

## Tutorial on plots

### Rough outline

The wonderful availability of up-to-date data on the coronavirus epidemic has led to an explosion of explanatory graphs, with accompanying claims about the similarity or dissimilarity of the trajectories of different countries, and how this relates to the measures they have taken against the virus. This tutorial looks at the different kinds of graph being used, what kind of questions they may help answer, and some guidance on avoiding over-interpreting apparent patterns.

### Data sources

A number of sites accumulate data from national sources, and produce various dashboards and plots. These include

- The Center for Systems Science and Engineering (CSSE) at Johns Hopkins University provides national data, and plots cumulative and daily cases on a linear scale

<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

- For the UK, Public Health England (PHE) graphs daily and cumulative confirmed cases, and total deaths

<https://www.arcgis.com/apps/opsdashboard/index.html#/f94c3c90da5b4e9f9a0b19484dd4bb14>

- Worldometer provides cumulative and daily, linear and log scales

<https://www.worldometers.info/coronavirus/#countries>

- The Financial Times FT – log cumulative cases and deaths, normalised to a common starting point, using data from Johns Hopkins.

<https://www.ft.com/coronavirus-latest>

World in Data

<https://ourworldindata.org/coronavirus>

data from ECDC

<https://ourworldindata.org/coronavirus-source-data>

There are numerous other sites.

### What do the numbers actually count?

What is a 'confirmed case'?

This is confusing. Is it someone who has a disease ? (COViD-19. Or is it Someone who has been tested for SARS-COV-2, .. and the test has been positive.

? PHE separate out these notifications [link]. But in their reporting, reporting positive test results but calling them covid-19. Fine when only symptomatic cases tested, but what when population testing comes in?

So people who have the virus and who have not been tested do not appear, which means this is only a minimum – the true number with the virus may be far higher, depending on the testing regime in the area. [link to World in Data discussion]. The number of people who have been infected, but not counted, will therefore vary from place to place: when everyone on the Diamond Princess cruise ship was tested, over 600 were positive but around half of these had not symptoms and would normally not have been tested. Numbers of confirmed cases should be treated with great caution, especially when comparing countries. But in countries with a fairly stable testing regime, they may be useful for assessing trends.

Important that the daily count is not necessarily the number who tested positive in a recent 24 hour window - it is the number of reports received centrally. So we can expect the daily counts to jump around, not only because of chance variability, but simply because of different local practices in submitting test results.

### **What is a 'Covid death'?**

Depends on the country. In the UK, it is someone who has tested positive and subsequently died. Vital to realise it is some who has died with the virus, and not necessarily of the virus. (16% causal figure from China?).

Just as with confirmed case, the count does not reflect deaths that occurred in the previous 24 hours, but deaths that were reported. There can be delays, for example to allow the family have been traced and informed[link]: for example a record 107 deaths were reported for March 26<sup>th</sup>, but some of these had occurred some days beforehand - the 11 deaths reported by

One NHS TRUST for the 26<sup>th</sup> March actually occurred on the 16<sup>th</sup>, 17<sup>th</sup>, 19<sup>th</sup> (2), 20<sup>th</sup>, 22<sup>nd</sup>, 23<sup>rd</sup> (6) of March.

Also may not include people who die outside hospital and were never tested.

