DOUGLAS FINCH

@douglasfinch

AIR QUALITY & PYTHON: DEVELOPING ONLINE ANALYSIS TOOLS

THE UNIVERSITY of EDINBURGH
School of GeoSciences

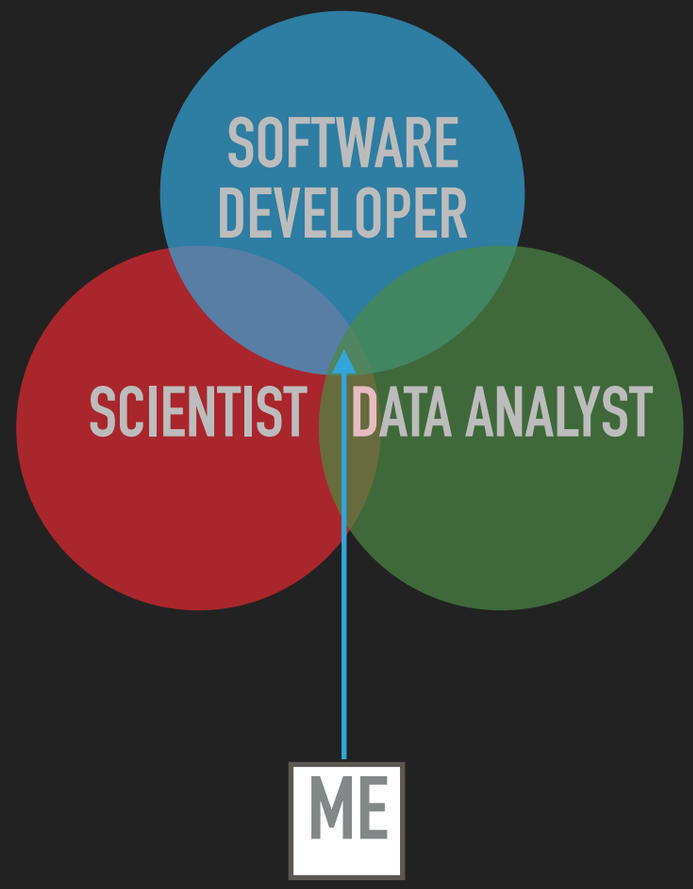
NERC
SCIENCE OF THE ENVIRONMENT

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 BRIEF INTRODUCTION TO AIR QUALITY

- A measure of how polluted the air we breathe is
- Specifically about pollution with direct health effects (eg. NO_2, ozone, particulate matter)
- Not CO_2 or CH_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

▲ Pollution casts a
Nick Ansell/PA

A record num
growing air p

Air pollution blamed after deaths from asthma soar

to clean up our air couldn't be clearer"

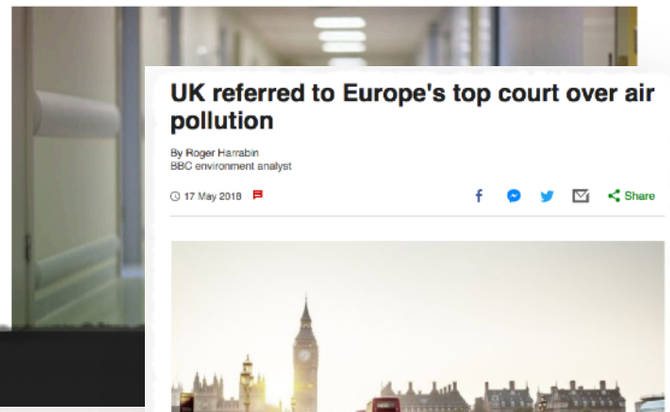
nd Wales have risen by more than 25 per cent in a
g rising air pollution levels.

Scottish study links air pollution spikes to hospital admissions

12.07.2018

Air pollution has been 'clearly linked' to spikes in breathing problem-related admissions to hospitals and visits to GPs, researchers in Scotland have claimed.

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.



