AIR QUALITY & PYTHON: DEVELOPING ONLINE ANALYSIS TOOLS
TALK OUTLINE

- Who I am/ what I do
- A case study of using python for science, data analysis & web development
  - Making air quality analysis more accessible for the public
- Quick and easy plots for the public & scientists
- Lessons learnt and future developments
ABOUT ME

- Post-doctoral researcher at the University of Edinburgh
- Background in atmospheric chemistry
- Started off in Fortran with atmospheric model development
- Self-taught Python to analyse the data output from models
- Now working as the research group coder/data wrangler - possibly ‘research data engineer’
A measure of how polluted the air we breathe is

Specifically about pollution with direct health effects (e.g. NO\textsubscript{2}, ozone, particulate matter)

Not CO\textsubscript{2} or CH\textsubscript{4} - these impact climate, not health directly

Generally emitted from traffic but also natural sources (e.g. forest fires)
Asthma deaths rise 25% amid growing air pollution crisis

Doctors urge ministers to act as 1,320 killed by asthma in England and Wales last year

Air pollution blamed after deaths from asthma soar

Air pollution causes nearly 15,000 cases of type 2 diabetes in UK each year, study suggests

Scottish study links air pollution spikes to hospital admissions

Researchers at the University of Dundee studied nearly 15 years of data for air pollution levels in Dundee, Perth and the surrounding area and matched it to medical records of 450 patients who suffer from bronchiectasis, a long-term chronic condition similar to COPD.

Illegal levels of air pollution linked to child’s death

Air pollution costs 200 people in Hounslow their lives each year

Just over 5% of deaths in the borough were directly linked to air pollution.

NEEDS TO BE MONITORED!
AIR QUALITY & PYTHON

**AIR QUALITY ➔ DATA ➔ PRODUCT**

- Numbers from the measurement sites are fairly meaningless
- Currently need to spend time and energy gathering and processing the data
- Daunting to people without the relevant skill set
- Time wasting to those with the relevant skill set
- Not considered by most people - out of sight out of mind

DATA ONLY HAS VALUE WHEN IT'S RELEVANT
(BORROWED FROM A TALK BY ALEXYS JACOB)
WHAT WE NEED . . .

▸ Something to combine data collection, analysis and visualisations
▸ A set of tools that anyone can use
▸ Easily accessible and understandable
▸ Useful for anyone - from school children to academics

THE SOLUTION . . .
### First Things First

**The Data**

| 10.0 | 10.3 | 10.4 | 10.5 | 10.6 | 10.7 | 10.8 | 10.9 | 11.0 | 11.1 | 11.2 | 11.3 | 11.4 | 11.5 | 11.6 | 11.7 | 11.8 | 11.9 | 12.0 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 72   | 69   | 68   | 69   | 70   | 71   | 72   | 73   | 74   | 75   | 76   | 77   | 78   | 79   | 80   | 81   | 82   | 83   | 84   |

| 10.0 | 10.1 | 10.2 | 10.3 | 10.4 | 10.5 | 10.6 | 10.7 | 10.8 | 10.9 | 11.0 | 11.1 | 11.2 | 11.3 | 11.4 | 11.5 | 11.6 | 11.7 | 11.8 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 85   | 86   | 87   | 88   | 89   | 90   | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  | 101  | 102  | 103  |

| 10.0 | 10.1 | 10.2 | 10.3 | 10.4 | 10.5 | 10.6 | 10.7 | 10.8 | 10.9 | 11.0 | 11.1 | 11.2 | 11.3 | 11.4 | 11.5 | 11.6 | 11.7 | 11.8 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 85   | 86   | 87   | 88   | 89   | 90   | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  | 101  | 102  | 103  |
Using data from DEFRA (UK government)

Sites (>150) across the UK taking hourly measurements of various pollutants

Some sites going since 1975

Pretty small data in the grand scheme of things
AIR QUALITY & PYTHON

Nearest to here is by Arthurs Seat

Arthurs Seat

Monitoring site
I need to know information about each and every site (e.g. co-ordinates, life span, pollutants measured)

No quick webpage or file with this information

Time for BeautifulSoup!

