

# Doing data science at scale in the public sector: the traffic cameras project

Arthur Turrell

Deputy Director, Data Science Campus, ONS

@arthurturrell, @datasciencecampus, @ONS

Innovation & Data4Good,  
October 2021



# Today's key takeaways

- Establishing a moderate level of data science tooling and skills presents big challenges in any large organisation, and the level of capability varies enormously across the public sector
- ONS and Data Science Campus are at heart of efforts to increase data and data science capability across the public sector
- Some of our recent projects demonstrate how to do data science at scale and in production!

Different parts of the public sector  
have different levels of data  
science maturity

# Some illustrative examples of different levels of data science maturity

## Low

- Analysis done only in Excel, and in a manual process
- No 'ground truth' data (data exchanged via email or many copies of data on shared drives)

## Medium

- Access to open source programming languages (eg Python) and their packages (eg pandas); access to modern integrated development environments (eg Visual Studio Code)
- Important data have ground truth versions
- Access to version control

## High

Everything in medium and...

- Access to Unix-based machines
- Access to advanced reproducibility tools like Docker
- Access to cloud computing, for research and production
- Ground truth data can be accessed via APIs

ONS and Data Science Campus are helping to build data science capability across the public sector

# Building capability across the public sector

## Community

[Government Data Science Partnership](#)

[x-Gov data science Slack channel](#)

[Economic data science seminar series](#)

ONS [coffee & coding](#) sessions

[UK Government Data Science GitHub](#)

[Communities of interest meet-ups](#) – mailing list

## Development

### Tools

Advice on data science tools

Integrated Data Service

### Training

[Taught courses and programmes](#) - online

Data Masterclass for senior leaders – with No 10

### Academic Programmes

[Degree Data Science Apprenticeship](#)

[Data Science Graduate Programme](#)

[Masters in Data Science for Government, MDataGov](#)

### Mentoring

Government [Data Science Accelerator](#) (cross-Govt)

International schemes

ONS' data science capability is still developing but *Traffic Cameras* shows how far we've come

# The Traffic Cameras Project

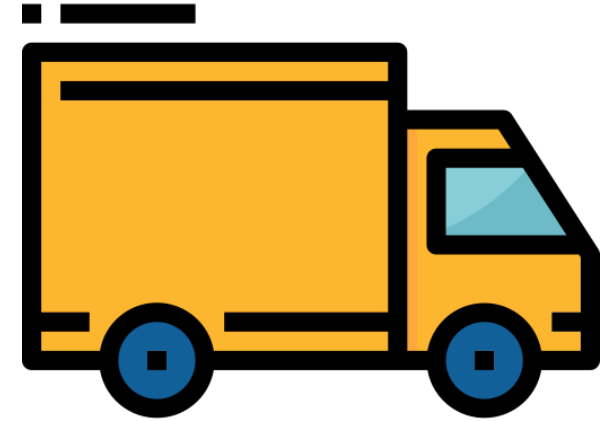
In collaboration with Newcastle University

Chen, Li, Ian Grimstead, Daniel Bell, Joni Karanka, Laura Dimond, Philip James, Luke Smith, and Alistair Edwardes. "Estimating vehicle and pedestrian activity from town and city traffic cameras." *Sensors* **21**, no. 13 (2021): 4564.



# Motivation – traffic cameras project

- **Produce indicators** of ‘busyness’ for urban areas
- Increase **public sector output** through use of under-utilised capital: low-definition traffic cameras
- Faster read on **economically important** traffic disruptions



**68%** of people travelled to work by road in the UK in 2017 (Labour Force Survey)

# The goal: from traffic cameras to time series

## Challenge:

- Pick up image data from 1,500 **publicly accessible** CCTV cameras located in urban centres every 10 minutes, 24/7, and process **the 216,000 images/day** (5GB/day) into **time series** of busyness

## Output:

- Daily count broken down by **vehicle (including type) or pedestrian** and **location**
- Make available at **daily frequency**

## Does not:

- Measure origins or destinations
- Cover highways or non-urban locations
- Provide insight as to *why* journeys are being undertaken



Example snapshot

# Automating the process: data cleaning

Lots of ways data can be problematic...