UK referred to Europe's top court over air pollution

By Roger Harrabin
BBC environment analyst

17 May 2018 Share

GETTY IMAGES

Persistent high levels of air pollution have seen the UK referred to the European Court of Justice

News | Science

News › Science

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

share Save 5

One in 10 cases of diabetes in Britain is caused by air pollution, say scientists CREDIT: MATT CARDY/GETTY

ror as they
ke from polluted air this summer

ktail of pollution and pollen in London kills people.
ns should spend a night on the wards to see the harm

▲ Air pollution levels in London outstrip l
Rousseau/PA

Illegal levels of air pollution linked to child's death

By Claire Marshall
BBC Environment Correspondent

3 July 2018 Share

Facebook

News › West London News › Hounslow London Borough Council

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 → DATA → PRODUCT

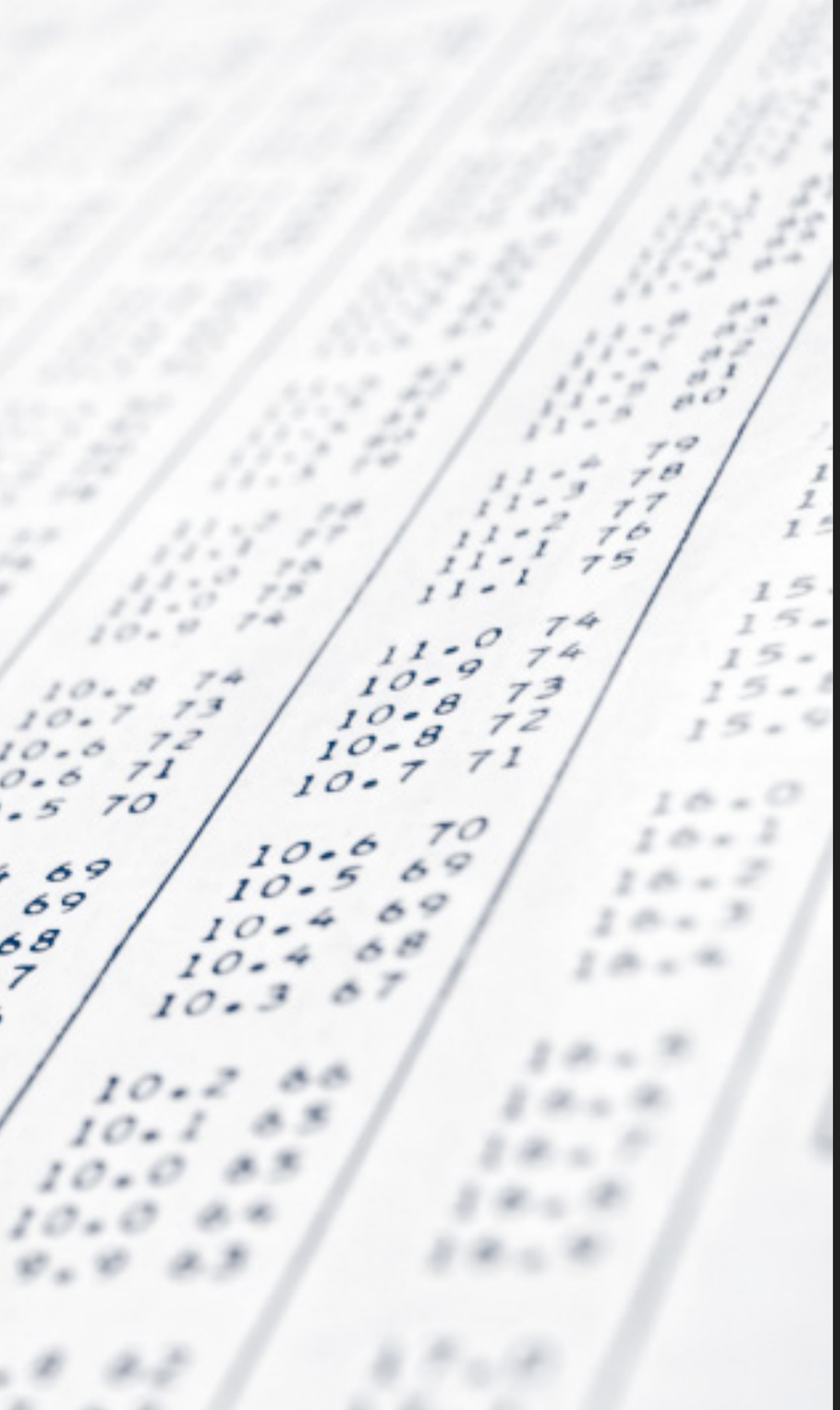
DATA ONLY HAS VALUE WHEN IT'S RELEVANT
(BORROWED FROM A TALK BY ALEXYS JACOB)

