WACL R Training

Training for air pollution data analysis in R

Will Drysdale and Jack Davison

11th & 12th Nov.

University of York

Welcome!

A course over two afternoons for beginners with R

- Introduction to R, RStudio and Programming for beginners
- Building a script; the benefits of programming over spreadsheets
- Reading, manipulating and visualising data, with tips and tricks to solve common problems
- Chance to practise skills with us on hand to help out

Approaches

- Authentic, live coding
- All course material will be made available
  - This will include all data and script files produced during this course
  - A bespoke self-teaching document will also be made available
  - Useful for post-course learning
- All material used in this course will be **entirely reproducible**
  - This means that you will be able to recreate all the outputs shown during the course (and afterwards)
- Questions are encouraged, and one of us will always be at hand to solve problems

Topics to be covered

Thursday 11th November, 13:00-17:00

- Introduction to R for Air Quality Data
  - Getting familiar with R and RStudio
  - Reading and interrogating data within R
  - Introducing statistical analysis; averages and trend lines
  - Using `openair` for air quality data analysis

Friday 12th November, 13:00-17:00

- Further uses of R in Data Science
  - Reading and combining multiple data streams
  - Further data handling; reshaping, grouping and summarising
  - Making publication standard visualisations with `ggplot2`
  - Real world data project

Who are we?

Jack Davison

I use R for:

- **Big data** analysis - far too big for Excel!
- **Statistical modelling** of data - R makes this easy.
- Developing **reproducible data tools** for others in academia and the private sector.

README.md

`gramsper`: Remote Sensing Emission Factor Development in R

Introduction

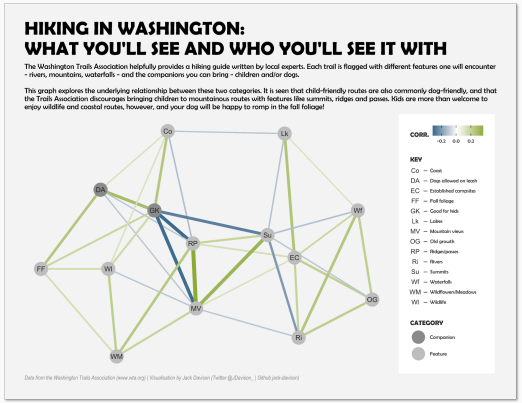
Functions to enhance the use of remote sensing at Ricardo Energy & Environment, particularly aiding with the calculation of emission factors (hence the name `gramsper`!).

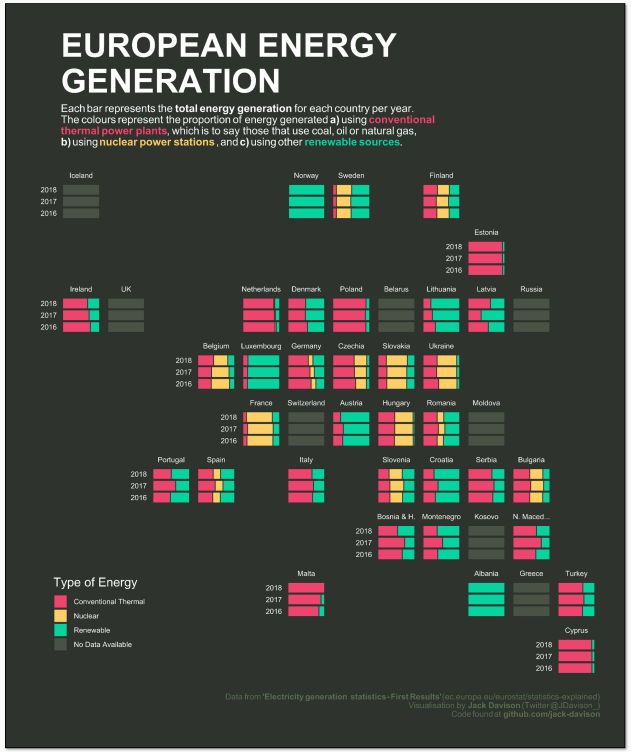
To install, type the below code into your R console:

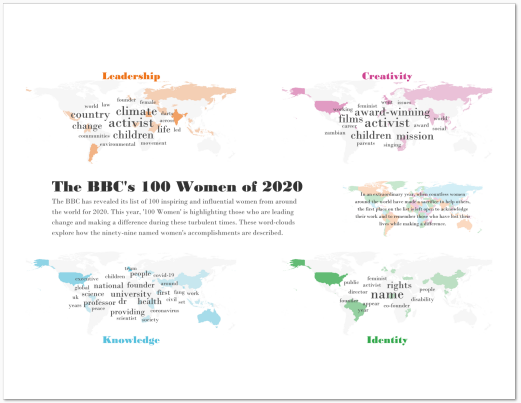
Who are we?