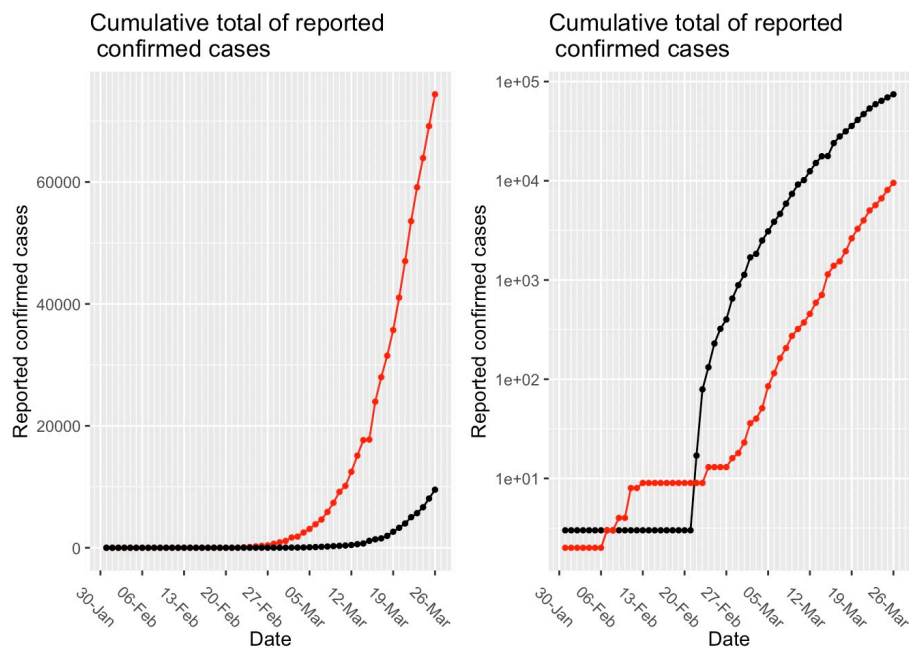
Low figures can also occur due to repriming decision: the low figure of X on March 24<sup>th</sup> can be attributed to changing from a 9am to 9am window to a 5pm to 5pm window, and so the deaths Yesterday's published figures only comprised the eight hour period from 9am 24<sup>th</sup> March to 5pm on 24<sup>th</sup> March.

Different countries may have different standards for concluding a Covid-19 death – Germany??

Knowing about the reporting process encourages great caution in interpreting daily counts.

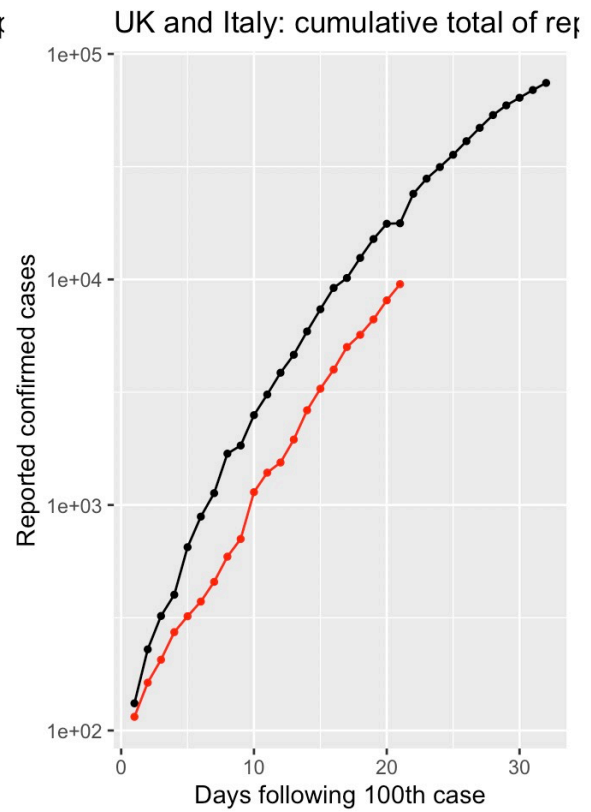
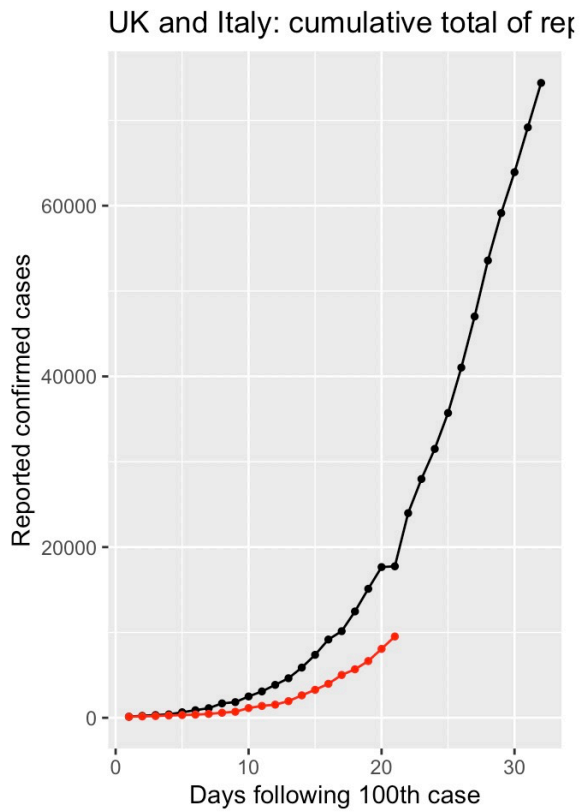
## Basic plotting