- 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

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

DATA COLLECTION

- 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

DATA SCRAPING

- 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

Site Name	Altitude (metres)	EU Site ID	Easting	Environment Type	Government Region	Latitude	Longitude	Northing
Aberdeen	20	GB0729A	394396	Urban Background	North East Scotland	57.15736	-2.094278	807392
Aberdeen Union Street Roadside	26	GB0923A	393656	Urban Traffic	North East Scotland	57.144555	-2.106472	805968
Aberdeen Wellington Road	1.5	GB1057A	394397	Urban Traffic	North East Scotland	57.133888	-2.094198	804779
Armagh Roadside	41	GB0996A	97684	Urban Traffic	Northern Ireland	54.353728	-6.654558	505347
Aston Hill	370	GB0031R	329899	Rural Background	North Wales	52.50385	-3.034178	290053
Auchencorth Moss	260	GB0048R	322166	Rural Background	Central Scotland	55.79216	-3.2429	656128
Ballymena Antrim Road	Not available	GB1074A	125704	Urban Traffic	Northern Ireland	54.851491	-6.274961	559121
Ballymena Ballykeel	59	GB0934A	127317	Urban Background	Northern Ireland	54.861595	-6.250873	560150
Barnsley 12	120	GB0600A	434204	Urban Background	Yorkshire & Humberside	53.55593	-1.485153	406713
Barnsley Gawber	100	GB0681A	432524	Urban Background	Yorkshire & Humberside	53.56292	-1.510436	407478
Barnstaple A39	13	GB1029A	257048	Urban Traffic	South West	51.074793	-4.041924	132591
Bath Roadside	27	GB0647A	375455	Urban Traffic	South West	51.391127	-2.354155	165847
Belfast Centre	10	GB0567A	146338	Urban Background	Northern Ireland	54.59965	-5.928833	529817
Belfast Clara St	20	GB0696A	148441	Suburban Background	Northern Ireland	54.591256	-5.89546	528763
Belfast East	20	GB0514A	148072	Urban Background	Northern Ireland	54.59653	-5.901667	529372
Belfast Stockman's Lane	9	GB1036A	143191	Urban Traffic	Northern Ireland	54.572586	-5.974944	526975
Billingham	10	GB0421A	446928	Urban Industrial	North East	54.60537	-1.275039	523597
Birkenhead Borough Road	7	GB1066A	331926	Urban Traffic	North West & Merseyside	53.388511	-3.025014	388453
Birmingham A4540 Roadside	109	GB1067A	408586	Urban Traffic	West Midlands	52.47609	-1.875024	286470