A really useful module to help extract data from html

Go through each DEFRA site webpage and get the data I want
GET THE POLLUTION DATA

- All site data available via a URL... if you know the URL
- Simple of task of matching the data you want with the URL
  - You need a site code and a year (site code gathered from site information)
  - e.g. ‘ED3’ & ‘2018’ for Edinburgh 2018
- This data is not in a useful structure
I arrived to pandas quite late

Started as an easy to read a .csv file of the web

A fantastic way to manage a lot of time series data

Filtering and resampling data becomes very quick

Great tutorials and documentation
import plotly.plotly as py
from plotly.graph_objs import *

trace0 = Scatter(
    x=[1, 2, 3, 4],
    y=[10, 15, 13, 17]
)
trace1 = Scatter(
    x=[1, 2, 3, 4],
    y=[16, 5, 11, 9]
)
data = Data([trace0, trace1])

py.plot(data, filename = 'basic-line')
DATA VISUALISATION

▸ Discovered plot.ly for nice graphics

▸ Interactive graphs - e.g. hover data & zoom
PUT IT ONLINE

INTO THE UNKNOWN

PUT IT ONLINE
PUTTING IT ONLINE – LEARNING THE ROPES

- Started out with Django
  - A web framework with a HUGE amount of documentation (a little daunting)
  - Luckily - a lot of tutorials (esp. Django Girls!)
  - Mainly focused on blogs - maybe not ideal for me
HOW IT WORKS

- Creates a number of python files (with basic templates)
- Files include:
  - urls.py - this is lists the website urls that will be visited and calls other modules
  - views.py - this both calls the processing modules and renders the webpage for viewing
  - models.py - this does the hard work, the processing bit.
  - static files - including html & css code
  - + others (including a settings file)
AIR QUALITY & PYTHON

FLOW

HTTP://WWW.UKATMOSPHERE.ORG

URLS.PY

VIEWS.PY

MODELS.PY

HTML & CSS

The website is currently under construction
A WEBSITE IS BORN

UK Atmosphere

Analysis Tools
Use the toolkit to plot air quality data

News
Coming Soon

About
Find more about how to use the tools

Download Tools
Coming Soon
AIR QUALITY & PYTHON

LIMITS

- Django is a great framework
- Not so easy to create multiple instances and interactive pages

PLOT.LY DASH

"Dash is a Python framework for building analytical web applications. No JavaScript required.

Built on top of Plotly.js, React, and Flask, Dash ties modern UI elements like dropdowns, sliders, and graphs to your analytical Python code."
Dash creates “apps” (which could be stand alone websites)

Every time a website is loaded a new app instance is created (eg. one per user)

Each app has a layout which contains the app structure (where the plots go, placement of buttons, dropdown menus etc)

Dash creates “callbacks” which detect a change by the user (by use of Python decorators) and then runs a function to update the page
UKATMOS.ORG

DJANGO WEB FRAMEWORK
NORMAL WEBPAGES GO HERE (E.G. HOMEPAGE)

DASH APP – WHERE ALL THE COOL STUFF HAPPENS

GETS THE DATA
PROCESSES THE DATA
DISPLAYS THE DATA
LETS THE USER CHANGE THE DATA

FOR EXAMPLE...
TOO MUCH DATA - TIME TO USE A DATABASE

- Website was calling .csv files from DEFRA at every request
- Fine for small data (<500 rows)
- The larger the data request the longer it will take…
  Until it crashes!

- A need for better data management - back to Django!
INTEGRATION OF A DATABASE

- Django very useful for SQL database management through Python
- Copy all the data from DEFRA to a new database
- Dash calls a Django model which calls a database (in this case Postgres)
- Allows access of any combination of millions of data points
- No longer relying on DEFRA - but needs constant updates
DEVELOPMENT OF THE ONLINE TOOLS

- Many many bugs fixes to address
- Integration of more data, e.g. European stations, local council stations, satellite data, models.
- Add more types of analysis & plots such as maps
- Get more feedback from users - what is actually useful?
LESSONS LEARNT

▸ Just jump in - you’ll never find the perfect tutorial
▸ Be adaptable
▸ Don’t be scared to make the wrong choice
▸ Take time to learn new things (Pandas!)
▸ Don’t get bogged down by the little things
▸ Keep an eye on the goal
▸ Don’t reinvent the wheel - use others code
▸ Go for a walk
THANKS FOR LISTENING!

@douglasfinch
www.ukatmosphere.org