-api.com)<sup>138</sup> is very simple to use. To get the country and city from an IP address we only need to make one web call. We'll use the python requests library for this, so first we'll have to add an import for this to our `views.py` file, and then write a function that can translate IP addresses to location information. Add the following import to your `views.py` file:

```
1 import requests
```

and above the `get_weather_from_ip()` function, add the `get_location_from_ip()` function as follows:

---

<sup>138</sup><http://ip-api.com>

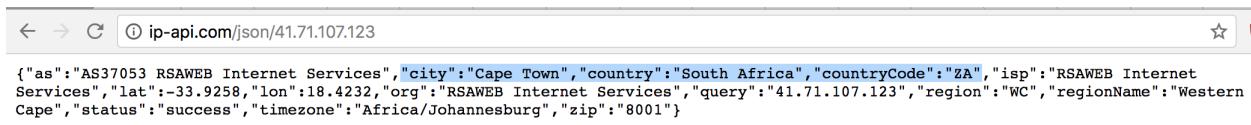
```

1 def get_location_from_ip(ip_address):
2     response = requests.get("http://ip-api.com/json/{}".format(ip_address))
3     return response.json()

```

Note: again we are diverging from best practice in the name of simplicity. Usually whenever you write any code that relies on networking (as above), you should add [exception handling<sup>139</sup>](#) so that your code can fail more gracefully if there are problems.

You can see the response that we'll be getting from this service by trying it out in your browser. Visit [http://ip-api.com/json/41.71.107.123<sup>140</sup>](http://ip-api.com/json/41.71.107.123) to see the JSON response for that specific IP address.



Take a look specifically at the highlighted location information that we'll need to extract to pass on to a weather service.

Before we set up the weather component, let's display the user's current location data instead of the hardcoded temperature that we had before. Change the `get_weather_from_ip()` function to call our new function and pass along some useful data as follows:

```

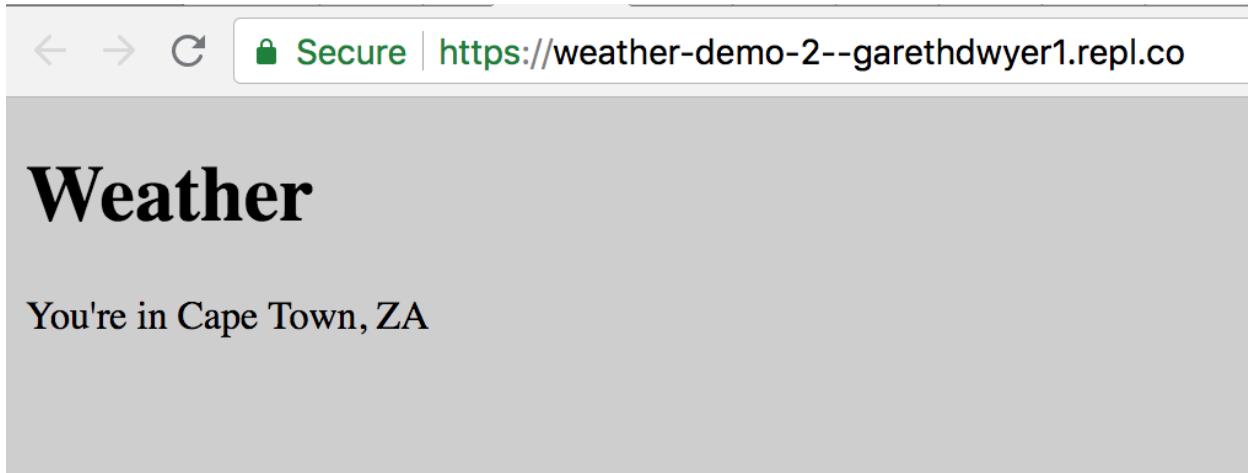
1 def get_weather_from_ip(request):
2     ip_address = request.GET.get("ip")
3     location = get_location_from_ip(ip_address)
4     city = location.get("city")
5     country_code = location.get("countryCode")
6     s = "You're in {}, {}".format(city, country_code)
7     data = {"weather_data": s}
8     return JsonResponse(data)

```

Now, instead of just printing the IP address that we get sent and making up some weather data, we use the IP address to guess the user's location, and pass the city and country code back to the template to be displayed. If you reload your app again, you should see something similar to the following (though hopefully with your location instead of mine).

<sup>139</sup><https://docs.python.org/3/tutorial/errors.html>

<sup>140</sup><http://ip-api.com/json/41.71.107.123>



weather app, location showing

That's the location component of our app done and dusted – let's move on to getting weather data for that location now.

## Getting weather data from OpenWeatherMap

To get weather data automatically from [OpenWeatherMap<sup>141</sup>](#), you'll need an API Key. This is a unique string that OpenWeatherMap gives to each user of their service and it's used mainly to restrict how many calls each person can make in a specified period. Luckily, OpenWeatherMap provides a generous “free” allowance of calls, so we won’t need to spend any money to build our app. Unfortunately, this allowance is not quite generous enough to allow me to share my key with every reader of this tutorial, so you’ll need to sign up for your own account and generate your own key.

Visit [openweathermap.org<sup>142</sup>](#), hit the “sign up” button, and register for the service by giving them an email address and choosing a password. Then navigate to the [API Keys<sup>143</sup>](#) section and note down your unique API key (or copy it to your clipboard).

<sup>141</sup><https://openweathermap.org/>

<sup>142</sup><https://openweathermap.org/>

<sup>143</sup>[https://home.openweathermap.org/api\\_keys](https://home.openweathermap.org/api_keys)

The screenshot shows the 'API keys' section of the OpenWeatherMap website. At the top, there's a navigation bar with links like 'Weather', 'Maps', 'API', 'Price', etc. Below that is a sub-navigation bar with 'Setup', 'API keys', 'My Services', etc. A red box highlights the 'cb932829eacb6a0e...' key in the 'Key' column. To its right, under 'Name', it says 'Default' with edit and delete icons. On the far right, there's a 'Create key' form with a placeholder 'Name' field.

OpenWeatherMap API Key page

This key is a bit like a password – when we use OpenWeatherMap’s service, we’ll always send along this key to indicate that it’s us making the call. Because Repl.it’s projects are public by default, we’ll need to be careful to keep this key private and prevent other people making too many calls using our OpenWeatherMap quota (potentially making our app fail when OpenWeatherMap starts blocking our calls). Luckily Repl.it provides a neat way of solving this problem using `.env` files<sup>144</sup>.

In your project, create a new file using the “New file” button as shown below. Make sure that the file is in the root of your project and that you name the file `.env` (in Linux, starting a filename with a `.` usually indicates that it’s a system or configuration file). Inside this file, define the `OPEN_WEATHER_TOKEN` variable as follows, but using your own token instead of the fake one below. Make sure not to have a space on either side of the `=` sign.

1 `OPEN_WEATHER_TOKEN=1be9250b94bf6803234b56a87e55f`

The screenshot shows a Repl.it project interface. At the top, it displays the repository details: '@GarethDwyer1/weather'. Below that is a toolbar with icons for 'Files' (highlighted with a red box), 'Upload', and 'Share'. A progress bar indicates 'saving...' next to the upload icon. The main area shows a list of files, with a red box highlighting the '+' icon in the 'Files' tab.

Creating a new file

Repl.it will load the contents of this file into our server’s environment variables<sup>145</sup>. We’ll be able to access this using the `os` library in Python, but when other people view or fork our Repl, they won’t

<sup>144</sup><https://repl.it/site/docs/secret-keys>

<sup>145</sup>[https://wiki.archlinux.org/index.php/environment\\_variables](https://wiki.archlinux.org/index.php/environment_variables)

see the `.env` file, keeping our API key safe and private.

To fetch weather data, we need to call the OpenWeatherMap api, passing along a search term. To make sure we're getting the city that we want, it's good to pass along the country code as well as the city name. For example, to get the weather in London right now, we can visit (again, you'll need to add your own API key in place of the string after `appid=`) <https://api.openweathermap.org/data/2.5/weather?q=London&units=metric&appid=cb932829ea...>

To test this, you can visit the URL in your browser first. If you prefer Fahrenheit to Celsius, simply change the `unit=metric` part of the url to `units=imperial`.



Let's write one last function in our `views.py` file to replicate this call for our visitor's city which we previously displayed.

First we need to add an import for the Python `os` (operating system) module so that we can access our environment variables. At the top of `views.py` add:

```
1 import os
```

Now we can write the function. Add the following to `views.py`:

```
1 def get_weather_from_location(city, country_code):
2     token = os.environ.get("OPEN_WEATHER_TOKEN")
3     url = "https://api.openweathermap.org/data/2.5/weather?q={},{}&units=metric&appi\
4 d={}".format(
5     city, country_code, token)
6     response = requests.get(url)
7     return response.json()
```

In line 2, we get our API key from the environment variables (note, you sometimes need to refresh the repl.it page with your repl in to properly load in the environment variables), and we then use this to format our URL properly in line 3. We get the response from OpenWeatherMap and return it as json.