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

NEXT STEP

ANALYSIS

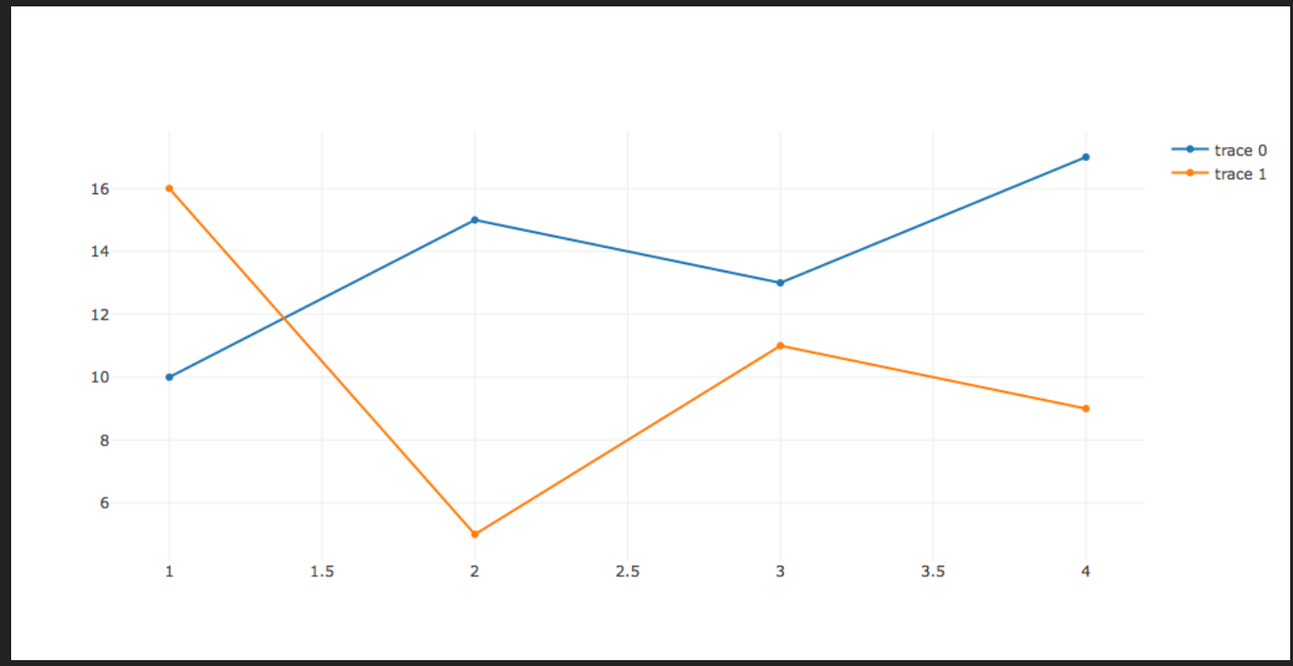
IMPORT PANDAS AS PD

- 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

DATA VISUALISATION

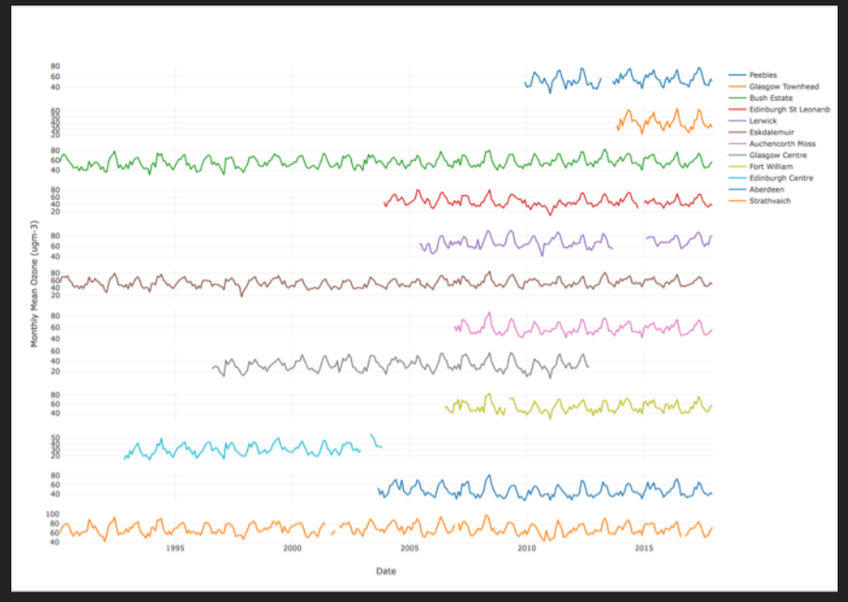
- plot.ly through python

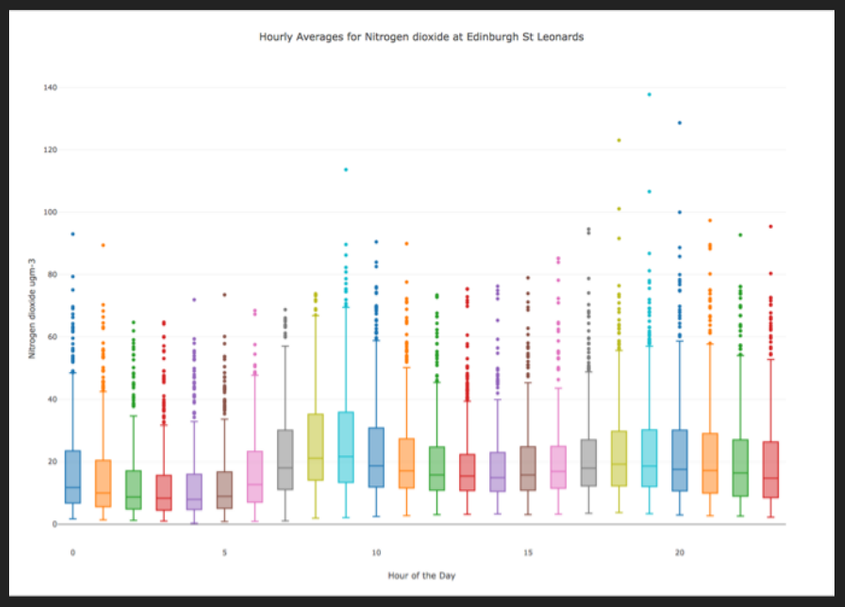
```
import plotly.plotly as py
from plotly.graph_objs import *

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')
```

