

The challenges and benefits of visualizing COVID-19 data

Colin Angus

School of Health & Related Research

The challenges and benefits of visualizing COVID-19 data

How Open Science (and Twitter) saved me from
madness in 2020

Colin Angus

School of Health & Related Research



The
University
Of
Sheffield.

About me

3



A screenshot of a Twitter profile card for Colin Angus. The header image shows a dark, stormy sea with several bright lights reflecting on the water. The profile picture is a circular photo of a smiling man with short hair, wearing a blue jacket, standing in a grassy field. To the right of the profile picture is a blue button that says "Edit profile". Below the profile picture, the name "Colin Angus" is displayed in bold, followed by the handle "@VictimOfMaths". The bio text reads: "Alcohol policy modeller in @sarg_scharr | Health inequalities | COVID-19 | Data visualisation | RStats | Cake | Cycling | Pedantry". Below the bio is a link to "github.com/VictimOfMaths" with a GitHub icon, followed by a calendar icon and the text "Joined November 2013". At the bottom, it shows "599 Following" and "4,929 Followers".

Colin Angus
@VictimOfMaths

Alcohol policy modeller in @sarg_scharr | Health inequalities | COVID-19 | Data visualisation | RStats | Cake | Cycling | Pedantry

github.com/VictimOfMaths Joined November 2013

599 Following 4,929 Followers

Before the world fell apart

My 'real' job involves modelling alcohol pricing policies and their impact on health, the economy and inequality...