•



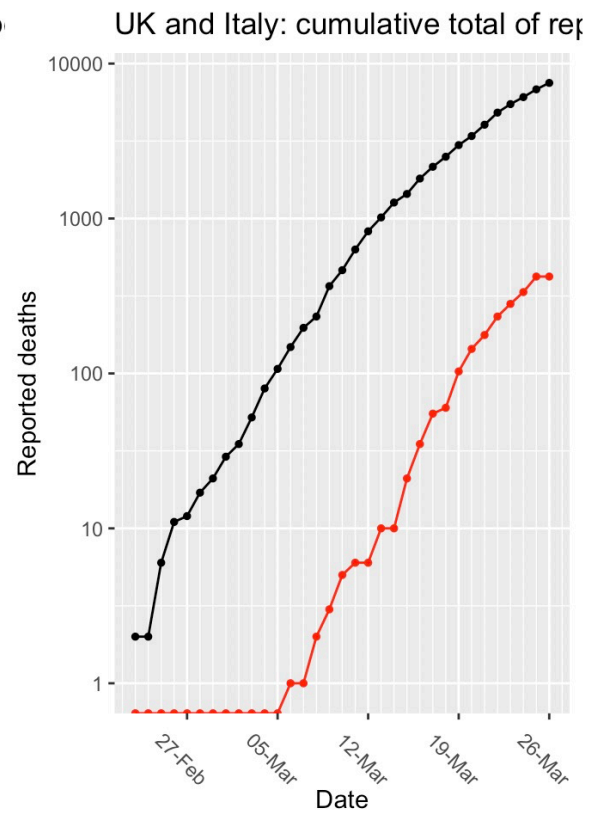
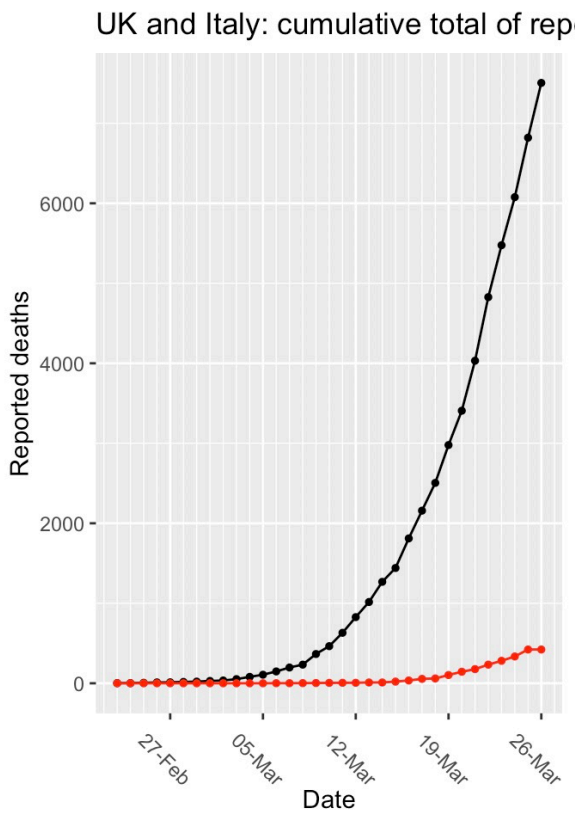
•