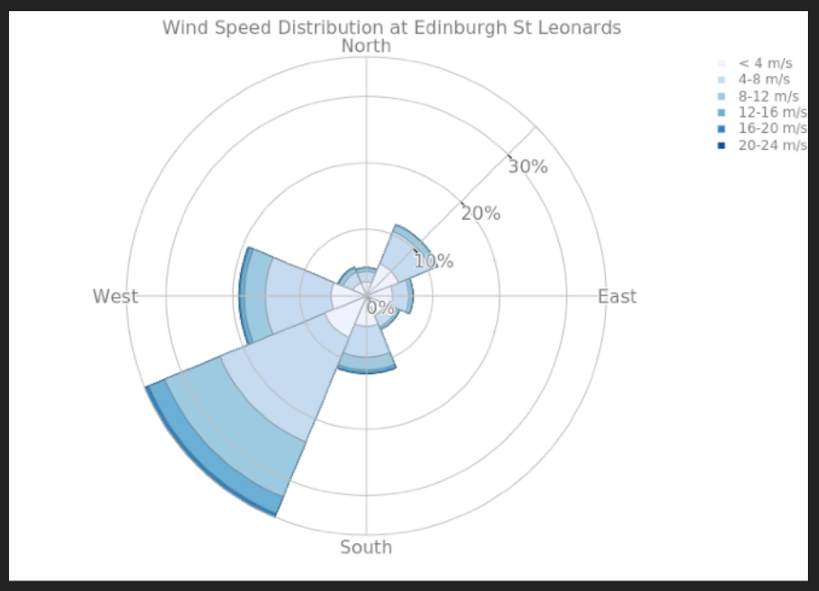


DATA VISUALISATION

- Discovered plot.ly for nice graphics
- Interactive graphs - e.g. hover data & zoom







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)