We can now use this function in our `get_weather_from_ip()` function by modifying it to look as follows:

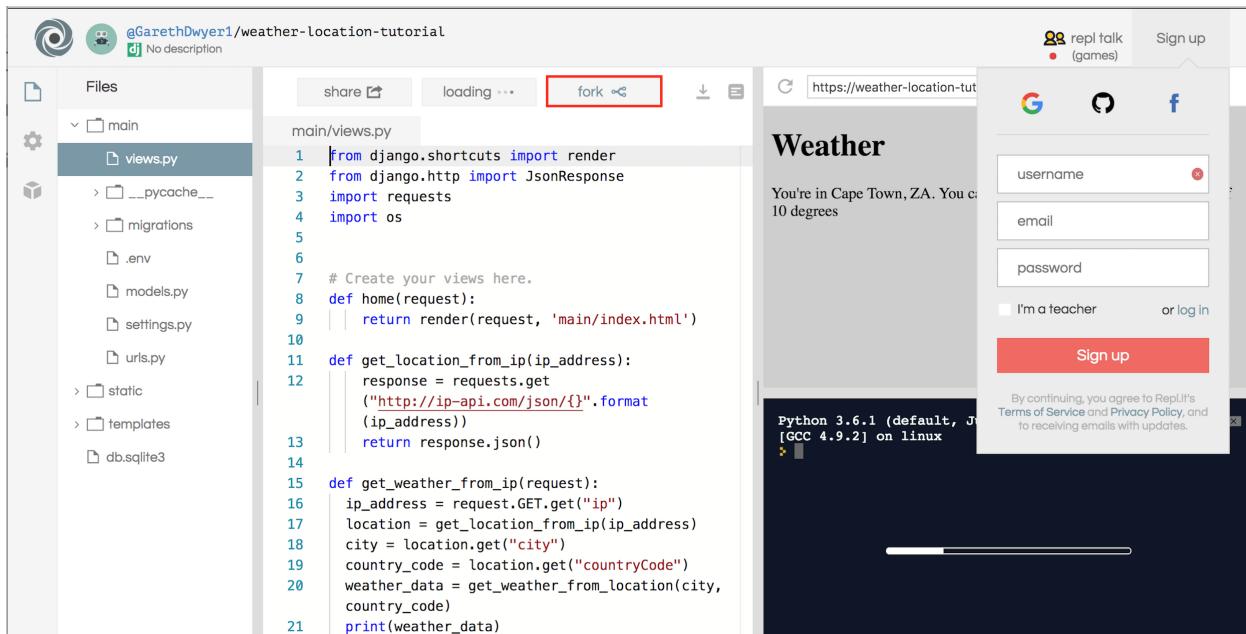
```

1 def get_weather_from_ip(request):
2     ip_address = request.GET.get("ip")
3     location = get_location_from_ip(ip_address)
4     city = location.get("city")
5     country_code = location.get("countryCode")
6     weather_data = get_weather_from_location(city, country_code)
7     description = weather_data['weather'][0]['description']
8     temperature = weather_data['main']['temp']
9     s = "You're in {}, {}. You can expect {} with a temperature of {} degrees".format(\n
10    city, country_code, description, temperature)
11    data = {"weather_data": s}
12    return JsonResponse(data)

```

We now get the weather data in line 6, parse this into a description and temperature in lines 7 and 8, and add this to the string we pass back to our template in line 9. If you reload the page, you should see your location and your weather.

and hit the “Fork” button. If you didn’t create an account at the beginning of this tutorial, you’ll be prompted to create one again. (You can even use a lot of Repl functionality without creating an account.)



Forking a Repl

If you’re stuck for ideas, some possible extensions are:

- Make the page look nicer by using [Bootstrap<sup>146</sup>](#) or another CSS framework in your template files.

<sup>146</sup><https://getbootstrap.com/>

- Make the app more customizable by allowing the user to choose their own location if the IP location that we guess is wrong
- Make the app more useful by showing the weather forecast along with the current weather. (This data is [also available<sup>147</sup>](#) from Open Weather Map).
- Add other location-related data to the web app such as news, currency conversion, translation, postal codes. See [---

<sup>147</sup><https://openweathermap.org/forecast5>](https://github.com/toddmotto/public-apis#geocoding<sup>148</sup></a> for a nice list of possibilities.</li></ul></div><div data-bbox=)

<sup>148</sup><https://github.com/toddmotto/public-apis#geocoding>

# Building a CRM app with NodeJS, Repl.it, and MongoDB

In this tutorial, we'll use NodeJS on Repl.it, along with a MongoDB database to build a basic [CRUD<sup>149</sup>](#) (Create, Read, Update, Delete) [CRM<sup>150</sup>](#) (Customer Relationship Management) application. A CRM lets you store information about customers to help you track the status of every customer relationship. This can help businesses keep track of their clients and ultimately increase sales. The application will be able to store and edit customer details, as well as keep notes about them.

This tutorial won't be covering the basics of Node.js, but each line of code will be explained in detail.

## Overview and requirements

All of the code will be written and hosted in Repl.it, so you won't need to install any additional software on your computer. In this tutorial, we'll be covering:

- Creating an account on [MongoDB Atlas<sup>151</sup>](#)
- Connecting our database to our Repl
- Creating a user interface to insert customer data
- Updating and deleting database entries

By the end of the tutorial, the application you will have created will be able to create, update, and delete documents in a MongoDB database. You will also have used a web application framework called [Express<sup>152</sup>](#) and the [Pug<sup>153</sup>](#) templating engine.

## Creating an account on MongoDB Atlas

MongoDB Atlas is a fully managed Database-as-a-Service. It provides a document database (often referred to as NoSQL), as opposed to a more traditional relational database like PostgreSQL.

Head over to [MongoDB Atlas<sup>154</sup>](#) and hit the “Try free” button. You should then sign up, clicking the “Get started free” button to complete the process.

---

<sup>149</sup>[https://en.wikipedia.org/wiki/Create,\\_read,\\_update\\_and\\_delete](https://en.wikipedia.org/wiki/Create,_read,_update_and_delete)

<sup>150</sup>[https://en.wikipedia.org/wiki/Customer\\_relationship\\_management](https://en.wikipedia.org/wiki/Customer_relationship_management)

<sup>151</sup><https://www.mongodb.com/cloud/atlas>

<sup>152</sup><https://expressjs.com/>

<sup>153</sup><https://pugjs.org/api/getting-started.html>

<sup>154</sup><https://www.mongodb.com/cloud/atlas>

After signing up, under “Shared Clusters”, press the “Create a Cluster” button.

You now have to select a provider and a region. For the purposes of this tutorial, we chose Google Cloud Platform as the provider and Iowa (us-central1) as the region, although it should work regardless of the provider and region.

### Cluster Region

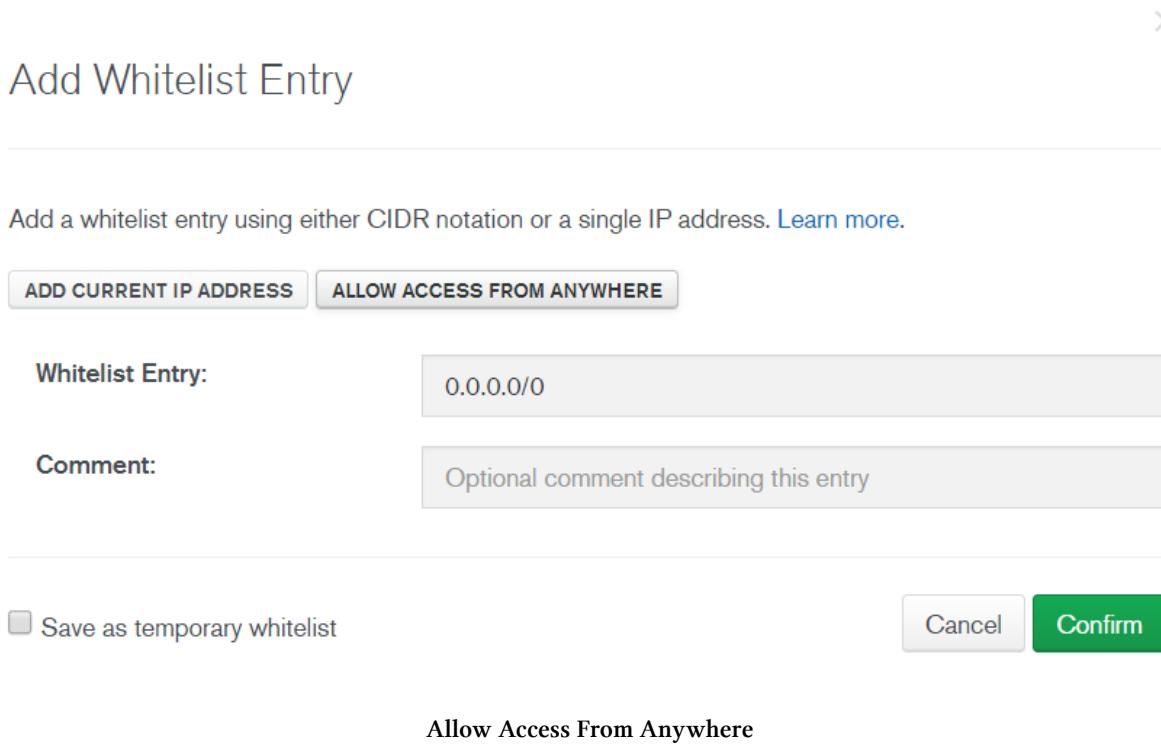
Under “Cluster Name” you can change the name of your cluster. Note that you can only change the name now - it can’t be changed once the cluster is created. After you’ve done that, click “Create Cluster”.