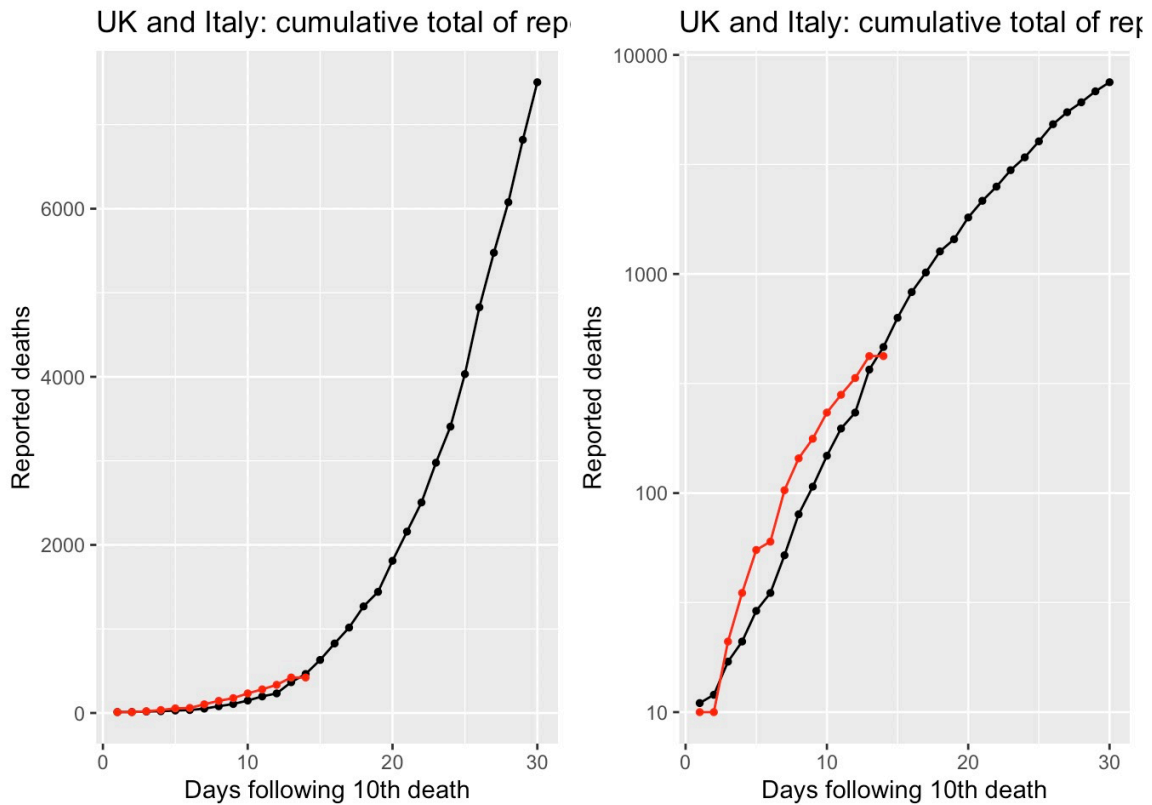
- Basic plot of cumulative UK and confirmed **cases** natural scale, [Italy and UK] , and log. Draw fan on log?? Show unreliability of early part of curve, warning not to over-interpret?
- Show lines of % increase / doubling time as on FT , BBC etc? Note that this is doubling of total count.
- Check what Burn-Murdoch says about log-scales (and what I say in my book!)
- Straight line Gradient is daily growth
- 
- <https://eagereyes.org/blog/2020/in-praise-of-the-diagonal-reference-line>



Plots aligned for 100 cze.