FLOW

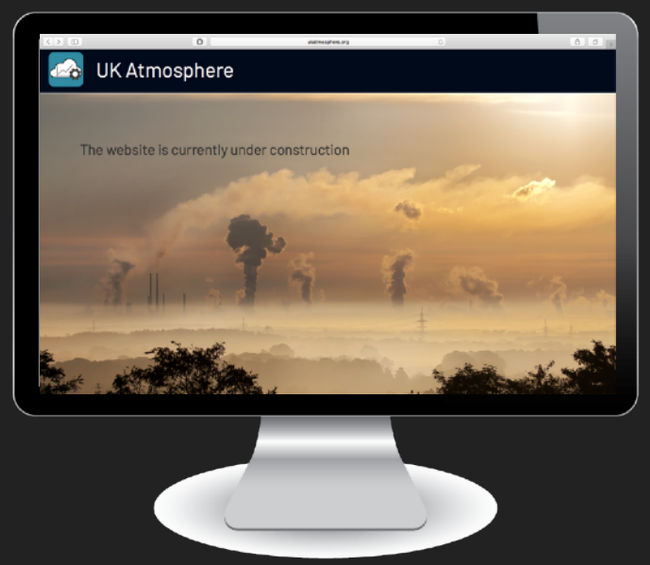
HTTP://WWW.UKATMOSPHERE.ORG

URLS.PY

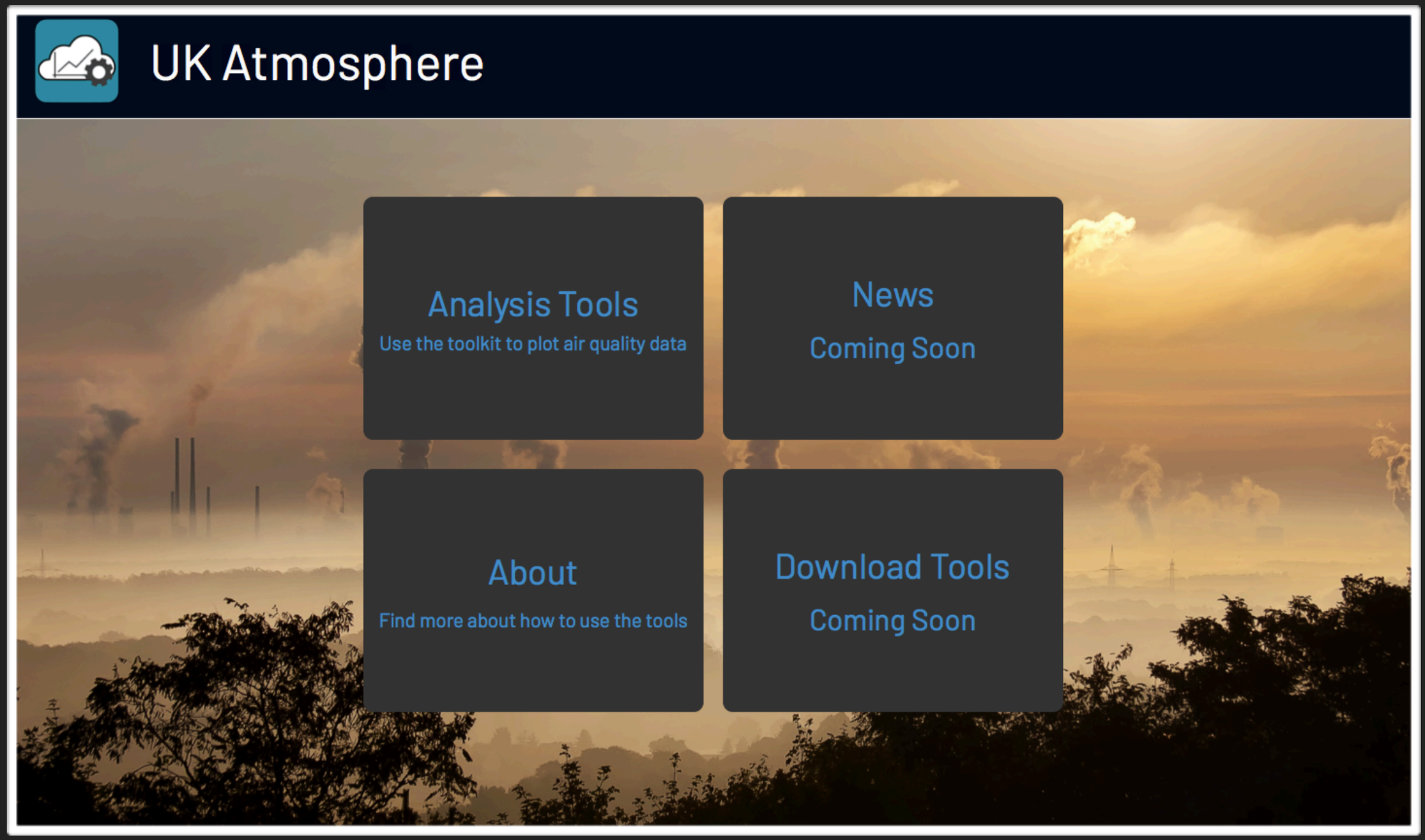
VIEWS.PY

MODELS.PY

HTML & CSS



A WEBSITE IS BORN



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."

PLOT.LY DASH

- 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!

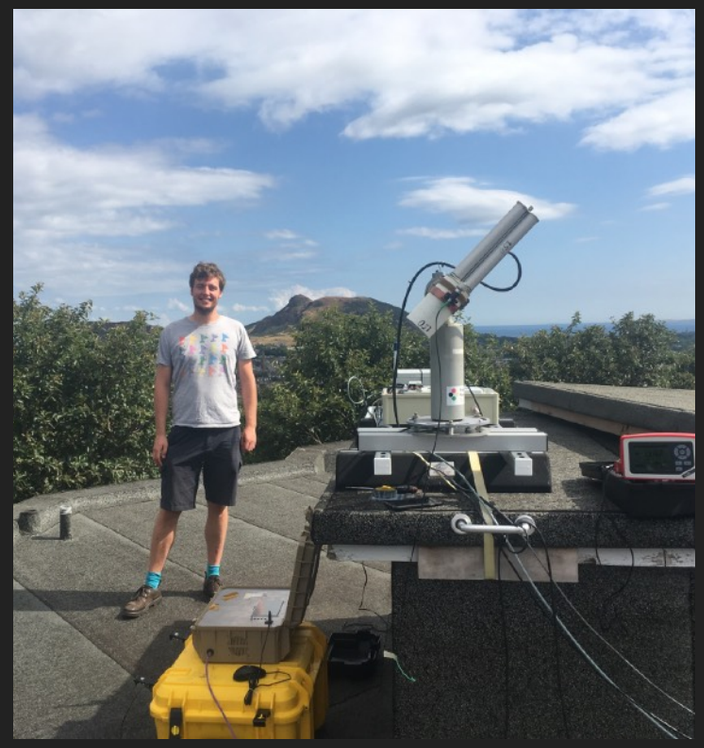
- A need for better data management - back to Django!

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