### Cluster Name

After a bit of time, your cluster will be created. Once it’s available, click on “Database Access” under the Security heading in the left-hand column and then click “Add New Database User”. You need a

database user to actually store and retrieve data. Enter a username and password for the user and make a note of those details - you'll need them later. Select "Read and write to any database" as the user privilege. Hit "Add User" to complete this step.

Next, you need to allow network access to the database. Click on "Network Access" in the left-hand column, and "Add IP Address". Because we won't have a static IP from Repl.it, we're just going to allow access from anywhere - don't worry, the database is still secured with the username and password you created earlier. In the popup, click "Allow Access From Anywhere" and then "Confirm".



Now select "Clusters", under "Data Storage" in the left-hand column. Click on "Connect" and select "Connect Your Application". This will change the pop-up view. Copy the "Connection String" as you will need it shortly to connect to your database from Repl.it. It will look something like this:

```
mongodb+srv://<username>:<password>@cluster0-zrtwi.gcp.mongodb.net/test?retryWrites=true&w=majority
```

## Creating a Repl and connecting to our Database

First, we need to create a new Node.js Repl to write the code necessary to connect to our shiny new Database. Navigate to repl.it and create a new Repl, selecting "Node.js" as the language.

A great thing about Repl is that it makes projects public by default. This makes it easy to share and is great for collaboration and learning, but we have to be careful not to make our database credentials available on the open Internet.

To solve this problem, we'll be using environment variables, as we have done in previous tutorials. We'll create a special file that Repl.it recognizes and keeps private for you, and in that file we declare variables that become part of our Repl.it development environment and are accessible in our code.

In your Repl, create a file called `.env` by selecting “Files” in the left-hand pane and then clicking the “Add File” button. Note that the spelling has to be exact or the file will not be recognized. Add your MongoDB database username and password (not your login details to MongoDB Atlas) into the file in the below format:

```
1 MONGO_USERNAME=username  
2 MONGO_PASSWORD=password
```

- Replace `username` and `password` with your database username and password
- Spacing matters. Make sure that you don't add any spaces before or after the `=` sign

Now that we have credentials set up for the database, we can move on to connecting to it in our code.

MongoDB is kind enough to provide a client that we can use. To test out our database connection, we're going to insert some customer data into our database. In your `index.js` file (created automatically and found under the Files pane), add the following code:

```
1 const MongoClient = require('mongodb').MongoClient;  
2 const mongo_username = process.env.MONGO_USERNAME  
3 const mongo_password = process.env.MONGO_PASSWORD  
4  
5 const uri = `mongodb+srv://${mongo_username}:${mongo_password}@cluster0-zrtwi.gcp.mo\n`  
ngodb.net/crmdb?retryWrites=true&w=majority`;  
6  
7 const client = new MongoClient(uri, { useNewUrlParser: true });
```

Let's break this down to see what is going on and what we still need to change:

- Line 1 adds the dependency for the MongoDB Client. As we have discussed before, Repl.it makes things easy by installing all the dependencies for us, so we don't have to use something like npm to do it manually.
- Line 2 & 3 we retrieve our MongoDB username and password from the environment variables that we set up earlier.
- Line 5 has a few very important details that we need to get right.
  - Replace the section between the @ and the next / with the same section of your connection string from MongoDB that we copied earlier. You may notice the  `${mongo_username}` and  `${mongo_password}` before and after the colon near the beginning of the string. These are called Template Literals. Template Literals allow us to put variables in a string, which Node.js will then helpfully replace with the actual values of the variables.

- Note crmdb after the / and before the ?. This will be the name of the database that we will be using. MongoDB creates the database if it doesn't exist for us. You can change this to whatever you want to name the database, but remember what you changed it to for future sections of this tutorial.
- Line 6 creates the client that we will use to connect to the database.

## Making a user interface to insert customer data

We're going to make an HTML form that will capture the customer data and send it to our Repl.it code, which will then insert it into our database.

In order to actually present and handle an HTML form, we need a way to process HTTP GET and POST requests. The easiest way to do this is to use a web application framework. A web application framework is designed to support the development of web applications - it gives you a standard way to build your application and lets you get to building your application fast without having to do the boilerplate code.

A really simple, fast and flexible Node.js web application framework is [Express<sup>155</sup>](#), which provides a robust set of features for the development of web applications.

The first thing we need to do is add the dependencies we need. Right at the top of your index.js file (above the MongoDB code), add the following lines:

```
1 let express = require('express');
2 let app = express();
3 let bodyParser = require('body-parser');
4 let http = require('http').Server(app);
5
6 app.use(bodyParser.json())
7 app.use(bodyParser.urlencoded({ extended: true }));
```

Let's break this down.

- Line 1 adds the dependency for Express. Repl.it will take care of installing it for us.
- Line 2 creates a new Express app that will be needed to handle incoming requests.
- Line 3 adds a dependency for 'body-parser'. This is needed for the Express server to be able to handle the data that the form will send, and give it to us in a useful format to use in the code.
- Line 4 adds a dependency for a basic HTTP server.
- Line 6 & 7 tell the Express app which parsers to use on incoming data. This is needed to handle form data.

Next, we need to add a way for the Express to handle an incoming request and give us the form that we want. Add the following lines of code below that which you just added:

---

<sup>155</sup><https://expressjs.com/>

```
1 app.get('/', function (req, res) {  
2   res.sendFile('/index.html', {root:'.'});  
3 });  
4  
5 app.get('/create', function (req, res) {  
6   res.sendFile('/create.html', {root:'.'});  
7 });
```

- `app.get` tells Express that we want it to handle a GET request.
- `'/'` tells Express that it should respond to GET requests sent to the root URL. A root URL looks something like ‘`https://crm.hawkiesza.repl.co`’ - note that there are no slashes after the URL.
- `'/create'` tells Express that it should respond to GET requests to `/create` after the root URL i.e. ‘`https://crm.hawkiesza.repl.co/create`’
- `res.sendFile` tells Express to send the given file as a response.

Before the server will start receiving requests and sending responses, we need to tell it to run. Add the following code below the previous line.

```
1 app.set('port', process.env.PORT || 5000);  
2 http.listen(app.get('port'), function() {  
3   console.log('listening on port', app.get('port'));  
4 });
```

- **Line 1** tells Express to set the port number to either a number defined as an environment variable, or 5000 if no definition was made.
- **Line 2-4** tells the server to start listening for requests.

Now we have an Express server listening for requests, but we haven’t yet built the form that it needs to send back if it receives a request.

Make a new file called `index.html` and paste the following code into it:

```
1 <!DOCTYPE html>
2 <html>
3 <body>
4 <form action="/create" method="GET">
5   <input type="submit" value="Create">
6 </form>
7
8 </body>
9 </html>
```

This is just a simple bit of HTML that puts a single button on the page. When this button is clicked it sends a GET request to /create, which the server will then respond to according to the code that we wrote above - in our case it will send back the `create.html` file which we will define now.

Make a new file called `create.html` and paste the following into it:

```
1 <!DOCTYPE html>
2 <html>
3 <body>
4
5 <h2>Customer details</h2>
6
7 <form action="/create" method="POST">
8   <label for="name" >Customer name *</label><br>
9   <input type="text" id="name" name="name" class="textInput" placeholder="John Smith\"
10 " required>
11   <br>
12   <label for="address" >Customer address *</label><br>
13   <input type="text" name="address" class="textInput" placeholder="42 Wallaby Way, S\
14 ydney" required>
15   <br>
16   <label for="telephone" >Customer telephone *</label><br>
17   <input type="text" name="telephone" class="textInput" placeholder="+275554202" req\
18 uired>
19   <br>
20   <label for="note" >Customer note</label><br>
21   <input type="text" name="note" class="textInput" placeholder="Needs a new pair of \
22 shoes">
23   <br><br>
24   <input type="submit" value="Submit">
25 </form>
26
27 </body>
28 </html>
```

We won't go in-depth into the above HTML. It is a very basic form with 4 fields (name, address, telephone, note) and a Submit button, which creates an interface that will look like the one below.

---

## Customer details

Customer name \*

John Smith

Customer address \*

42 Wallaby Way, Sydney

Customer telephone \*

+275554202

Customer note

Needs a new pair of shoes

Submit

### Customer Details

When the user presses the submit button a POST request is made to /create with the data in the form - we still have to handle this request in our code as we're currently only handling a GET request to /.