- Then deaths, natural and log scale,





Maths at bottom

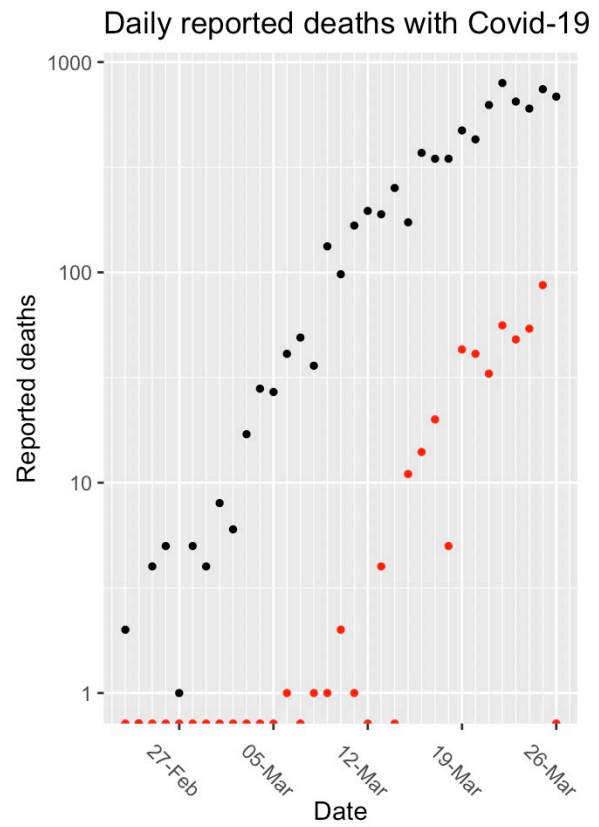
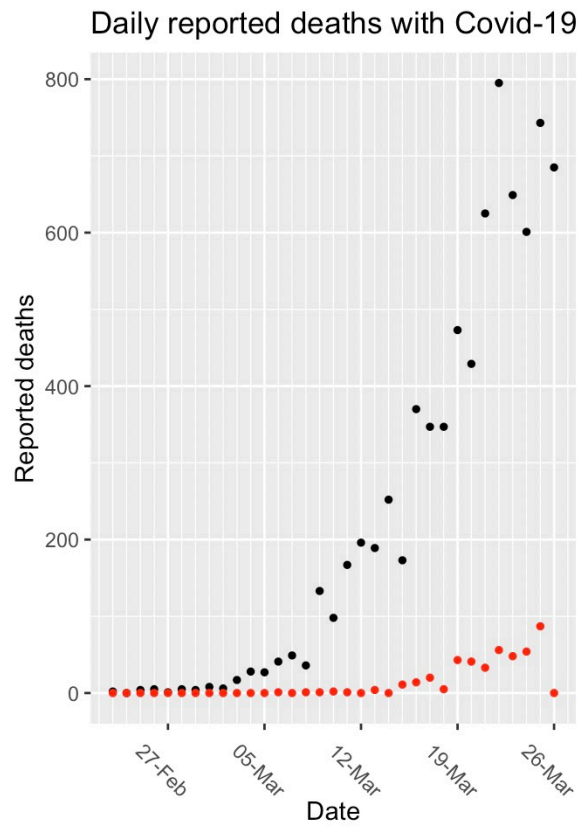
- After 10 deaths , natural and log
- Caution in interpreting early part of curve – very big uncertainty about ‘true’ risk gradient.

### Daily data

- Basic plot of daily Uk and Italian cases as points, natural scale and log
- Plot of daily deaths , natural and log [not aligned]

Discussion of why look at daily data – independent observations, quicker detection of changes, but great care not to over-interpret. Jumpy.  
can fit models (but note problems) – see future blog.

Crucial that straight lines here represent fixed % increase / doubling time of *daily count*, not *total count*.



## BLOG 2

### Fitting models

- Straight line
- Straight line with error bars

### Fitting models

- Curves..