Artefacts



Missing data

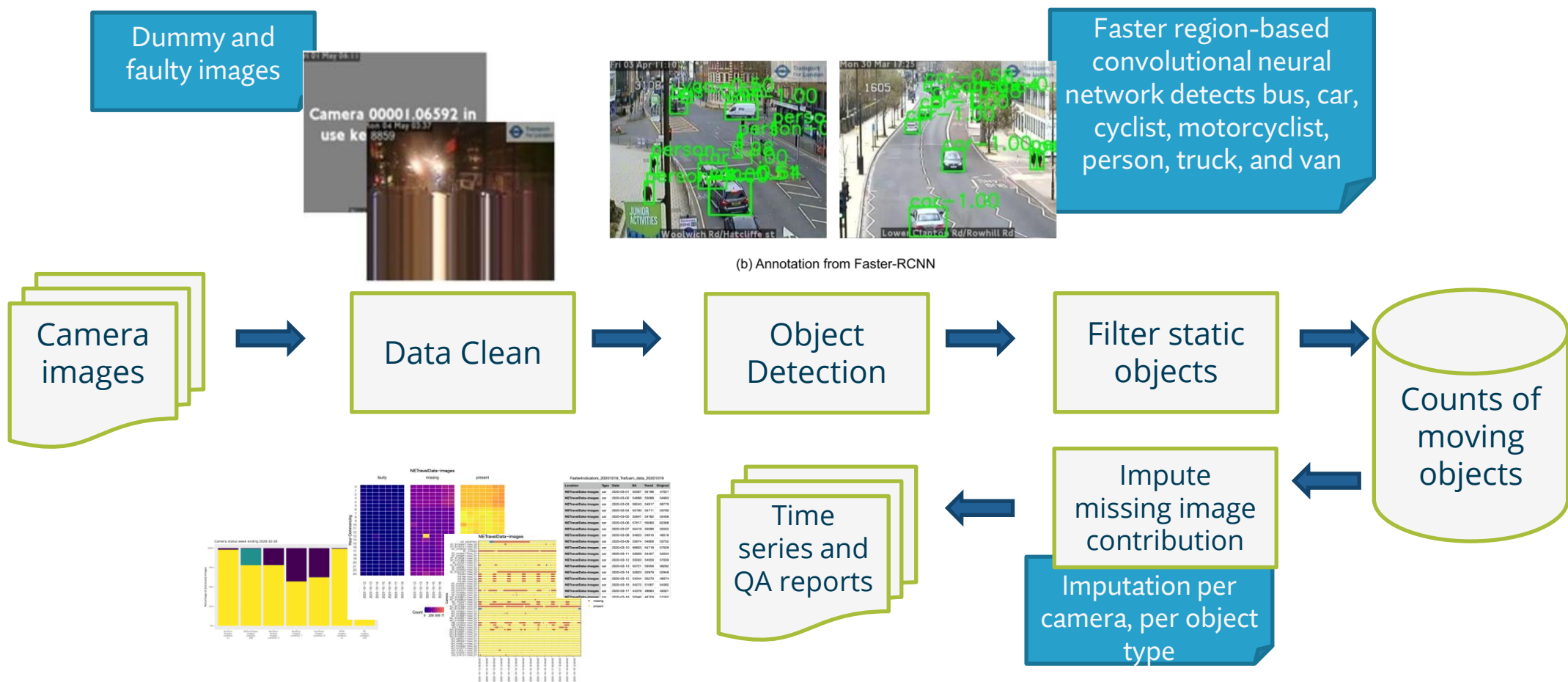


Parked cars



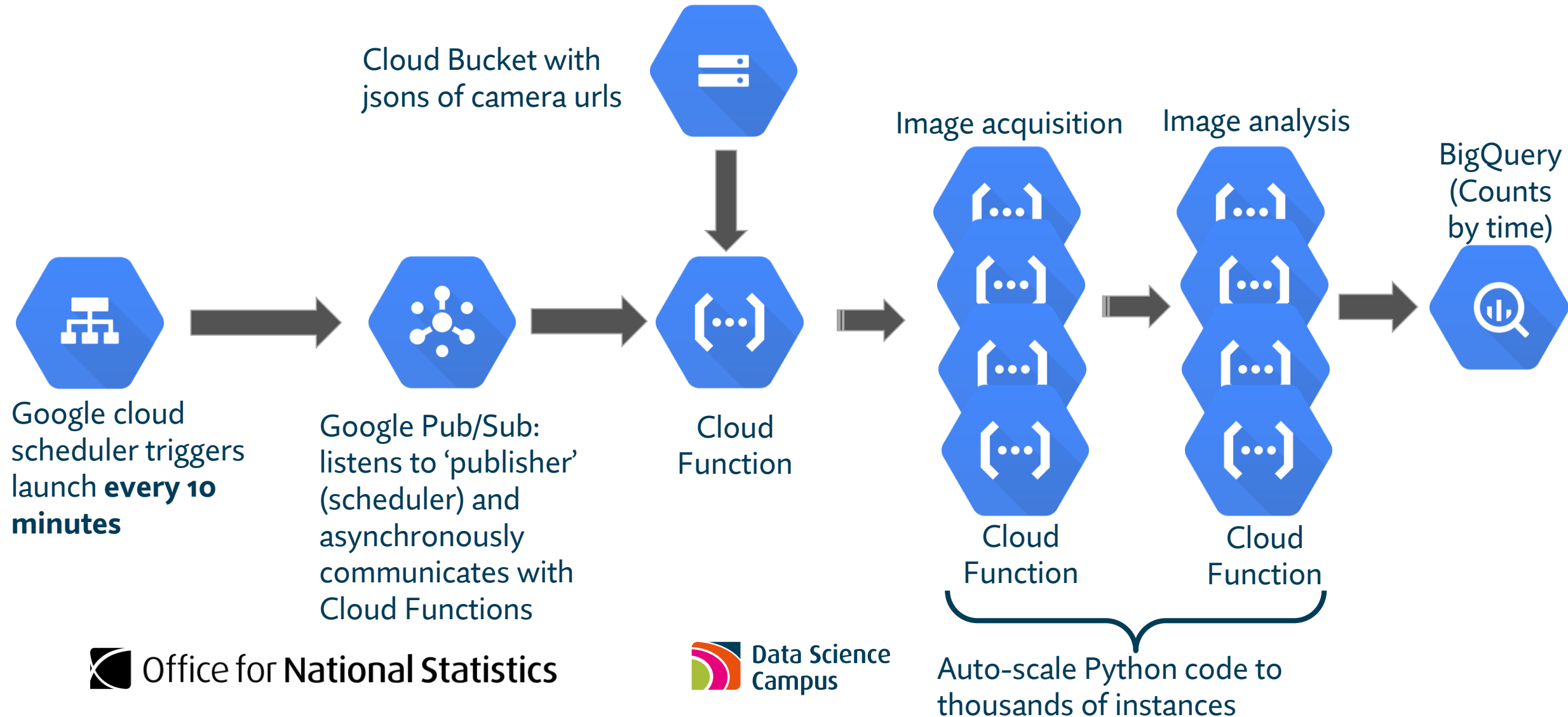
Traffic lights that look like pedestrians!

# Automating the process: the cloud pipeline



# Automating the process: scaling up

## Cloud architecture for image acquisition and object counts



# Automating the