If you now start up your application (click the “run” button) a new window should appear on the right that displays the “create” button we defined just now in “create.html”. You can also navigate to [https://<repl\\_name>.your\\_username.repl.co](https://<repl_name>.your_username.repl.co) (replace <repl\_name> with whatever you named your Repl (but with no underscores or spaces) and <your\_username> with your Repl username) to see the form. You will be able to see this URL in your Repl itself.

If you select “create” and then fill in the form and hit submit, you’ll get a response back that says Cannot POST /create. This is because we haven’t added the code that handles the form POST request, so let’s do that.

Add the following code into your `index.js` file, below the `app.get` entry that we made above.

```
1 app.post('/create', function (req, res, next) {
2   client.connect(err => {
3     const customers = client.db("crmdb").collection("customers");
4
5     let customer = { name: req.body.name, address: req.body.address, telephone: req.\n6 body.telephone, note: req.body.note };
7     customers.insertOne(customer, function(err, res) {
8       if (err) throw err;
9       console.log("1 customer inserted");
10      });
11    })
12    res.send('Customer created');
13  })
```

- Line 1 defines a new route that listens for an HTTP ‘POST’ request at `/create`.
- Line 2 connects to the database. This happens asynchronously, so we define a callback function that will be called once the connection is made.
- Line 3 creates a new collection of customers. Collections in MongoDB are similar to Tables in SQL.
- Line 5 defines customer data that will be inserted into the collection. This is taken from the incoming request. The form data is parsed using the parsers that we defined earlier and is then placed in the `req.body` variable for us to use in the code.
- Line 6 inserts the customer data into the collection. This also happens asynchronously, and so we define another callback function that will get an error if an error occurred, or the response if everything happened successfully.
- Line 7 throws an error if the above insert had a problem.
- Line 8 gives us some feedback that the insert happened successfully.

If you now run the Repl (you may need to refresh it) and submit the filled-in form, you’ll get a message back that says “Customer created”. If you then go and look in your cluster in MongoDB and select the “collections” button, you’ll see a document has been created with the details that we submitted in the form.

## Updating and deleting database entries

As a final step in this tutorial, we want to be able to update and delete database documents in our collection. To make things simpler, we're going to make a new HTML page where we can request a document and then update or delete it.

First, let's make the routes to our new page. In your `index.js`, add the following code below the rest of your routing code (ie. before the MongoDB code):

```

1 app.get('/get', function (req, res) {
2   res.sendFile('/get.html', {root: '.'});
3 });
4
5 app.get('/get-client', function (req, res) {
6   client.connect(err => {
7     client.db("crmdb").collection("customers").findOne({name: req.query.name}, f\ 
8 unction(err, result) {
9     if (err) throw err;
10    res.render('update', {oldname: result.name, oldaddress: result.address, ol\ 
11 dtelephone: result.telephone, oldnote: result.note, name: result.name, address: resu\ 
12 lt.address, telephone: result.telephone, note: result.note});
13  });
14 });
15 });

```

- Line 1-3 as before, this tells Express to respond to incoming GET requests on `/get` by sending the `get.html` file which we will define below.
- Line 5-12 this tells Express to respond to incoming GET requests on `/get-client`.
  - Line 7 makes a call to the database to fetch a customer by name. If there are more than 1 with the same name, then the first one found will be returned.
  - Line 9 tells Express to render the `update` template, replacing variables with the given values as it goes. Important to note here is that we are also replacing values in the hidden form fields we created earlier with the current values of the customer details. This is to ensure that we update or delete the correct customer.

In your `index.html` file, add the following code after the `</form>` tag:

```
1 <br>
2 <form action="/get" method="GET">
3   <input type="submit" value="Update/Delete">
4 </form>
```

This adds a new button that will make a GET request to /get, which will then return get.html.

Create

Update/Delete

## Index

Make a new file called get.html with the following contents:

```
1 <!DOCTYPE html>
2 <html>
3 <body>
4   <form action="/get-client" method="GET">
5     <label for="name" >Customer name *</label><br>
6     <input type="text" id="name" name="name" class="textInput" placeholder="John Smi\
7 th" required>
8     <input type="submit" value="Get customer">
9   </form>
10 </body>
11 </html>
```

This makes a simple form with an input for the customer's name and a button.

---

Customer name \*

#### Get Customer

Clicking this button will then make a GET call to `/get-client` which will respond with the client details where we will be able to update or delete them.

To actually see the customer details on a form after requesting them, we need a templating engine to render them onto the HTML page and send the rendered page back to us. With a templating engine, you define a template - a page with variables in it - and then give it the values you want to fill into the variables. In our case, we're going to request the customer details from the database and tell the templating engine to render them onto the page.

We're going to use a templating engine called [Pug<sup>156</sup>](#). Pug is a simple templating engine that integrates fully with Express. The syntax that Pug uses is very similar to HTML. One important difference in the syntax is that spacing is very important as it determines your parent/child hierarchy.

First, we need to tell Express which templating engine to use and where to find our templates. Put the following line above your route definitions (i.e. after the other app. lines in `index.js`):

```
1 app.engine('pug', require('pug').__express)
2 app.set('views', '.')
3 app.set('view engine', 'pug')
```

Now create a new file called `update.pug` with the following content:

---

<sup>156</sup><https://pugjs.org/api/getting-started.html>

```

1  html
2    body
3      p #{message}
4      h2= 'Customer details'
5      form(method='POST' action='/update')
6          input(type='hidden' id='oldname' name='oldname' value=oldname)
7          input(type='hidden' id='oldaddress' name='oldaddress' value=oldaddress)
8          input(type='hidden' id='oldtelephone' name='oldtelephone' value=oldtelephone)
9          input(type='hidden' id='oldnote' name='oldnote' value=oldnote)
10         label(for='name') Customer name:
11         br
12         input(type='text', placeholder='John Smith' name='name' value=name)
13         br
14         label(for='address') Customer address:
15         br
16         input(type='text', placeholder='42 Wallaby Way, Sydney' name='address' value=a\
17 ddress)
18         br
19         label(for='telephone') Customer telephone:
20         br
21         input(type='text', placeholder='+275554202' name='telephone' value=telephone)
22         br
23         label(for='note') Customer note:
24         br
25         input(type='text', placeholder='Likes unicorns' name='note' value=note)
26         br
27         button(type='submit' formaction="/update") Update
28         button(type='submit' formaction="/delete") Delete

```

This is very similar to the HTML form we created previously for `create.html`, however this is written in the Pug templating language. We're creating a hidden element to store the "old" name, telephone, address, and note of the customer - this is for when we want to do an update.

Using the old details to update the customer is an easy solution, but not the best solution as it makes the query cumbersome and slow. If you add extra fields into your database you would have to remember to update your query as well, otherwise it could lead to updating or deleting the wrong customer if they have the same information. A better, but more complicated way is to use the unique ID of the database document as that will only ever refer to one customer.

We have also put in placeholder variables for name, address, telephone, and note, and we have given the form 2 buttons with different actions.

If you now run the code, you will have an index page with 2 buttons. Pressing the 'Update/Delete' button will take you to a new page that asks for a Customer name. Filling the customer name and

pressing ‘Get customer’ will, after a little time, load a page with the customer’s details and 2 buttons below that say ‘Update’ and ‘Delete’. Make sure you enter a customer name you have entered before.

## Customer details

Customer name:

Customer address:

Customer telephone:

Customer note:

### Update-Delete

Our next step is to add the ‘Update’ and ‘Delete’ functionality. Add the following code below your routes in `index.js`:

```

1 app.post('/update', function(req, res) {
2   client.connect(err => {
3     if (err) throw err;
4     let query = { name: req.body.oldname, address: req.body.oldaddress, telephone: r\
5 eq.body.oldtelephone, note: req.body.oldnote };
6     let newvalues = { $set: {name: req.body.name, address: req.body.address, telepho\
7 ne: req.body.telephone, note: req.body.note } };
8     client.db("crmdb").collection("customers").updateOne(query, newvalues, function(\
9 err, result) {
10       if (err) throw err;
11       console.log("1 document updated");
12       res.render('update', {message: 'Customer updated!', oldname: req.body.name, \
13 oldaddress: req.body.address, oldtelephone: req.body.telephone, oldnote: req.body.no\
14 te, name: req.body.name, address: req.body.address, telephone: req.body.telephone, n\
15 ote: req.body.note});
16     });
17   });
18 })
19
20 app.post('/delete', function(req, res) {

```

```
21 client.connect(err => {
22   if (err) throw err;
23   let query = { name: req.body.name, address: req.body.address ? req.body.address \ 
24 : null, telephone: req.body.telephone ? req.body.telephone : null, note: req.body.no\ 
25 te ? req.body.note : null };
26   client.db("crmdb").collection("customers").deleteOne(query, function(err, obj) {
27     if (err) throw err;
28     console.log("1 document deleted");
29     res.send(`Customer ${req.body.name} deleted`);
30   });
31 });
32 })
```

This introduces 2 new ‘POST’ handlers - one for /update, and one for /delete.