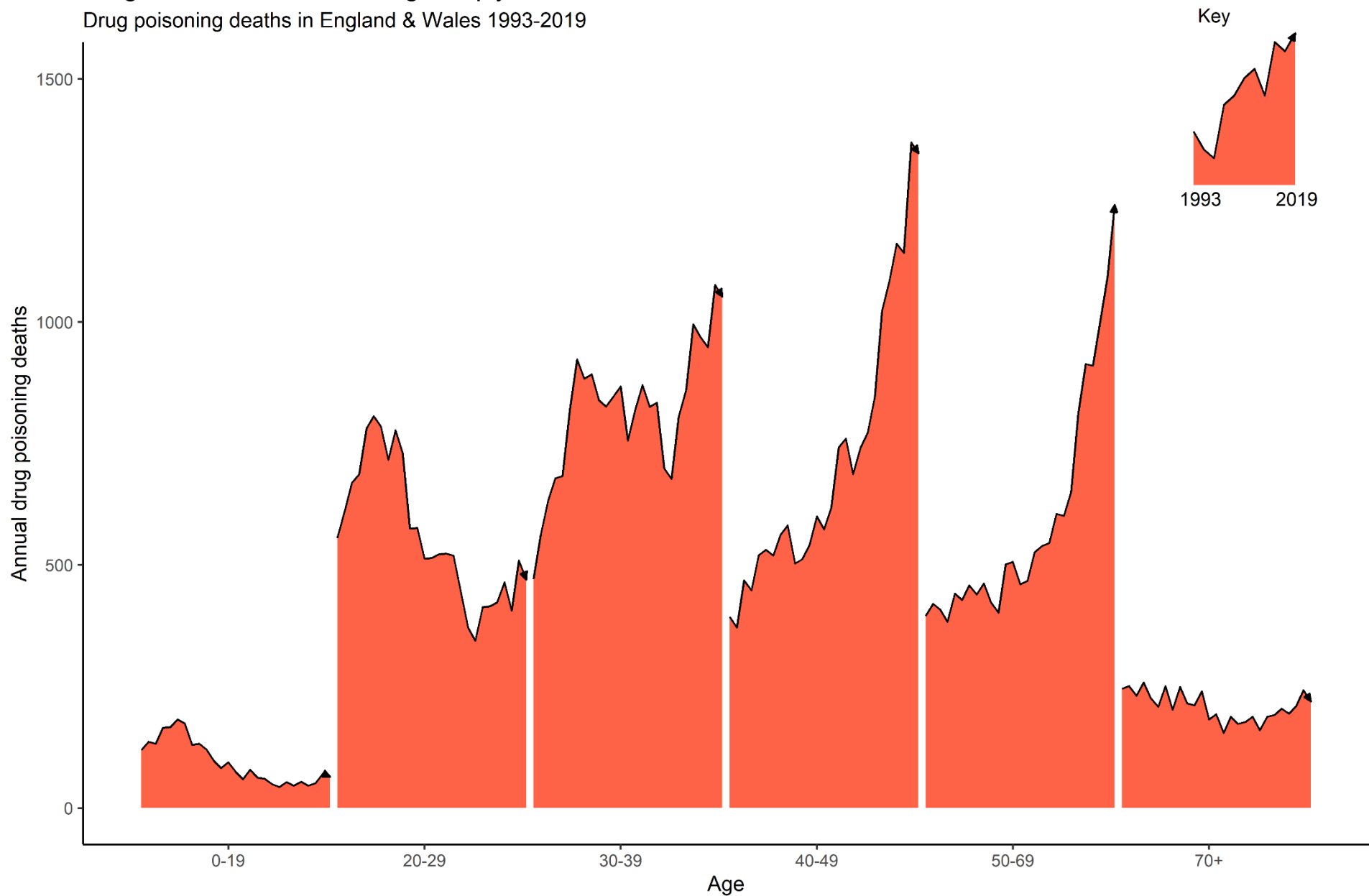
Before the world fell apart

My 'real' job involves modelling alcohol pricing policies and their impact on health, the economy and inequality...

...but I also had a sideline in doing little bits of analysis and visualising data related to alcohol or inequality that I found interesting and posting these on Twitter.

Drug-related deaths are rising sharply in older adults

Drug poisoning deaths in England & Wales 1993-2019



Data from Office for National Statistics | Plot by @VictimOfMaths



The
University
Of
Sheffield.

Then March happened

7

Then March happened

- I started drawing a few graphs of COVID-related things, largely just to help me get my own head around what was going on

Then March happened

- I started drawing a few graphs of COVID-related things, largely just to help me get my own head around what was going on
- These included maps of potential COVID risk, analysis of case numbers and estimates of excess mortality

Then March happened

- I started drawing a few graphs of COVID-related things, largely just to help me get my own head around what was going on
- These included maps of potential COVID risk, analysis of case numbers and estimates of excess mortality
- Wherever possible, this was broken down by age, sex, deprivation and/or local area

Then March happened

- I made sure to link every plot to the R code that generated it

Then March happened

- I made sure to link every plot to the R code that generated it
- I posted Twitter threads discussing what I thought the graphs might show and what they might not show

Then March happened

- I made sure to link every plot to the R code that generated it on GitHub
- I posted Twitter threads discussing what I thought the graphs might show and what they might not show
- With the help of the lovely RDM folks, I archived the latest versions of all of my COVID-related plots on Figshare



The
University
Of
Sheffield.

What this looked like...

14

So what happened?

The good

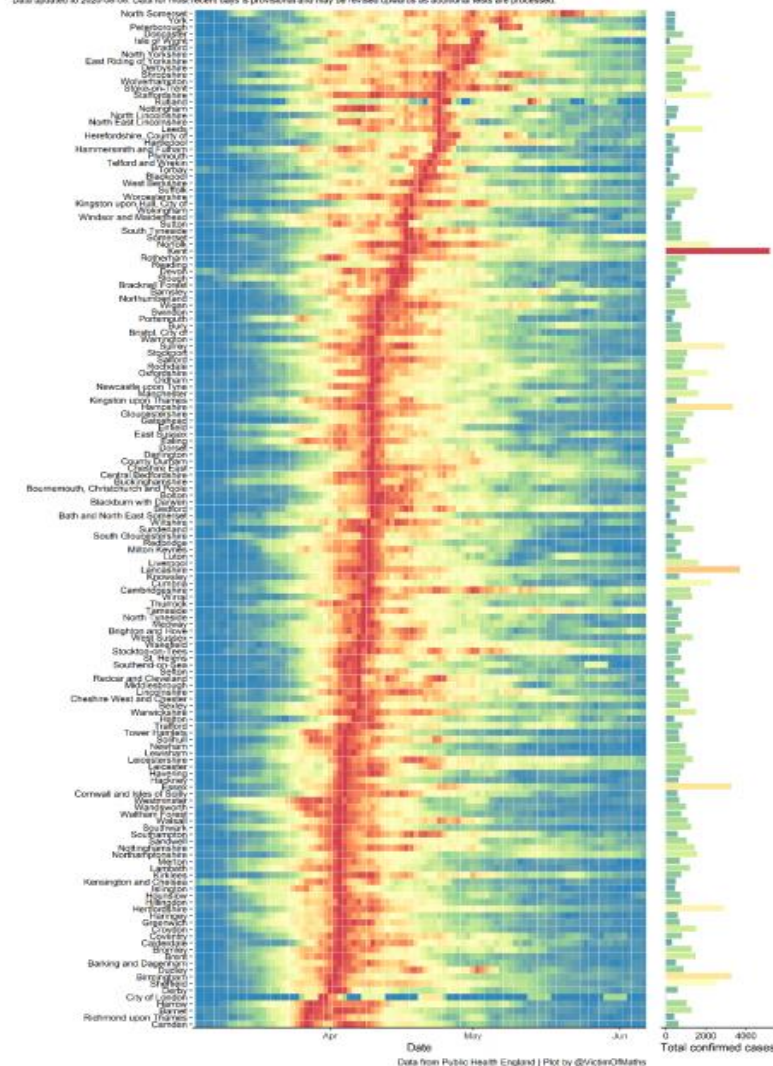
- One of my graphs was used in a report by the IFS