process: scale and cost

Clearing a 3-month image backlog saw 15,000 Cloud Function processes running simultaneously, achieving impressive **scale**—but scales down to zero when not in use....

run\_model\_on\_image

Version 5, deployed at 30 Jun 2020, 13:57:58

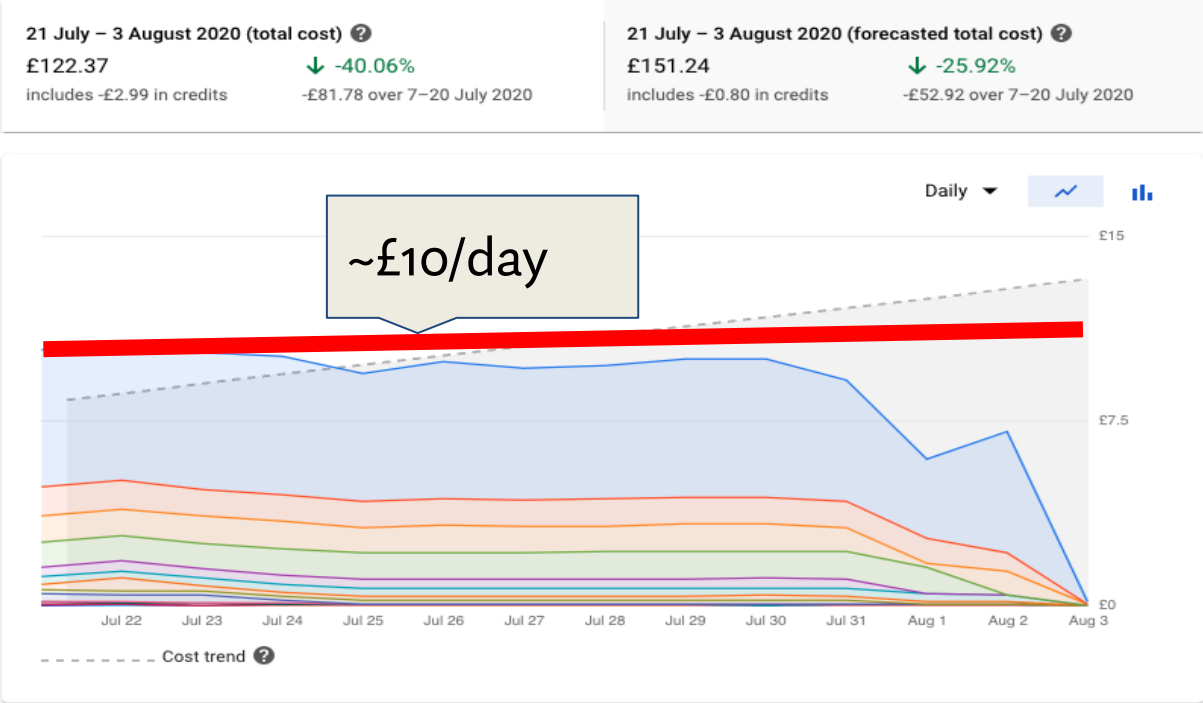
GENERAL TRIGGER SOURCE PERMISSION TESTING