- Line 2 connects to our MongoDB database.
- Line 3 throws an error if there was a problem connecting to the database.
- Line 4 defines a query that we will use to find the document to update. In this case, we are using the details of the customer *before* it was updated. We saved this name earlier in a hidden field in the HTML. Trying to find the customer by its updated name obviously won’t work because it hasn’t been updated yet. Also, note that we are setting some of the fields to null if they are empty. This is so that the database returns the correct document when we update or delete - if we search for a document that has no address with an address of “” (empty string), then our query won’t return anything.
- Line 5 defines the new values that we want to update our customer with.
- Line 6 updates the customer with the new values using the query
- Line 7 throws an error if there was a problem with the update.
- Line 8 logs that a document was updated.
- Line 9 re-renders the update page with a message saying that the customer was updated, and displays the new values.
- Line 15 connects to our MongoDB database.
- Line 16 throws an error if there was a problem connecting to the database.
- Line 17 defines a query that we will use to find the document to delete. In this case we are using all the details of the customer *before* any changes were made on the form to make sure we delete that specific customer.
- Line 18 we connect to the database and delete the customer.
- Line 19 throws an error if there was a problem with the delete.
- Line 20 logs that a document was deleted.
- Line 21 sends a response to say that the customer was deleted.

## Putting it all together

If you run your application now, you'll be able to create, update, and delete documents in a MongoDB database. This is a very basic CRUD application, with a very basic and unstyled UI, but it should give you the foundation to build much more sophisticated applications.

Some ideas for this are:

\*You could add fields to the database to classify customers according to which stage they are in your sales [pipeline<sup>157</sup>](#) so that you can track if a customer is potentially stuck somewhere and contact them to re-engage.

\*You could then integrate some basic marketing automation with a page allowing you to send an email or SMS to customers (though don't spam clients!).

\*You could also add fields to keep track of customer purchasing information so that you can see which products do well with which customers.

If you want to start from where this tutorial leaves off, fork the Repl at [In the next chapter, we'll be introducing using machine learning to classify text in Repl.it.](https://repl.it/@GarethDwyer1/nodejs-crm<sup>158</sup>.</a></p></div><div data-bbox=)

---

<sup>157</sup><https://www.bitrix24.com/glossary/what-is-pipeline-management-definition-crm.php>

<sup>158</sup><https://repl.it/@GarethDwyer1/nodejs-crm>

# Introduction to Machine Learning with Python and Repl.it

In this tutorial, we're going to walk through how to set up a basic Python Repl<sup>159</sup> that can learn the difference between two categories of sentences, positive and negative. For example, if you had the sentence "I love it!", we want to train a machine to know that this sentence is associated with happy and positive emotions. If we have a sentence like "it was really terrible", we want the machine to label it as a negative or sad sentence.

The maths, specifically calculus and linear algebra, behind machine learning gets a bit hairy. We'll be abstracting this away with the Python library scikit-learn<sup>160</sup>, which makes it possible to do advanced machine learning in a few lines of Python.

At the end of this tutorial, you'll understand the fundamental ideas of automatic classification and have a program that can learn by itself to distinguish between different categories of text. You'll be able to use the same code to learn new categories (e.g. spam/not-spam, or clickbait/non-clickbait).

## Overview and requirements

To follow along with this tutorial, you should have at least basic knowledge of Python or a similar programming language. Ideally, you should also sign up for a Repl.it<sup>161</sup> account so that you can modify and extend the bot we build, but it's not completely necessary.

In this tutorial, we will:

- Create some simple mock data - text to classify as positive or negative
- Explain vectorisation of the dataset
- Cover how to classify text using a machine learning classifier
- Compare this to a manual classifier

## Setting up

Create a new Python Repl and open the `main.py` file that Repl created for you automatically. Add the following two imports to the top and run the Repl so that these dependencies are installed.

---

<sup>159</sup><https://repl.it>

<sup>160</sup><https://scikit-learn.org/>

<sup>161</sup><https://repl.it>

```
1 from sklearn import tree
2 from sklearn.feature_extraction.text import CountVectorizer
```

In line 1, we import the `tree` module, which will give us a Decision Tree classifier that can learn from data. In line 2, we import a vectoriser – something that can turn text into numbers. We'll describe each of these in more detail soon!

Throughout the next steps, you can hit the big green “run” button to run your code, check for bugs, and view output along the way (you should do this every time you add new code).

## Creating some mock data

Before we get started with the exciting part, we'll create a very simple dataset – too simple in fact. You might not see the full power of machine learning at first as our task will look so easy, but once we've walked through the concepts, it'll be a simple matter of swapping the data out for something bigger and more complicated.

On the next lines of `main.py` add the following lines of code.

```
1 positive_texts = [
2     "we love you",
3     "they love us",
4     "you are good",
5     "he is good",
6     "they love mary"
7 ]
8
9 negative_texts = [
10    "we hate you",
11    "they hate us",
12    "you are bad",
13    "he is bad",
14    "we hate mary"
15 ]
16
17 test_texts = [
18    "they love mary",
19    "they are good",
20    "why do you hate mary",
21    "they are almost always good",
22    "we are very bad"
23 ]
```

We've created three simple datasets of five sentences each. The first one contains positive sentences; the second one contains negative sentences; and the last contains a mix of positive and negative sentences.

It's immediately obvious to a human which sentences are positive and which are negative, but can we teach a computer to tell them apart?

We'll use the two lists `positive_texts` and `negative_texts` to *train* our model. That is, we'll show these examples to the computer along with the correct answers for the question "is this text positive or negative?". The computer will try to find rules to tell the difference, and then we'll test how well it did by giving it `test_texts` without the answers and ask it to guess whether each example is positive or negative.

## Understanding vectorization

The first step in nearly all machine learning problems is to translate your data from a format that makes sense to a human to one that makes sense to a computer. In the case of language and text data, a simple but effective way to do this is to associate each unique word in the dataset with a number, from 0 onwards. Each text can then be represented by an array of numbers, representing how often each possible word appears in the text.

Let's go through an example to see how this works. If we had the two sentences

```
["nice pizza is nice"], ["what is pizza"]
```

then we would have a dataset with four unique words in it. The first thing we'd want to do is create a vocabulary mapping to map each unique word to a unique number. We could do this as follows:

```
1 {
2     "nice": 0,
3     "pizza": 1,
4     "is": 2,
5     "what": 3
6 }
```

To create this, we simply go through both sentences from left to right, mapping each new word to the next available number and skipping words that we've seen before. We can now convert our sentences into bag of words vectors as follows, where we indicate the frequency of occurrence of each of the words in our vocabulary:

```

1  [
2      [2, 1, 1, 0], # two "nice", one "pizza", one "is", zero "what"
3      [0, 1, 1, 1] # zero "nice", one "pizza", one "is", one "what"
4  ]

```

Each sentence vector is always the same length as the *total* vocabulary size. We have four words in total (across all of our sentences), so each sentence is represented by an array of length four. Each position in the array represents a word, and each value represents how often that word appears in that sentence.

The first sentence contains the word “nice” twice, while the second sentence does not contain the word “nice” at all. According to our mapping, the zeroth element of each array should indicate how often the word nice appears, so the first sentence contains a 2 in the beginning and the second sentence contains a 0 there.

This representation is called “bag of words” because we lose all of the information represented by the *order* of words. We don’t know, for example, that the first sentence starts and ends with “nice”, only that it contains the word “nice” twice.

With real data, these arrays get *very* long. There are millions of words in most languages, so for a big dataset containing most words, each sentence needs to be represented by a very long array, where nearly all values are set to zero (all the words not in that sentence). This could take up a lot of space, but luckily scikit-learn uses a clever sparse-matrix implementation to overcome this. This doesn’t quite look like the above, but the overall concept remains the same.

Let’s see how to achieve the above using scikit-learn’s optimised vectoriser.

First we want to combine all of our “training” data (the data that we’ll show the computer along with the correct labels of “positive” or “negative” so that it can learn), so we’ll combine our positive and negative texts into one array. Add the following code below the datasets you created.

```

1 training_texts = negative_texts + positive_texts
2 training_labels = ["negative"] * len(negative_texts) + ["positive"] * len(positive_t\
3 exts)

```

Our dataset now looks like this:

```

1 ['we hate you', 'they hate us', 'you are bad', 'he is bad', 'we hate mary', 'we love\
2 you', 'they love us', 'you are good', 'he is good', 'they love mary']
3 ['negative', 'negative', 'negative', 'negative', 'negative', 'positive', 'positive', \
4 'positive', 'positive', 'positive']

```

The two arrays (texts and labels) are associated by index. The first text in the first array is negative, and corresponds to the first label in the second array, and so on.