The good

Figure 4. Timelines for COVID-19 cases by local authority in England as of 6 June

Timelines for COVID-19 cases in English Local Authorities

The heatmap represents the 5-day rolling average of the number of new confirmed cases, normalised to the maximum value within the Local Authority. (As are ordered by the date at which they reached their peak number of new cases). Bars on the right represent the absolute number of cases in each LA. Data updated to 2020-06-06. Data for more recent dates is provisional and may be revised upwards as additional tests are processed.



Source: Angus, 2020.

The good

- One of my graphs was used in a report by the IFS
- I was quoted on BBC Radio 4's More or Less

The good

- One of my graphs was used in a report by the IFS
- I was quoted on BBC Radio 4's More or Less
- My code has been adapted and used by:
 - Public Health teams in the UK

The good

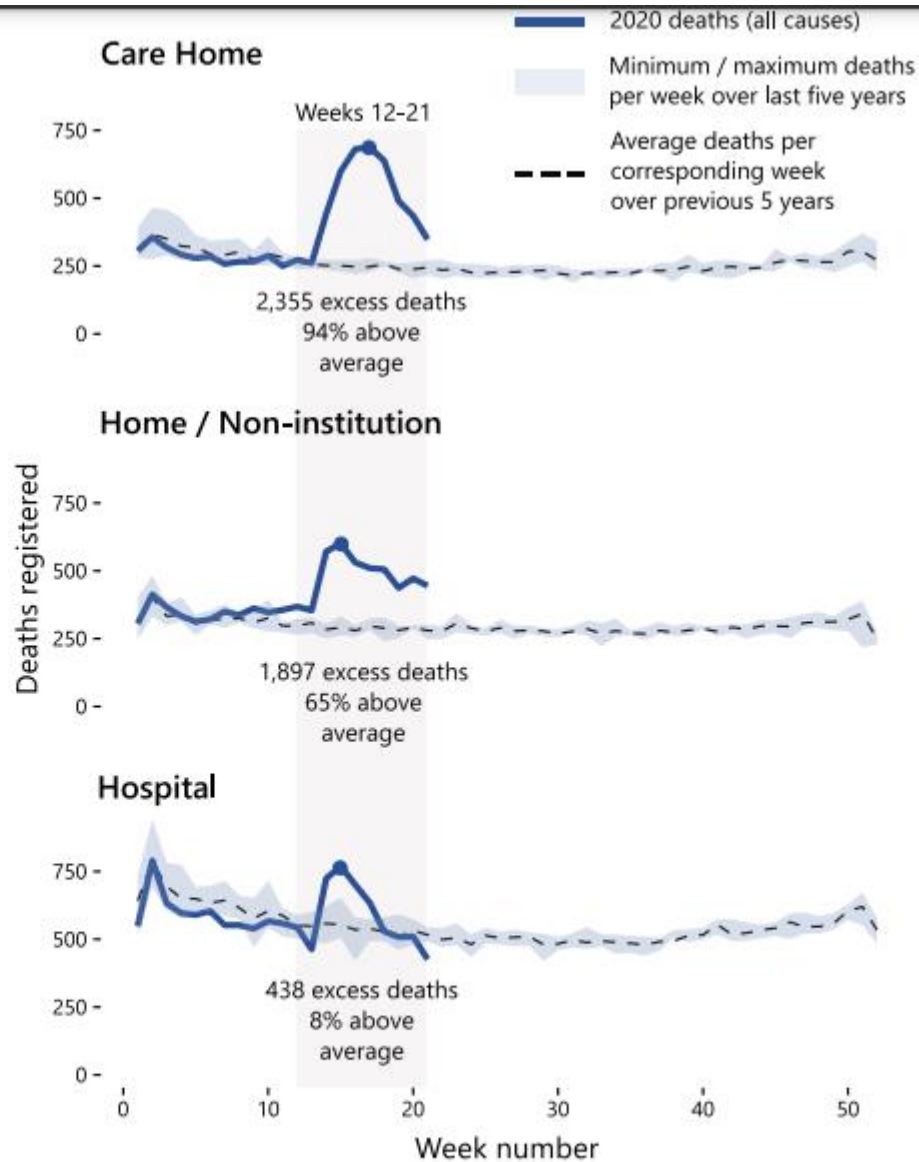
- One of my graphs was used in a report by the IFS
- I was quoted on BBC Radio 4's More or Less
- My code has been adapted and used by:
 - Public Health teams in the UK
 - Official statistical agencies

