

Causal Analysis and John Snow's cholera theory: Part 1

2 min

In the world of data, we'll hear time and time again that "correlation does not equal causation." In other words, while two events might be connected or related, that doesn't mean they're in a cause-and-effect relationship.

A "causal link" means proving that one event causes another. One of the most important ways this has been applied in the last few centuries has been in epidemiology, the study of diseases. Discovering correct causal links has meant big things for the prevention and treatment of diseases.

Let's take a look at one of the earliest instances of successful causal analysis in medicine, which starts with a man called John Snow. (Not the fantasy-famous Lord of the North, but a real-life nineteenth century London doctor.)

Until Dr. Snow's discovery in the mid-nineteenth century, people believed that cholera was caused by vapors rising from the burial grounds of plague victims from two centuries earlier. (A good try, but cholera is actually a waterborne disease caused by bacteria found in sewage. It causes severe dehydration and has a fatality rate of over 50% when untreated.)

By studying earlier cholera epidemics and organizing his data analysis around his hunch that cholera was waterborne, Dr. Snow was able to link an 1854 cholera outbreak in London to a contaminated water pump – effectively proving a causal relationship between contaminated water and cholera **before humans even knew that bacteria existed!**

How'd he do it? Let's find out in Part 2.

Instructions

The image on the right is Dr. Snow's data visualization solution – [click here](#) for a full-size, zoomable version, and check out the next slide for more information on it!