Now we need a vectoriser to transform the texts into numbers. We can create one in scikit-learn with

```
1 vectorizer = CountVectorizer()
```

Before we can use our vectorizer, it needs to run once through all the data we have so it can build the mapping from words to indices. This is referred to as “fitting” the vectoriser, and we can do it like this:

```
1 vectorizer.fit(training_texts)
```

If we want, we can see the mapping it created (which might not be in order, as in the examples we walked through earlier, but each word will have its own index). We can inspect the vectoriser’s vocabulary by adding the line

```
1 print(vectorizer.vocabulary_)
```

(Note the underscore at the end. Scikit-learn uses this as a convention for “helper” attributes. The mapping is explicit only for debugging purposes and you shouldn’t need to use it in most cases). My vocabulary mapping looked as follows:

```
1 {'we': 10, 'hate': 3, 'you': 11, 'they': 8, 'us': 9, 'are': 0, 'bad': 1, 'he': 4, 'i\\
2 s': 5, 'mary': 7, 'love': 6, 'good': 2}
```

Behind the scenes, the vectoriser inspected all of our texts, did same basic preprocessing like making everything lowercase, split the text into words using a built-in *tokenization* method, and produced a vocabulary mapping specific to our dataset.

Now that we have a vectorizer that knows what words are in our dataset, we can use it to transform our texts into vectors. Add the following lines of code to your Repl:

```
1 training_vectors = vectorizer.transform(training_texts)
2 testing_vectors = vectorizer.transform(test_texts)
```

The first line creates a list of vectors which represent all of the training texts, still in the same order, but now each text is a vector of numbers instead of a string.

The second line does the same with the test vectors. The machine learning part isn’t looking at our test texts (that would be cheating) – it’s just mapping the words to numbers so that it can work with them more easily. Note that when we called `fit()` on the vectoriser, we only showed it the training texts. Because there are words in the test texts that don’t appear in the training texts, these words will simply be ignored and will not be represented in `testing_vectors`.

Now that we have a vectorised representation our problem, let’s take a look at how we can solve it.

## Understanding classification

A classifier is a statistical model that tries to predict a label for a given input. In our case, the input is the text and the output is either “positive” or “negative”, depending on whether the classifier thinks that the input is positive or negative.

A machine learning classifier can be “trained”. We give it labelled data and it tries to learn rules based on that data. Every time it gets more data, it updates its rules slightly to account for the new information. There are many kinds of classifiers, but one of the simplest is called a Decision Tree.

Decision trees learn a set of yes/no rules by building decisions into a tree structure. Each new input moves down the tree, while various questions are asked one by one. When the input filters all the way to a leaf node in the tree, it acquires a label.

If that’s confusing, don’t worry! We’ll walk through a detailed example with a picture soon to clarify. First, let’s show how to get some results using Python.

Add the following lines to `main.py`:

```
1 classifier = tree.DecisionTreeClassifier()
2 classifier.fit(training_vectors, training_labels)
3 predictions = classifier.predict(testing_vectors)
4 print(predictions)
```

Similarly to the vectoriser, we first create a classifier by using the module we imported at the start. Then we call `fit()` on the classifier and pass in our training vectors and their associated labels. The decision tree is going to look at both and attempt to learn rules that separate the two kinds of data.

Once our classifier is trained, we can call the `predict()` method and pass in previously unseen data. Here we pass in `testing_vectors` which is the list of vectorized test data that the computer didn’t look at during training. It has to try and apply the rules it learned from the training data to this new “unseen” data. Machine learning is pretty cool, but it’s not magic, so there’s no guarantee that the rules we learned will be any good yet.

The code above produces the following output:

```
1 ['positive' 'positive' 'negative' 'positive' 'negative']
```

Let’s take a look at our test texts again to see if these predictions match up to reality.

```
1 "they love mary"
2 "they are good"
3 "why do you hate mary"
4 "they are almost always good"
5 "we are very bad"
```

The output maps to the input by index, so the first output label (“positive”) matches up to the first input text (“they love mary”), and the last output label (“negative”) matches up to the last input text (“we are very bad”).

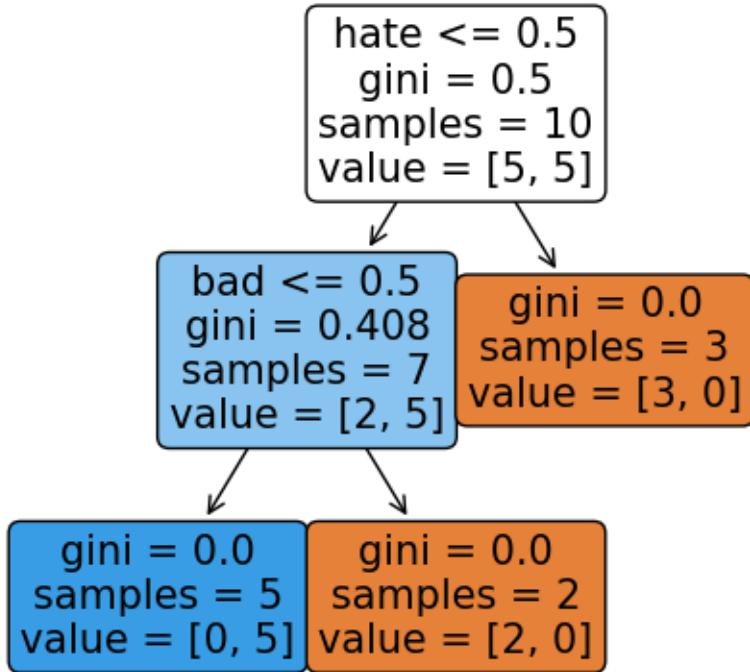
It looks like the computer got every example right! It’s not a difficult problem to solve. The words “bad” and “hate” appear only in the negative texts and the words “good” and “love”, only in the positive ones. Other words like “they”, “mary”, “you” and “we” appear in both good and bad texts. If our model did well, it will have learned to ignore the words that appear in both kinds of texts, and focus on “good”, “bad”, “love” and “hate”.

Decision Trees are not the most powerful machine learning model, but they have one advantage over most other algorithms: after we have trained them, we can look inside and see exactly how they work. More advanced models like deep neural networks are nearly impossible to make sense of after training.

The Scikit-learn tree module contains a useful function to assist in visualising trees. Add the following code to the end of your Repl:

```
1 import matplotlib.pyplot as plt
2 fig = plt.figure(figsize=(5,5))
3 tree.plot_tree(classifier, feature_names = vectorizer.get_feature_names(), rounded = \
4 True, filled = True)
5 fig.savefig('tree.png')
```

In the left-hand pane, you should see a file called ‘tree.png’. If you open it, your tree graph should looks as follows:



A visualised decision tree

The above shows a decision tree that only learned two rules. The first rule (top square) is about the word “hate”. The rule is “is the number of times ‘hate’ occurs in this sentence less than or equal to 0.5”. None of our sentences contain duplicate words, so each rule will really be only about whether the word appears or not (you can think of the  $\leq 0.5$  rules as  $< 1$  in this case).

For each question in our training dataset, we can ask if the first rule is True or False. If the rule is True for a given sentence, we’ll move that sentence down the tree **left**. If not, we’ll go **right**.

Once we’ve asked this first question for each sentence in our dataset, we’ll have three sentences for which the answer is “False”, because three of our training sentences contain the word “hate”. These three sentences go right in the decision tree and end up at first leaf node (an end node with no arrows coming out the bottom). This leaf node has `value = [3, 0]` in it, which means that three samples reach this node, and three belong to the negative class and zero to the positive class.

For each sentence where the first rule is “True” (the word “hate” appears less than 0.5 times, or in our case 0 times), we go down the left of the tree, to the node where `value = [2, 5]`. This isn’t a leaf node (it has more arrows coming out the bottom), so we’re not done yet. At this point we have two negative sentences and all five positive sentences still.

The next rule is “bad  $\leq 0.5$ ”. In the same way as before, we’ll go down the right path if we have more than 0.5 occurrences of “bad” and left if we have fewer than 0.5 occurrences of “bad”. For the last two negative sentences that we are still evaluating (the two containing “bad”), we’ll go *right* and end up at the node with `value=[2, 0]`. This is another leaf node and when we get here we have two negative sentences and zero positive ones.

All other data will go left, and we’ll end up at `[0, 5]`, or zero negative sentences and five positive ones.

As an exercise, take each of the test sentences (not represented in the annotated tree above) and try to follow the set of rules for each one. If it ends up in a bucket with more negative sentences than positive ones (either of the right branches), it’ll be predicted as a negative sentence. If it ends up in the left-most leaf node, it’ll be predicted as a positive sentence.

## Building a manual classifier

When the task at hand is this simple, it’s often easier to write a couple of rules manually rather than using Machine Learning. For this dataset, we could have achieved the same result by writing the following code.

```

1 def manual_classify(text):
2     if "hate" in text:
3         return "negative"
4     if "bad" in text:
5         return "negative"
6     return "positive"
7
8 predictions = []
9 for text in test_texts:
10    prediction = manual_classify(text)
11    predictions.append(prediction)
12 print(predictions)
```

Here we have replicated the decision tree above. For each sentence, we check if it contains “hate” and if it does we classify it as negative. If it doesn’t, we check if it contains “bad”, and if it does, we classify it as negative. All other sentences are classified as positive.