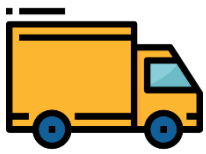
Active instances

14 days 30 days



...meaning it's very efficient, with a **cost** to run of £10 per day in total





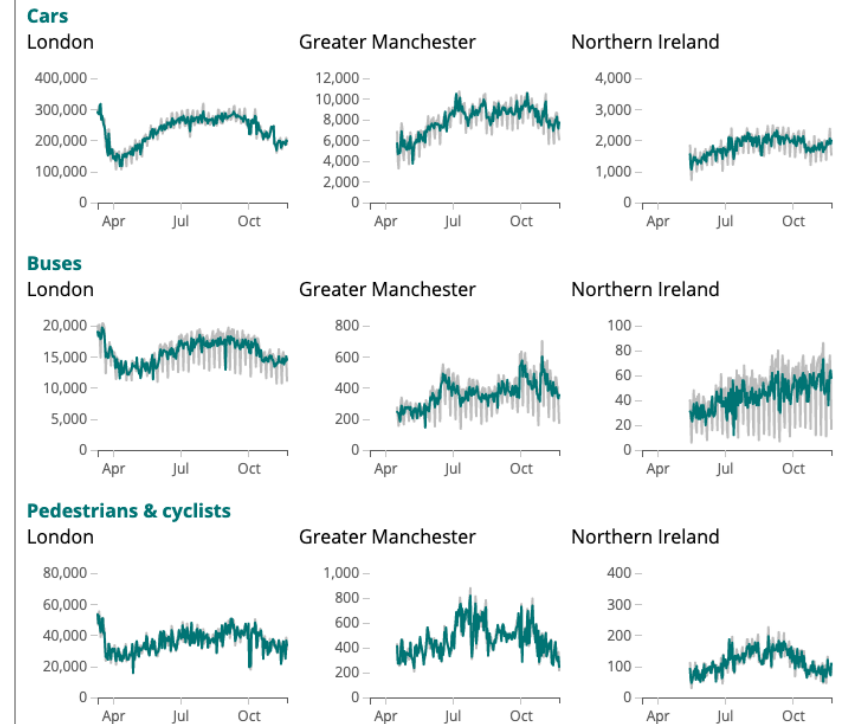
# Outcomes from the Traffic Cameras Project:

- Traffic busyness now makes up one of ONS' "Economic activity and social change in the UK, real-time indicators", available on the website weekly
- Paper with methods published in *Sensors*
- Will be trialled elsewhere, eg with Statistics Sweden
- Several prizes at Geography in Government awards
- Code open sourced: [https://github.com/datasciencecampus/chrono\\_lens](https://github.com/datasciencecampus/chrono_lens)

**Figure 11: In the week ending 22 November 2020, average counts of all traffic camera activity increased in London when compared with the previous week**

Activity in selected areas, daily counts of cars, buses, pedestrians and cyclists, seasonally adjusted, March to November 2020, UK

— Daily number (not seasonally adjusted)  
— Daily number (seasonally adjusted)



Source: Transport for London, Transport for Greater Manchester, TrafficWatchNI

# Today's key takeaways

- Establishing a moderate level of data science tooling and skills presents big challenges in any large organisation, and the level of capability varies enormously across the public sector
- ONS and Data Science Campus are at heart of efforts to increase data and data science capability across the public sector
- Some of our recent projects demonstrate how to do data science at scale and in production!





# More info

- Chen, Li, Ian Grimstead, Daniel Bell, Joni Karanka, Laura Dimond, Philip James, Luke Smith, and Alistair Edwardes. "Estimating vehicle and pedestrian activity from town and city traffic cameras." *Sensors* **21**, no. 13 (2021): 4564. <https://doi.org/10.3390/s21134564>
- Code for Traffic Cams: [https://github.com/datasciencecampus/chrono\\_lens](https://github.com/datasciencecampus/chrono_lens)
- Data Science Campus: <https://datasciencecampus.ons.gov.uk/>
- ONS' faster indicators pack:  
<https://www.ons.gov.uk/economy/economicoutputandproductivity/output/bulletins/economicactivityandsocialchangeintheukrealtimeindicators/26august2021>