have occurred in care homes

Between weeks 12 and 21 (16 March to 24 May) there were 2,355 (94%) more deaths in care homes than average. Excess deaths peaked in week 17 and have been falling ever since, but deaths are still higher than average for this time of year.

In the same period, there were 1,897 excess deaths which took place at home or in a non-institutional setting (65% above average).

Excess deaths in hospitals peaked in week 15 and have now fallen to below average levels. The total excess over weeks 12 to 21 is 8% above average.



- One of my graphs was used in a report by the IFS
- I was quoted on BBC Radio 4's More or Less
- My code has been adapted and used by:
 - Public Health teams in the UK
 - Official statistical agencies
 - International public health organisations



Painel Brasil

Painel Estados

Linha do tempo

Casos diários

Casos acumulados

SRAG

Sintomas

Relatório municipal

Duplicação de casos e óbitos

Fator de crescimento

Mapa Brasil

Medidas de combate

População em risco

Notas técnicas

Sobre o projeto

Atualização dos dados

16/11/2020 21:22:31

Sobre dados e discrepâncias

Linha do tempo

Nesta aba apresentamos um indicador de concentração de casos e óbitos por semana epidemiológica.

Os dados de incidência diária de casos são agregados por semana epidemiológica (encerradas) para todas as Unidades da Federação (UF) e municípios. Em seguida, é realizada uma padronização dos dados, calculando o valor percentual de cada semana em relação a semana com o maior valor de casos. Desta forma, os gráficos apresentam em cor vermelha, a semana epidemiológica com a maior concentração de casos, e em cores mais claras as outras semanas epidemiológicas. A mesma lógica foi aplicada na incidência de óbitos.