So what’s the difference between machine learning and traditional rule-based models like this one? The main advantage of learning the rules directly is that it doesn’t really get more difficult as the dataset grows. Our dataset was trivially simple, but a real-world dataset might need thousands or millions of rules, and while we could write a more complicated set of if-statements “by hand”, it’s much easier if we can teach machines to learn these by themselves.

Also, once we've perfected a set of manual rules, they'll still only work for a single dataset. But once we've perfected a machine learning model, we can use it for many tasks, simply by changing the input data!

In the example we walked through, our model was a perfect student and learned to correctly classify all five unseen sentences, this is not usually the case for real-world settings. Because machine learning models are based on probability, the goal is to make them as accurate as possible, but in general you will not get 100% accuracy. Depending on the problem, you might be able to get higher accuracy by hand-crafting rules yourself, so machine learning definitely isn't the correct tool to solve all classification problems.

Try the code on bigger datasets to see how it performs. There is no shortage of interesting data sets to experiment with. For example, you could have a look at positive vs negative movie reviews from IMDB using the dataset [here<sup>162</sup>](#). See if you can load the dataset from there into the classifier we built in this tutorial and compare the results.

You can fork this Repl here: [In the next chapter, we'll be investigating the Quicksort algorithm. Whether you're applying for jobs or just like algorithms, it's useful to understand how sorting works. In real projects, most of the time you'll just call `.sort\(\)`, but here you'll build a sorter from scratch and understand how it works.](https://repl.it/@GarethDwyer1/machine-learning-intro<sup>163</sup></a> to keep hacking on it (it's the same code as we walked through above but with some comments added.) If you prefer, the entire program is shown at the end of this tutorial, so you can copy paste it and work from there.</p>
</div>
<div data-bbox=)

—

```

1  from sklearn import tree
2  from sklearn.feature_extraction.text import CountVectorizer
3  import matplotlib.pyplot as plt
4
5  positive_texts = [
6      "we love you",
7      "they love us",
8      "you are good",
9      "he is good",
10     "they love mary"
11 ]
12
13 negative_texts = [
14     "we hate you",
15     "they hate us",
16     "you are bad",
17     "he is bad",

```

---

<sup>162</sup><https://keras.io/api/datasets/imdb/>

<sup>163</sup><https://repl.it/@GarethDwyer1/machine-learning-intro>

```
18     "we hate mary"
19 ]
20
21 test_texts = [
22     "they love mary",
23     "they are good",
24     "why do you hate mary",
25     "they are almost always good",
26     "we are very bad"
27 ]
28
29 training_texts = negative_texts + positive_texts
30 training_labels = ["negative"] * len(negative_texts) + ["positive"] * len(positive_t\
31 exts)
32
33 vectorizer = CountVectorizer()
34 vectorizer.fit(training_texts)
35 print(vectorizer.vocabulary_)
36
37 training_vectors = vectorizer.transform(training_texts)
38 testing_vectors = vectorizer.transform(test_texts)
39
40 classifier = tree.DecisionTreeClassifier()
41 classifier.fit(training_vectors, training_labels)
42
43 print(classifier.predict(testing_vectors))
44
45 fig = plt.figure(figsize=(5,5))
46 tree.plot_tree(classifier, feature_names = vectorizer.get_feature_names(), rounded = \
47 True, filled = True)
48 fig.savefig('tree.png')
49
50 def manual_classify(text):
51     if "hate" in text:
52         return "negative"
53     if "bad" in text:
54         return "negative"
55     return "positive"
56
57 predictions = []
58 for text in test_texts:
59     prediction = manual_classify(text)
60     predictions.append(prediction)
```

```
61 print(predictions)
```

# Quicksort tutorial: Python implementation with line by line explanation

In this tutorial, we'll be going over the [Quicksort<sup>164</sup>](#) algorithm with a line-by-line explanation. We'll go through how the algorithm works, build it in Repl.it and then time it to see how efficient it is.

## Overview and requirements

We're going to assume that you already know at least something about [sorting algorithms<sup>165</sup>](#), and have been introduced to the idea of Quicksort. By the end of this tutorial, you should have a better understanding of how it works.

We're also going to assume that you've covered some more fundamental computer science concepts, especially [recursion<sup>166</sup>](#), on which Quicksort relies.

To recap, Quicksort is one of the most efficient and most commonly used algorithms to sort a list of numbers. Unlike its competitor, Mergesort, Quicksort can sort a list in place, without the need to create a copy of the list, and therefore saving on memory requirements.

The main intuition behind Quicksort is that if we can efficiently *partition* a list, then we can efficiently sort it. Partitioning a list means that we pick a *pivot* item in the list, and then modify the list to move all items larger than the pivot to the right and all smaller items to the left.

Once the pivot is done, we can do the same operation to the left and right sections of the list recursively until the list is sorted.

Here's a Python implementation of Quicksort. Have a read through it and see if it makes sense. If not, read on below!

---

<sup>164</sup><https://en.wikipedia.org/wiki/Quicksort>

<sup>165</sup>[https://en.wikipedia.org/wiki/Sorting\\_algorithm](https://en.wikipedia.org/wiki/Sorting_algorithm)

<sup>166</sup>[https://en.wikipedia.org/wiki/Recursion#In\\_computer\\_science](https://en.wikipedia.org/wiki/Recursion#In_computer_science)

```

1 def partition(xs, start, end):
2     follower = leader = start
3     while leader < end:
4         if xs[leader] <= xs[end]:
5             xs[follower], xs[leader] = xs[leader], xs[follower]
6             follower += 1
7             leader += 1
8         xs[follower], xs[end] = xs[end], xs[follower]
9     return follower
10
11 def _quicksort(xs, start, end):
12     if start >= end:
13         return
14     p = partition(xs, start, end)
15     _quicksort(xs, start, p-1)
16     _quicksort(xs, p+1, end)
17
18 def quicksort(xs):
19     _quicksort(xs, 0, len(xs)-1)

```

## The Partition algorithm

The idea behind the partition algorithm seems intuitive, but the actual algorithm to do it efficiently is pretty counter-intuitive.

Let's start with the easy part – the idea. We have a list of numbers that isn't sorted. We pick a point in this list, and make sure that all larger numbers are to the *right* of that point and all the smaller numbers are to the *left*. For example, given the random list:

```
1 xs = [8, 4, 2, 2, 1, 7, 10, 5]
```

We could pick the *last* element (5) as the pivot point. We would want the list (after partitioning) to look as follows:

```
1 xs = [4, 2, 2, 1, 5, 7, 10, 8]
```

Note that this list isn't sorted, but it has some interesting properties. Our pivot element, 5, is in the correct place (if we sort the list completely, this element won't move). Also, all the numbers to the left are smaller than 5 and all the numbers to the right are greater.

Because 5 is the in the correct place, we can ignore it after the partition algorithm (we won't need to move it again). This means that if we can sort the two smaller sublists to the left and right of

5() [4, 2, 2, 1] and [7, 10, 8]) then the entire list will be sorted. Any time we can efficiently break a problem into smaller sub-problems, we should think of *recursion* as a tool to solve our main problem. Using recursion, we often don't even have to think about the entire solution. Instead, we define a base case (a list of length 0 or 1 is always sorted), and a way to divide a larger problem into smaller ones (e.g. partitioning a list in two), and almost by magic the problem solves itself!

But we're getting ahead of ourselves a bit. Let's take a look at how to actually implement the partition algorithm on its own, and then we can come back to using it to implement a sorting algorithm.

## A bad partition implementation

You could probably easily write your own `partition` algorithm that gets the correct results without referring to any textbook implementations or thinking about it too much. For example:

```

1 def bad_partition(xs):
2     smaller = []
3     larger = []
4     pivot = xs.pop()
5     for x in xs:
6         if x >= pivot:
7             larger.append(x)
8         else:
9             smaller.append(x)
10    return smaller + [pivot] + larger

```

In this implementation, we set up two temporary lists (`smaller` and `larger`). We then take the `pivot` element as the last element of the list (`pop` takes the last element and removes it from the original `xs` list).

We then consider each element `x` in the list `xs`. The ones that are smaller than the pivot, we store in the `smaller` temporary list, and the others go to the `larger` temporary list. Finally, we combine the two lists with the pivot item in the middle, and we have partitioned our list.

This is much easier to read than the implementation at the start of this post, so why don't we do it like this?

The primary advantage of Quicksort is that it is an *in place* sorting algorithm. Although for the toy examples we're looking at, it might not seem like much of an issue to create a few copies of our list, if you're trying to sort terabytes of data, or if you are trying to sort any amount of data on a very limited computer (e.g a smartwatch), then you don't want to needlessly copy arrays around.

In Computer Science terms, this algorithm has a space-complexity of  $O(2n)$ , where  $n$  is the number of elements in our `xs` array. If we consider our example above of `xs = [8, 4, 2, 2, 1, 7, 10, 5]`, we'll need to store all 8 elements in the original `xs` array as well as three elements ([7, 10, 8]) in the `larger` array and four elements ([4, 2, 2, 1]) in the `smaller` array. This is a waste of space! With some clever tricks, we can do a series of swap operations on the original array and not need to make any copies at all.

