Intro to Python

Paul Bradshaw

This week:

- Why Python? Use cases
- Intro to Colab and Python notebooks
- How to store information: variables

What about you?

- What coding experience do you have?
- What do you want/expect to use Python for?

You are planning a story...

1. ...on zero hours contracts with agencies. How can you establish just **how many agency jobs are advertised as full time**?
2. ...**on discrimination against people on benefits in the rental sector** - but how do you establish how big the problem is?
3. ...on **FOI rejections** and you want to put out a story within 30 minutes of a new dataset coming out - how?

Why Python for journalism?

- Repeat analysis - or conduct in advance
- Scraping - or querying APIs
- Transparency, clarity (internal + external)
- Use JSON, XML, SQL, text, PDFs, ‘big data’

The INDEPENDENT

parental control and Digital Vault

NEWS | OPINION | SPORT | OLYMPICS | LIFE | PROPERTY | ARTS & ENTS | TRAVEL | MO

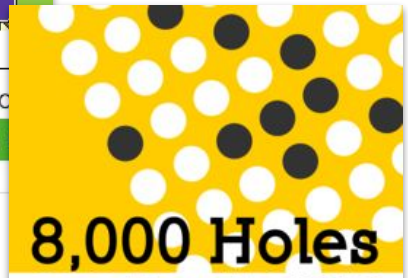
Athletics | Cricket | Football | Golf | Motor Racing | Olympics | Racing | Rugby League | Rugby Union | Sailing

Hot Topics | Euro 2012 | Wimbledon | Olympics

Olympics Sport > Olympics

Seb Coe promised an 'uplifting torch relay to inspire a generation'. So are these really the rol models to do it?

The intention was that 8,000 local heroes with tales to motivate young people wou carry the flame around Britain. But the 2012 sponsors had their own ideas. Tom Peck reports



No DSS: Most flat shares refuse benefit claimants

9 March 2017 | Comments

Coronavirus: Prisoners' children 'forgotten' during pandemic

By Paul Lynch & Paul Bradshaw
BBC Shared Data Unit

30 March

Coronavirus pandemic

Colab notebooks - why

- Run Python inside Google Drive!
- Share your code (including outputs)
- A transparent methodology
- Renders nicely on GitHub

Colab notebooks - how

- Go to https://colab.research.google.com
- Click ‘new notebook’
- Give it a name - but keep **.ipynb**
- Two types of block: code, and text

Text blocks: markdown

- Simple formatting language:
- # Heading 1
- **Italic** ****Bold**** ******Bold and italic******
- [Link](http://link.com)
- (Or just use the formatting buttons)

Text blocks: headings

- Start a text block with a # Heading and it will be added to navigation on left
- Headings are indented, i.e. ### Level 3 headings are indented under ## Level 2 headings

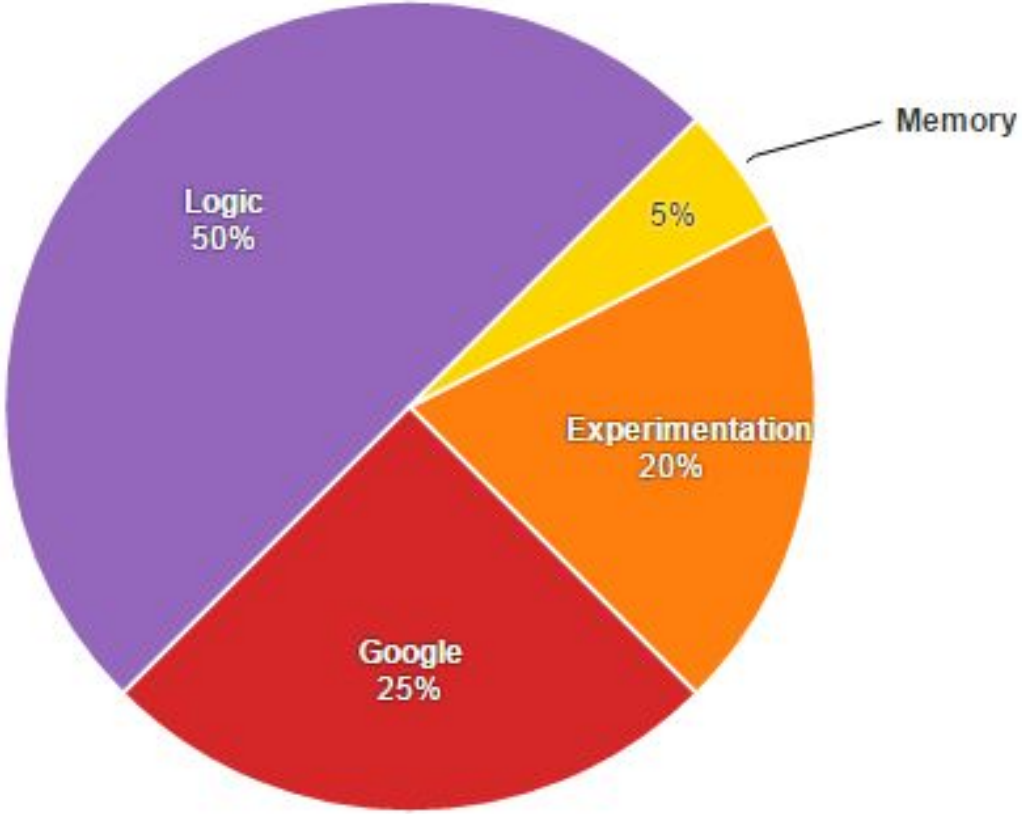
Code blocks

- Write code and run with ▶️ button
- Successful: green tick ✅
- Error: red button and exclamation ⚠️
- Any outputs shown underneath