Brasil ou algum estado

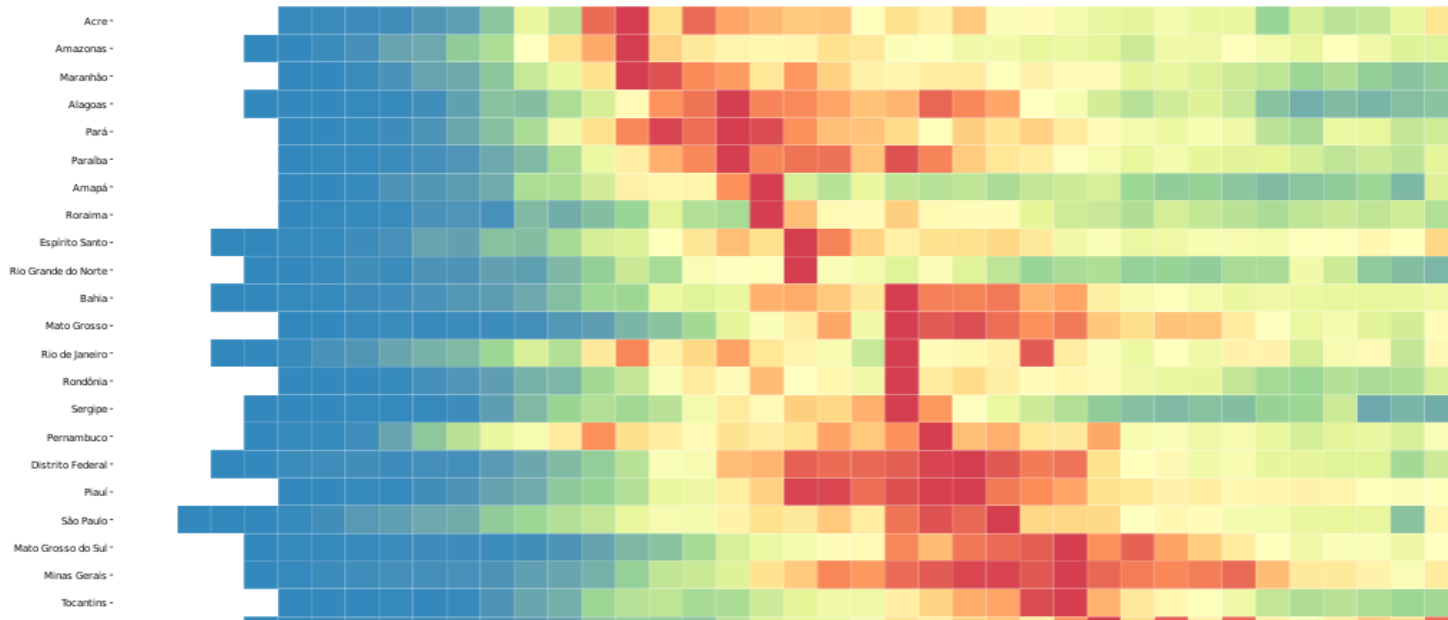
Brasil

Tipo

Casos

Linhas do tempo para casos de COVID-19 nos estados

O mapa de calor representa a quantidade de casos, normalizado pelo valor máximo no estado. Os estados foram ordenados segundo a semana epidemiológica com o maior número de casos.



- One of my graphs was used in a report by the IFS
- I was quoted on BBC Radio 4's More or Less
- My code has been adapted and used by:
 - Public Health teams in the UK
 - Official statistical agencies
 - International public health organisations
 - Other academics



David Henderson @_davidhen · 19h

For forthcoming @LTCcovid UK and international reports. A look at excess deaths by location of death and % change in each week compared to average of last 5 years across UK countries where available. Excess deaths at home still high in recent weeks.

Percentage change in 2020 weekly all-cause deaths from previous 5 year average



1

10

8

↑



David Henderson
@_davidhen

Big shout to @VictimOfMaths for code to wangle (very messy) E & W data....

1:26 PM · Jun 22, 2020 · Twitter for Mac

More good

- I've been invited to write blog posts and journal articles about my COVID work
- I've forced myself to learn and adopt *loads* of new skills and better coding and open science practices
- I've made *loads* of useful new contacts
- I've learned more about the vagaries of UK mortality data than anybody ever needs to know

The bad

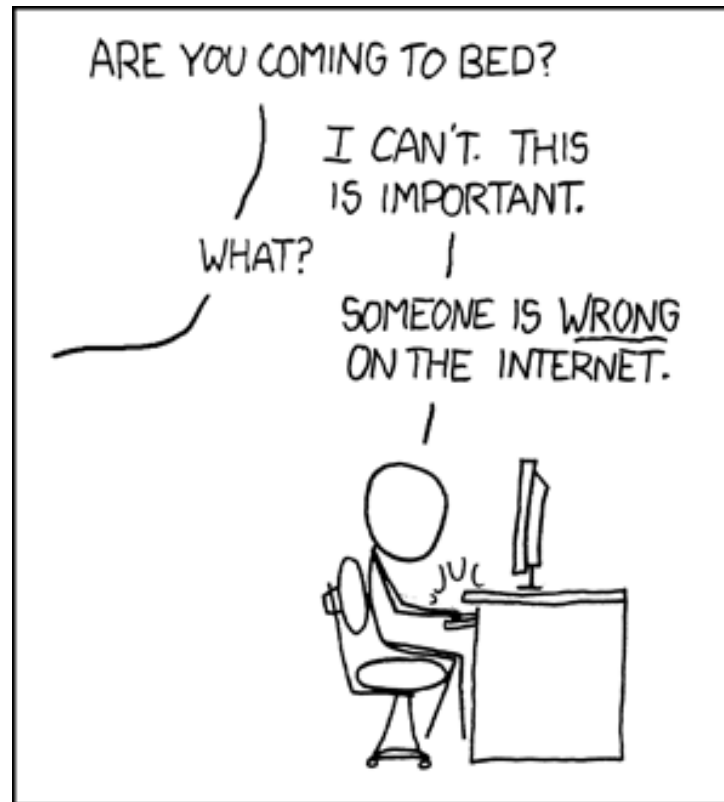
- Making data more accessible to a wider audience can be time consuming and hard work