## Overview of the actual partition implementation

Let's pull out a few key parts of the good partition function that might be especially confusing before getting into the detailed explanation. Here it is again for reference.

```

1 def partition(xs, start, end):
2     follower = leader = start
3     while leader < end:
4         if xs[leader] <= xs[end]:
5             xs[follower], xs[leader] = xs[leader], xs[follower]
6             follower += 1
7             leader += 1
8     xs[follower], xs[end] = xs[end], xs[follower]
9     return follower

```

In our good `partition` function, you can see that we do some swap operations (lines 5 and 8) on the `xs` that is passed in, but we never allocate any new memory. This means that the storage remains constant to the size of `xs`, or  $O(n)$  in Computer Science terms. That is, this algorithm has *half* the space requirement of the “bad” implementation above, and should therefore allow us to sort lists that are twice the size using the same amount of memory.

The confusing part of this implementation is that although everything is based around our pivot element (the last item of the list in our case), and although the pivot element ends up somewhere in the middle of the list at the end, we don’t actually touch the pivot element *until the very last swap*.

Instead, we have two other counters (`follower` and `leader`) which move around the smaller and bigger numbers in a clever way and implicitly keep track of where the pivot element should end up. We then switch the pivot element into the correct place at the end of the loop (line 8).

The `leader` is just a loop counter. Every iteration it increments by one until it gets to the pivot element (the end of the list). The `follower` is more subtle, and it keeps count of the number of swap iterations we do, moving up the list more slowly than the `leader`, tracking where our pivot element should eventually end up.

The other confusing part of this algorithm is on line 4. We move through the list from left to right. All numbers are currently to the *left* of the pivot but we eventually want the “big” items to end up on the *right*.

Intuitively, you would then expect us to do the swapping action when we find an item that is *larger* than the pivot, but in fact, we do the opposite. When we find items that are *smaller* than the pivot, we swap the `leader` and the `follower`.

You can think of this as pushing the small items further to the left. Because the `leader` is always ahead of the `follower`, when we do a swap, we are swapping a small element with one further left in the list. The `follower` only looks at “big” items (ones that the `leader` has passed over without action), so when we do the swap, we’re swapping a small item (`leader`) with a big one (`follower`), meaning that small items will move towards the left and large ones towards the right.

## Line by line examination of partition

We define `partition` with three arguments, `xs` which is the list we want to sort, `start` which is the index of the first element to consider and `end` which is the index of the last element to consider.

We need to define the `start` and `end` arguments because we won't always be partitioning the entire list. As we work through the sorting algorithm later, we are going to be working on smaller and smaller sublists, but because we don't want to create new copies of the list, we'll be defining these sublists by using indexes to the original list.

In line 2, we start off both of our pointers – `follower`, and `leader` – to be the same as the beginning of the segment of the list that we're interested in. The leader is going to move faster than the follower, so we'll carry on looping until the leader falls off the end of the list segment (`while leader < end`).

We could take any element we want as a pivot element, but for simplicity, we'll just choose the last element. In line 4 then, we compare the `leader` element to the pivot. The leader is going to step through each and every item in our list segment, so this means that when we're done, we'll have compared the partition with every item in the list.

If the `leader` element is smaller or equal to the pivot element, we need to send it further to the left and bring a larger item (tracked by `follower`) further to the right. We do this in lines 4-5, where if we find a case where the `leader` is smaller or equal to the pivot, we swap it with the `follower`. At this point, the follower is pointing at a small item (the one that was `leader` a moment ago), so we increment `follower` by one in order to track the next item instead. This has a side effect of counting how many swaps we do, which incidentally tracks the exact place that our pivot element should eventually end up.

Whether or not we did a swap, we want to consider the next element in relation to our pivot, so in line 7 we increment `leader`.

Once we break out of the loop (line 8), we need to swap the pivot item (still on the end of the list) with the `follower` (which has moved up one for each element that was smaller than the pivot). If this is still confusing, look at our example again:

```
1  xs = [8, 4, 2, 2, 1, 7, 10, 5]
```

In `xs`, there are 4 items that are smaller than the pivot. Every time we find an item that is smaller than the pivot, we increment `follower` by one. This means that at the end of the loop, `follower` will have incremented 4 times and be pointing at index 4 in the original list. By inspection, you can see that this is the correct place for our pivot element (5).

The last thing we do is return the `follower` index, which now points to our pivot element in its *correct* place. We need to return this as it defines the two smaller sub-problems in our partitioned list - we now want to sort `xs[0:4]` (the first 4 items, which form an unsorted list) and the `xs[5:]` (the last 3 items, which form an unsorted list).

```
1 xs = [4, 2, 2, 1, 5, 7, 10, 8]
```

If you want another way to visualise exactly how this works, going over some examples by hand (that is, writing out a short randomly ordered list with a pen and paper, and writing out the new list at each step of the algorithm) is very helpful. You can also watch [this detailed YouTube video<sup>167</sup>](#) where KC Ang demonstrates every step of the algorithm using paper cups in under 5 minutes!

## The Quicksort function

Once we get the partition algorithm right, sorting is easy. We'll define a helper `_quicksort` function first to handle the recursion and then implement a prettier public function after.

```
1 def _quicksort(xs, start, end):
2     if start >= end:
3         return
4     p = partition(xs, start, end)
5     _quicksort(xs, start, p-1)
6     _quicksort(xs, p+1, end)
```

To sort a list, we partition it (line 4), sort the left sublist (line 5: from the start of the original list up to the pivot point), and then sort the right sublist (line 6: from just after the pivot point to the end of the original list). We do this recursively with the `end` boundary moving left, closer to `start`, for the left sublists and the `start` boundary moving right, closer to `end`, for the right sublists. When the `start` and `end` boundaries meet (line 2), we're done!

The first call to Quicksort will always be with the entire list that we want sorted, which means that `0` will be the start of the list and `len(xs)-1` will be the end of the list. We don't want to have to remember to pass these extra arguments in every time we call Quicksort from another program (e.g. in any case where it is not calling itself), so we'll make a prettier wrapper function with these defaults to get the process started.

```
1 def quicksort(xs):
2     return _quicksort(xs, 0, len(xs)-1)
```

Now we, as users of the sorting function, can call `quicksort([4,5,6,2,3,9,10,2,1,5,3,100,23,42,1])`, passing in only the list that we want sorted. This will in turn go and call the `_quicksort` function, which will keep calling itself until the list is sorted.

## Testing our algorithm

We can write some basic driver code to take our newly implemented Quicksort out for a spin. Create a new Python Repl and add the following code to `main.py`. Then insert the code listed at the beginning of this tutorial after the imports.

---

<sup>167</sup>[https://www.youtube.com/watch?v=MZaf\\_9IZCrc](https://www.youtube.com/watch?v=MZaf_9IZCrc)

```

1 from datetime import datetime
2 import random
3
4 # create 100000 random numbers between 1 and 1000
5 xs = [random.randrange(1000) for _ in range(10)]
6
7 # look at the first few and last few
8 print(xs[:10])
9 #apply the algorithm
10 quicksort(xs)
11 # have a look at the results
12 print(xs[:10])

```

If you run this code, you will see the sorted list. This does what we expect, but it doesn't tell us about how efficient Quicksort is - so let's take a closer look. Replace the code in `main.py` with the following, and again add the code listed at the beginning of this tutorial after the imports on line 3.

```

1 from datetime import datetime
2 import random
3
4 # create 100000 random numbers between 1 and 1000
5 xs = [random.randrange(1000) for _ in range(100000)]
6
7 # look at the first few and last few
8 print(xs[:10], xs[-10:])
9
10 # start the clock
11 t1 = datetime.now()
12 quicksort(xs)
13 t2 = datetime.now()
14 print("Sorted list of size {} in {}".format(len(xs), t2 - t1))
15
16 # have a look at the results
17 print(xs[:10], xs[-10:])

```

The code generates a random list of 100 000 numbers and sorts this list in around 5 seconds. You can compare the performance of Quicksort to some other common sorting algorithms using [this Repl](#)<sup>168</sup>.

If you want to try the code from the tutorial out, visit the Repl at <https://repl.it/@GarethDwyer1/quicksort><sup>169</sup>. You'll be able to run the code, see the results, and even fork it to continue developing or testing it on your own.

---

<sup>168</sup><https://repl.it/@GarethDwyer1/sorting>

<sup>169</sup><https://repl.it/@GarethDwyer1/quicksort?language=python3>

If you need help, the folk over at the [Repl discord server<sup>170</sup>](#) are very friendly and keen to help people learn.

---

<sup>170</sup><https://repl.it/discord>

# Closing note

We have now come to the end of the series of tutorials. You have learnt the basics of the Repl.it IDE, worked with more advanced features and gone through a number of practical projects. This doesn't mean the end of fun projects, for you should now be equipped to tackle your own projects, which you can start from scratch or use the code from the tutorials as a basis.