

Causal Analysis and John Snow's cholera theory: Part 2

3 min

Dr. John Snow's causal analysis breakthrough started with how he visualized his data: he organized cholera death records by location rather than by time, which was more common. He made a map, and discovered that the deaths centered around a water pump on Broad Street.

From there, Dr. Snow used death records that seemed to contradict his theory to strengthen his explanation. For instance, a woman who died of cholera in a completely different neighborhood had just visited her aunt's house near Broad Street and drunk water from the pump.

Dr. Snow also found that a workhouse and a brewery near the pump both had few or no cholera deaths. Upon investigation, he learned that the workhouse had its own water supply, and that the brewers not only had access to a well at the brewery, but that they drank only malt liquor and never visited the Broad Street pump.

Snow advised that the handle be taken off the Broad Street pump to prevent people from drinking the contaminated water. The handle was removed, and this action coincided with the end of that outbreak. The number of deaths was already trailing off (more than 75% of residents had left the area to avoid "choleric vapors"), but this public health intervention prevented the disease from recurring as people returned, and the epidemic ended.

The built-in test cases helped Snow to isolate variables and prove that the key variable was that people who developed cholera had drunk water from the contaminated pump. From there, repeated studies of cholera and modern lab experiments have only confirmed the causal link he discovered.

In modern lab science, we use controlled experiments to isolate variables and prove causation. Controlled experiments are often not possible outside of lab settings, though, so data scientists do the best they can to isolate and control variables and get comfortable working with some amount of error.

Instructions

The image on the right is Dr. Snow's data visualization solution, a map with cholera death data visualized directly on it in a "geographically-distributed bar chart". ([Click here](#) for a larger, zoomable version.)

There's a small dot on Broad Street, in the center of the map, labeled "PUMP." Notice how the bars (each representing an instance of cholera) are generally concentrated around this part of the map. There are other pumps labeled on the map— for instance, one just south of Broad Street, on the corner of Brewer Street and Bridle Street. All these other pumps have fewer cases of cholera near them than the Broad Street pump.

