

🕒 This article is more than 11 years old

John Snow's data journalism: the cholera map that changed the world

John Snow's map of cholera outbreaks from nineteenth century London changed how we saw a disease - and gave data journalists a model of how to work today

[Interactive map](#)

[Download the data](#)

[More data journalism and data visualisations from the Guardian](#)



📷 John Snow's cholera map of Soho. Click image to embiggen

Simon Rogers

Fri 15 Mar 2013 09:30 GMT

How often does a map change the world? In 1854, one produced by Doctor John Snow, altered it forever.

In the world of the 1850s, cholera was believed to be spread by [miasma in the air](#), germs were not yet understood and the sudden and serious outbreak of cholera in London's Soho was a mystery.

So Snow did something data journalists often do now: he mapped the cases. The map essentially represented each death as a bar, and you can see them in the smaller image above.



📷 Dr John Snow, anaesthetist. And data journalist? Photograph: Centre for Sexual & Reproductive Health

It became apparent that the cases were clustered around the pump in Broad (now Broadwick) street.

There were some outliers though and Snow wrote that:

■ ■ *In some of the instance , where the deaths are scattered a little further from the rest on the map, the malady was probably contracted at a nearer point to the pump*

One 59-year-old woman sent daily for water from the Broad street pump because she liked its taste. Wrote Snow:

■ ■ *I was informed by this lady's son that she had not been in the neighbourhood of Broad Street for many months. A cart went from broad Street to West End every day and it was the custom to take out a large bottle of the water from the pump in Broad Street, as she preferred it. The water was taken on Thursday 31st August., and she drank of it in the evening, and also on Friday. She was seized with cholera on the evening of the latter day, and died on Saturday*

At a local brewery, the workers were allowed all the beer they could drink - it was believed they didn't drink water at all. But it had its own water supply too and there were consequently fewer cases.

In nearby Poland street, a workhouse was surrounded by cases but appeared unaffected: this was because, again, it had its own water supply.

It turned out that the water for the pump was polluted by sewage from a nearby cesspit where a baby's nappy contaminated with cholera had been dumped. But he didn't just produce a map; it was one part of a detailed statistical analysis.


As the [Public Health Perspectives blog says](#), it changed how we see data visualisations, and how we see microbes. Snow was born 200 years ago this week and is the [subject of an exhibiton at the London School of Hygiene and Tropical Medicine](#).

But how would those deaths look for a data journalist today?

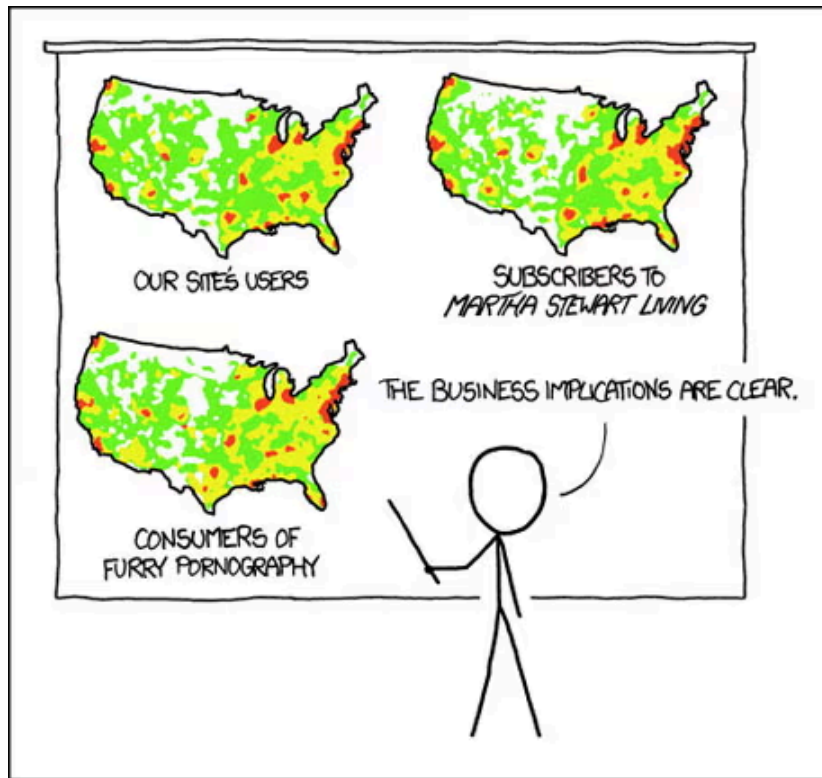
Thanks to [Robin Wilson at Southampton University, we have the data](#). Robin painstakingly georeferenced every cholera death and pump location, so we could recreate the map on a modern layout of London. We wondered what would happen if we tried to recreate the map using a modern tool, opting to try [CartoDB](#), using the lovely [Stamen](#) 'toner' projection to at least keep the background in common with Snow's London.

An interactive version

 Water pumps  Cholera deaths

 Cholera map key Photograph: Guardian

As XKCD have pointed out, heatmaps or dotmaps have flaws, not least that they tend to show where the people are.



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

📷 XKCD on heatmaps. Image: XKCD

And the alternative is usually to aggregate the data, so that you could show, say, the incidence of cholera by geographical area - a choropleth. But in this case, would that have worked?

The cluster of dots around the Broad street pump were what alerted Snow to the cause of the outbreak.

[Edward Tufte is interesting on this](#). He points out that

■ ■ *The big problem is that dot maps fail to take into account the number of people living in an area and at risk to get a disease ... Snow's dot map does not assess varying densities of population in the area around the pump*

But, as Tufte points out, this part of Soho was incredibly thickly populated. And "aggregations by area can sometimes mask and even distort the true story of the data". A choropleth map of the area might show that there was a cluster of cholera cases, but it might not, depending on where the

boundaries are drawn. Mark Monmonier, author of [How to lie with maps has examined this](#).

But there's another key point here: in the event of an outbreak like this now, it's inconceivable that the government would publish the data on grounds of privacy; that the victims' addresses were personal data.

As data journalists, we agonise over how to represent the true impact of an event. **Maps** are often the first thing to reach for because it's easy: the tools are now just so easy to use and so much data is geographic. Although they are often mightily popular with readers, it's probably not always the right choice. Trying harder to show the data in different ways is an honourable objective.

But when they work, maps can tell a story in a language that everyone can understand.

Maybe Snow's map had such a huge impact on its own because it was simply a great data visualisation.

Robin Wilson has given us links to the data below. What can you do with it?

Download the data

[DATA: download the full spreadsheet as a Google Fusion table](#)
[Available in more formats here](#)

NEW! Buy our book

[Facts are Sacred: the power of data](#)

More open data

[Data journalism and data visualisations from the Guardian](#)

World government data

[Search the world's government data with our gateway](#)

Development and aid data

[Search the world's global development data with our gateway](#)

Can you do something with this data?

Flickr Please post your visualisations and mash-ups on our [Flickr group](#)
Contact us at data@guardian.co.uk

Get the A-Z of data

More at the Datastore directory

Follow us on Twitter

Like us on Facebook

You've read *21 articles* in the last year

Article count **on**

This is what we're up against