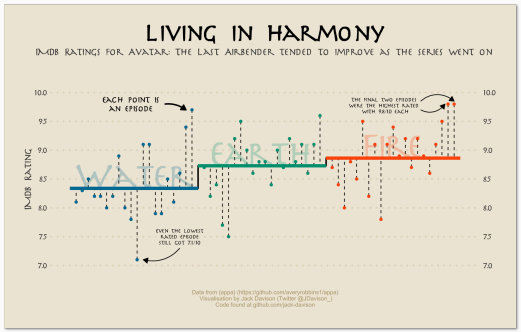
Jack Davison

I also use R extensively for data visualisation!







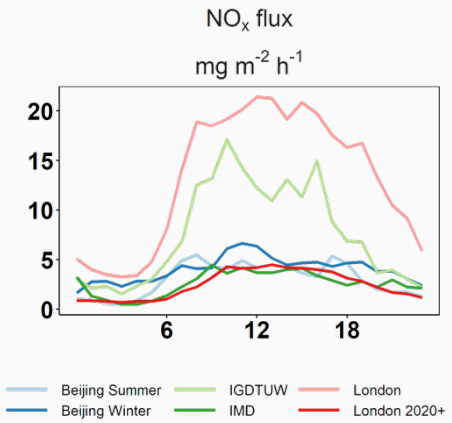


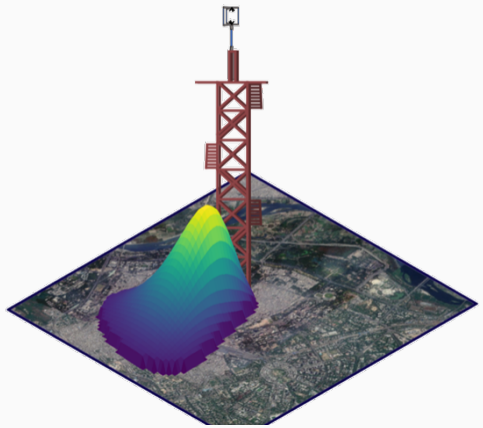
Who are we?

Will Drysdale

I use R for:

- **Eddy Covariance** - processing of high time resolution data (5 - 20 Hz) to calculate emissions using *eddy4R*
  - Perform analysis **automatically** and **reproducibly**
  - **Collaborate** with developers to add our own tools



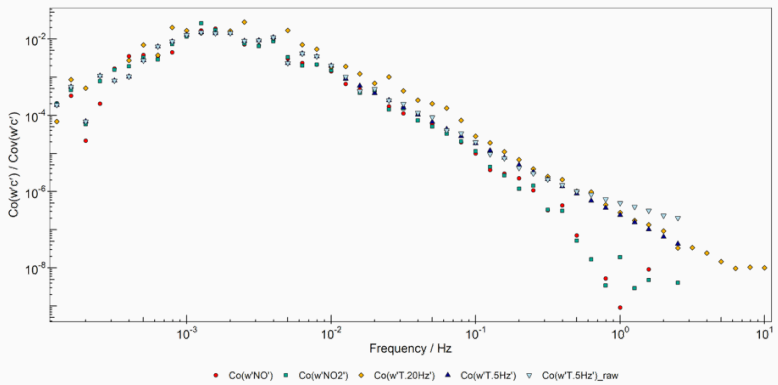


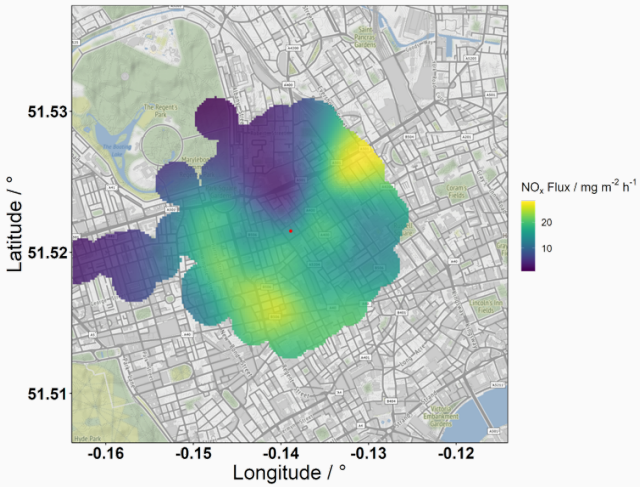
Who are we?

Will Drysdale

I also use R in many other aspects of my work:

- **Instrument data** work up
- Producing **Figures**
- **Mapping** spatial data





Who are you?

Introductions

- What is your name?
- What do you do?
- What kind of data do you use?
  - Big? Small? From the lab? Fieldwork? Modelled? Time-series? Categorical?
- What are you hoping to get out of these sessions?

Further Help

Learning R does not finish at the end of this short course

- There are many R users in WACL who are happy to help, including ourselves.
- There are lots of resources online that we'll point you to.
- WACL has a programming Slack channel for help with R & Python.
- If there is interest, we'll look to do shorter sessions on more specific problems