- Making data more accessible to a wider audience can be time consuming and hard work
- The data landscape is always changing, so you constantly have to revisit and update things to keep them working

The bad

- Making data more accessible to a wider audience can be time consuming and hard work
- The data landscape is always changing, so you constantly have to revisit and update things to keep them working
- The internet is full of lovely, helpful, creative people, but also plenty of strange ones with strange opinions

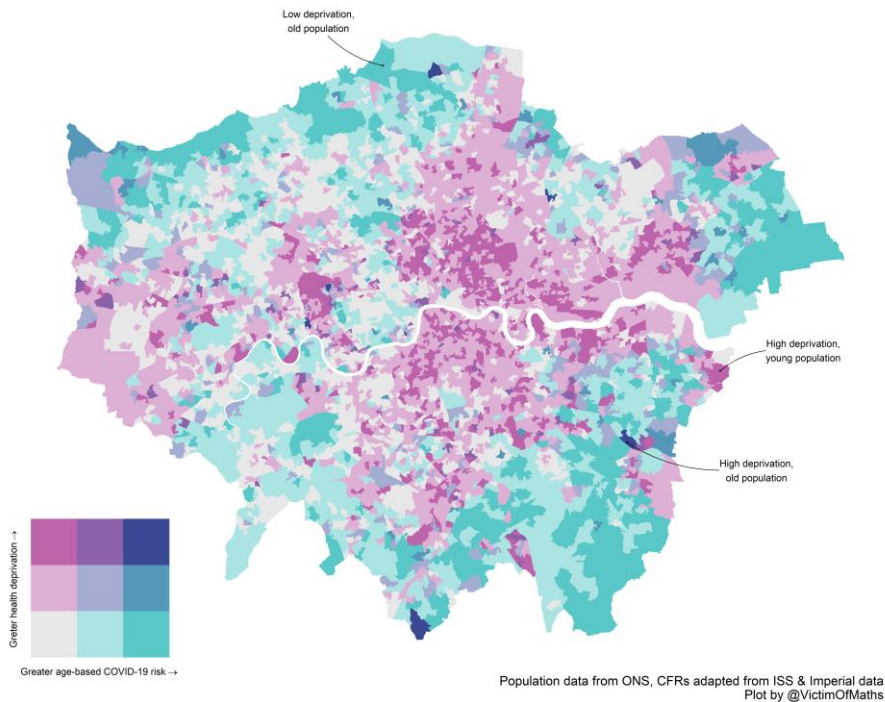
The bad



© Randall Munroe xkcd.com

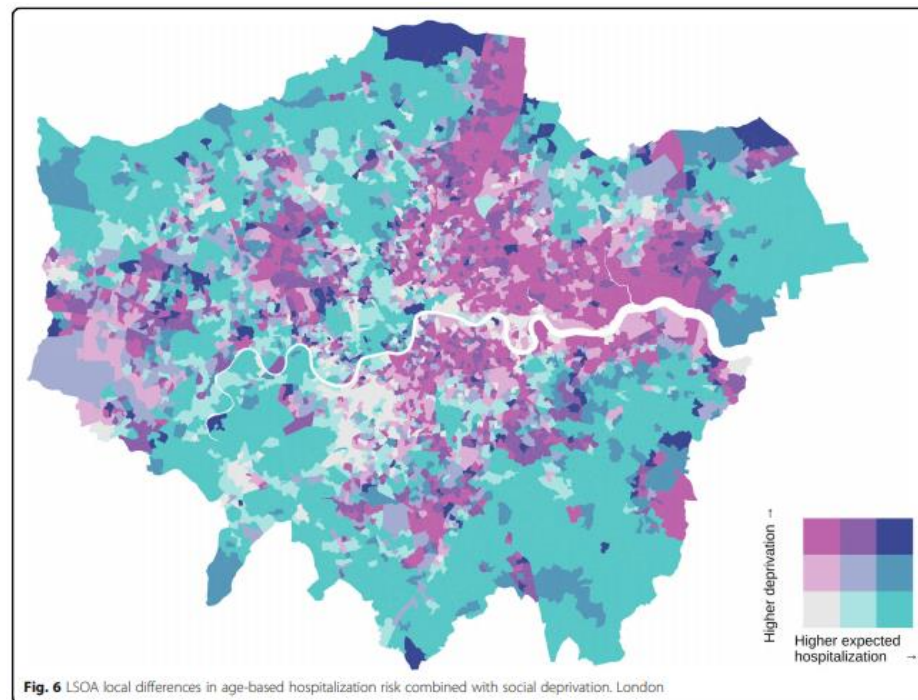
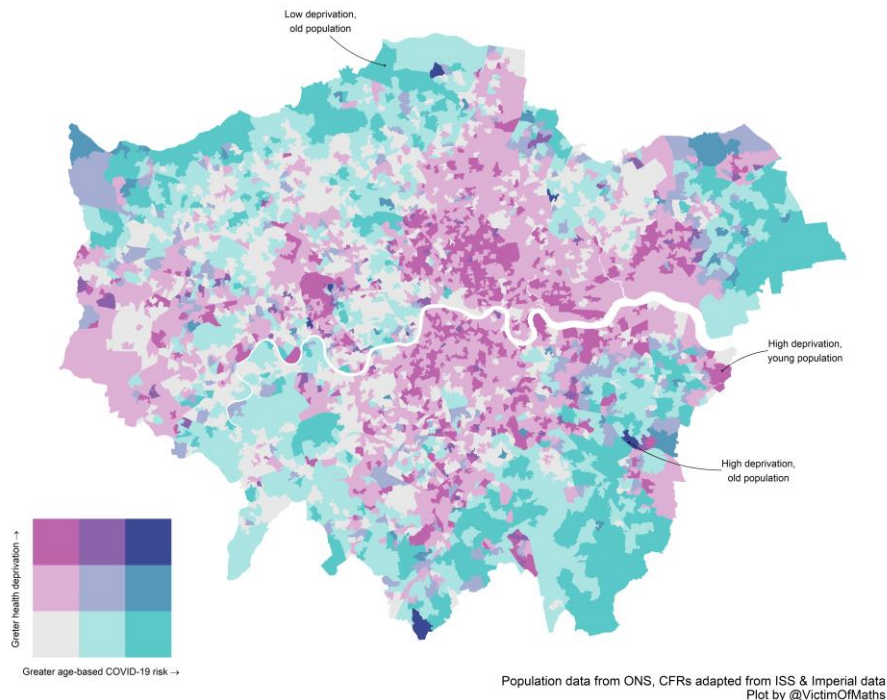
Mapping potential COVID-19 risk across London

LSOA-level health deprivation and potential COVID-19 mortality risk based on age-sex structure of population



Mapping potential COVID-19 risk across London

LSOA-level health deprivation and potential COVID-19 mortality risk based on age-sex structure of population




Verhagen et al. *BMC Medicine* (2020) 18:203
<https://doi.org/10.1186/s12916-020-01646-2>

BMC Medicine

RESEARCH ARTICLE

Open Access

Forecasting spatial, socioeconomic and demographic variation in COVID-19 health care demand in England and Wales

Mark D. Verhagen^{1*}, David M. Brazeal¹, Jennifer Beam Dowd¹, Ilya Kashnitsky^{2,3} and Melinda C. Mills^{1*} 



What I've learned

33

- Making the code available hugely increases trust in what you are presenting

What I've learned

34

- Making the code available hugely increases trust in what you are presenting
- It also allows people to help you improve it...

What I've learned

- Making the code available hugely increases trust in what you are presenting
- It also allows people to help you improve it...
- ...including chasing down bugs you'd missed...

What I've learned

36

- Making the code available hugely increases trust in what you are presenting
- It also allows people to help you improve it...
- ...including chasing down bugs you'd missed...
- ...and solving the ones you hadn't...

What I've learned

37

- Making the code available hugely increases trust in what you are presenting
- It also allows people to help you improve it...
- ...including chasing down bugs you'd missed...
- ...and solving the ones you hadn't...
- ...*and* do their own cool things that build on your work

What I've learned

- Archiving things on Figshare means that people can cite your work *and you can find out about it*
- Regular archiving of evolving work means you have a historic record of past versions

tl;dr

Open science can be a faff sometimes, but it's pretty great and we should all do more of it.



The
University
Of
Sheffield.

Thanks for listening

c.r.angus@sheffield.ac.uk



[@VictimOfMaths](https://twitter.com/VictimOfMaths)



[VictimOfMaths](https://github.com/VictimOfMaths)