Code blocks - numbering

- Once a code block has been run, a number will appear next to it, indicating whether it was the 1st code block to be run, or 2nd, 3rd, etc.
- If you get an error, check you've run all the blocks in order

How programming is done



Emma Wedekind 🐞
@EmmaWedekind

Follow

The hardest thing about coding is knowing what to Google 👩‍💻

11:23 AM - 29 Apr 2019

1,538 Retweets **8,282** Likes

218 1.5K 8.3K

Emma Wedekind 🐞
@EmmaWedekind

Follow

The hardest thing about coding is knowing what to Google 🤷‍♀️

11:23 AM - 29 Apr 2019

1,538 Retweets **8,282** Likes

218 1.5K 8.3K

Pete Sherlock
@petesherlock79

Following

Replying to @paulbradshaw

5% memory, 15% googling, 80% ask Paul

Head to the repo...

github.com/paulbradshaw/pythonin12parts

- Click on the part1 folder
- Open 01pythonVariables.ipynb
- (If you have a GitHub account you can copy)

Moving from spreadsheets to Python

This notebook will explain some basic concepts in Python by explaining how common spreadsheet workflows have parallels in coding.

But why do it in Python if you can do it in a spreadsheet? There are two reasons: first this is just a stepping stone to things that you *can't* do in a spreadsheet. And second: using notebooks can actually make your process much more transparent and easy to communicate (and remember) than using a spreadsheet.

Let's begin.

Cell references are like variables

If you've done a calculation in a spreadsheet, chances are you've written a formula like this:

`A1+A2`

In that situation, you are asking the spreadsheet to add together the number in cell A1 with the number in cell A2.

(A very common example of this would be dividing a number by a grand total to calculate a percentage.)

First, of course, the numbers must actually *be* in those cells - and for that to happen, you or someone else would have had to have typed the number into that cell.

In Python, the same principle exists with **variables**.

A variable is created with the equals sign, `=`, like this:

In [2]:
```
A1 = 34
A2 = 101
```

This is called **assigning** a variable.


Variables

Variables

- To store and manipulate information
- Initiate/change with =, e.g. **myage = 28**
- Different types of variable: string, integer, Boolean, list, dictionary

Head to the repo...

github.com/paulbradshaw/pythonin12parts

- Click on the part1 folder
- Open 02textVariables.ipynb

More variables in Python: text and other beasts

In a previous notebook we looked at how to create variables in Python and perform a calculation with those.

Those variables all happened to be numbers. We're going to need them again so let's recreate them in this notebook first:

In [ ]:

```
#store the numbers of requests in a new variable called 'fcorequests'
fcorequests = 48
#store the numbers of refusals in a new variable called 'sec27refusals'
sec27refusals = 28
#calculate a percentage by dividing the part (refusals) by the whole (requests)
#and store in a new variable called 'percrefused'
percrefused = sec27refusals/fcorequests
#print it
print(percrefused)
#multiply by 100 to make it easier to 'read' as a percentage
print(percrefused*100)
```

```
0.5833333333333334
58.333333333333336
```

(A quick recap: these figures are taken from Freedom of Information statistics: April to June 2021 bulletin - the data tables link is here and go to the sheet called '10_Exemptions'.)

Creating text variables: 'strings'

Data doesn't just come in the form of numbers, however. It's almost certain that at some point we're going to need to store some **text**.

For example, we might need to store the recipients of money being spent, in order to calculate which one received the most spending.

In our FOI story we might want to store the names of organisations, or the sections of the FOI Act that are being used as the basis for refusals, or a note from the spreadsheet.

Text variables

- Use single, double or three single quotes
- E.g. "Paul", 'Paul', '''Paul'''
- Allows you to include apostrophes etc.
- If you forget quotes it will think it’s a command (and throw an error)

Curly quote DANGER!

- If you copy quotation marks from docs or slides it may cause an error
- Instead type directly in to ensure you get ‘straight’ quotation marks

Lists and dictionaries

- Lists use [square brackets] and can contain multiple items separated by a comma
- Dictionaries use {curly brackets} and contain key-value pairs separated by a comma
- Used to store data, or loop through data
- More next week!

Key points

- Use Python when you hit the limits of Excel
- Use variables to store info to question
- Coding is problem-solving, not memorising
- Break down problems and use Google

Next...

- Create your own Colab notebook and store 5 different types of variables
- Google how to transform string variables from lower case to upper case, etc.
- Share the notebook with paulonhismobile@gmail.com