Cholera Epidemiology: Analysis of the 1854 Outbreak, London, England

Epidemiology experienced a major breakthrough in 1854 in a groundbreaking study of Cholera outbreaks in the poverty stricken streaks of London, England. John Snow, a physician, coupled geography of death and a knowledge of epidemiology to save hundreds if not thousands of lives.

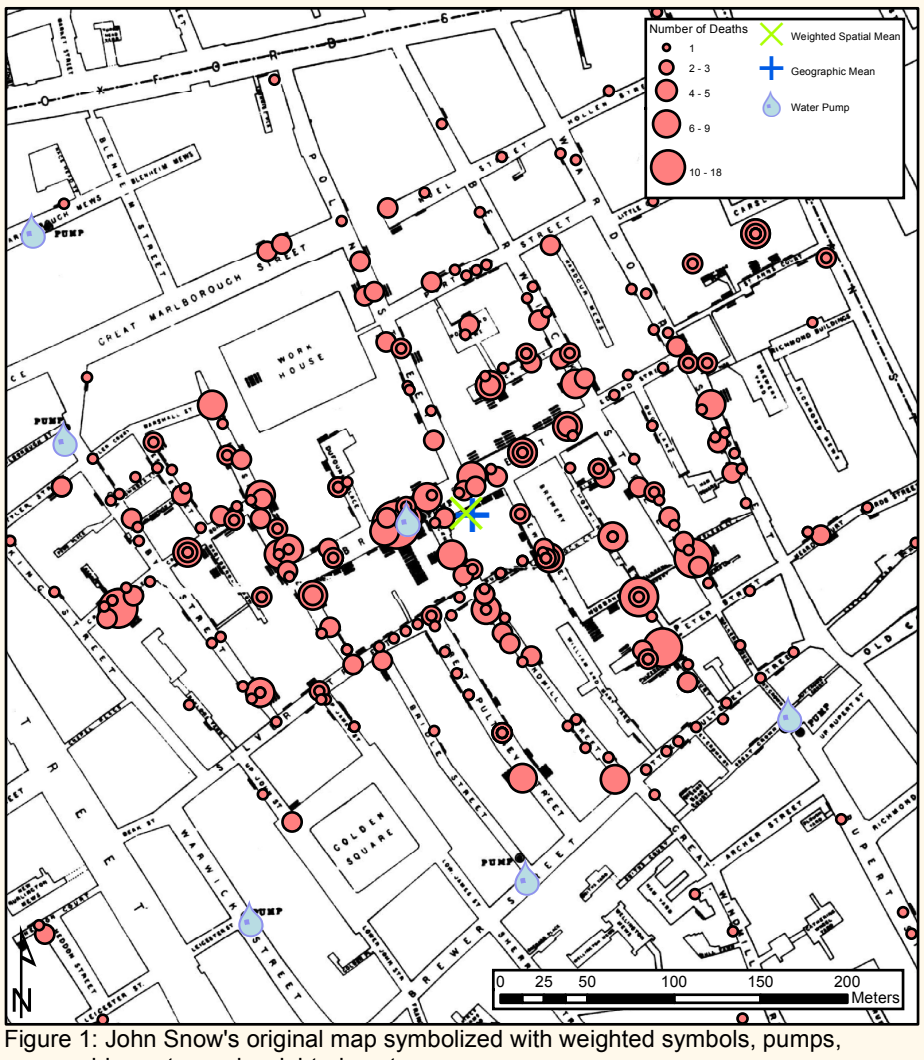

Figure 1: John Snow's original map symbolized with weighted symbols, pumps, geographic center and weighted center.

John Snow's original map is digitized on the left with the original data also transcribed into a graduated symbology where more deaths are shown by larger symbols. Each circle is placed at an address where a death was recorded, and the symbol grows the more deaths happen that address. Water pumps are shown with a water drop symbol. This is essentially the only data Snow had when he came to his conclusion. Once the deaths had been plotted geographically, Snow recognized a clustered pattern of deaths around the water pumps in the area.

This conclusion is illustrated by the geographic and weighted centers. The geographic center is the mean location of all death addresses, and the weighted mean uses the number of deaths to find the center. Both centers, but especially the weighted center are remarkably close to a water pump. Snow noticed this pattern and concluded the pump should be shut down because it was spreading cholera.

Using Snow's original data and locations of death we can use modern geographic and statistical techniques to confirm that Snow's hypothesis and patterns were truly significant. It also allows us to identify patterns that Snow may not have identified. These are techniques that are used in modern epidemiology studies to identify incidence, distribution and potential origins.

Thiessen polygons are created as a first basic step toward digital analysis of the data. Thiessen polygons are polygons that only include one feature, in this case pumps, and any location in the polygon is closer to its point than to any other point. These polygons identify possible areas of influence for each point. The standard distance circle shows how dispersed the deaths are. The size of the circle, centered on the weighted center contains 1 standard deviation (~68%) of the deaths. Smaller circles represent more compact spatial distribution, and larger circles represent a larger spatial distribution. These techniques are less based on statistics, more on easy pattern identification. The thiessen polygons and standard distance in concert with the number of deaths at each address makes interpreting Broadwick and Carnaby pumps as infected much easier.

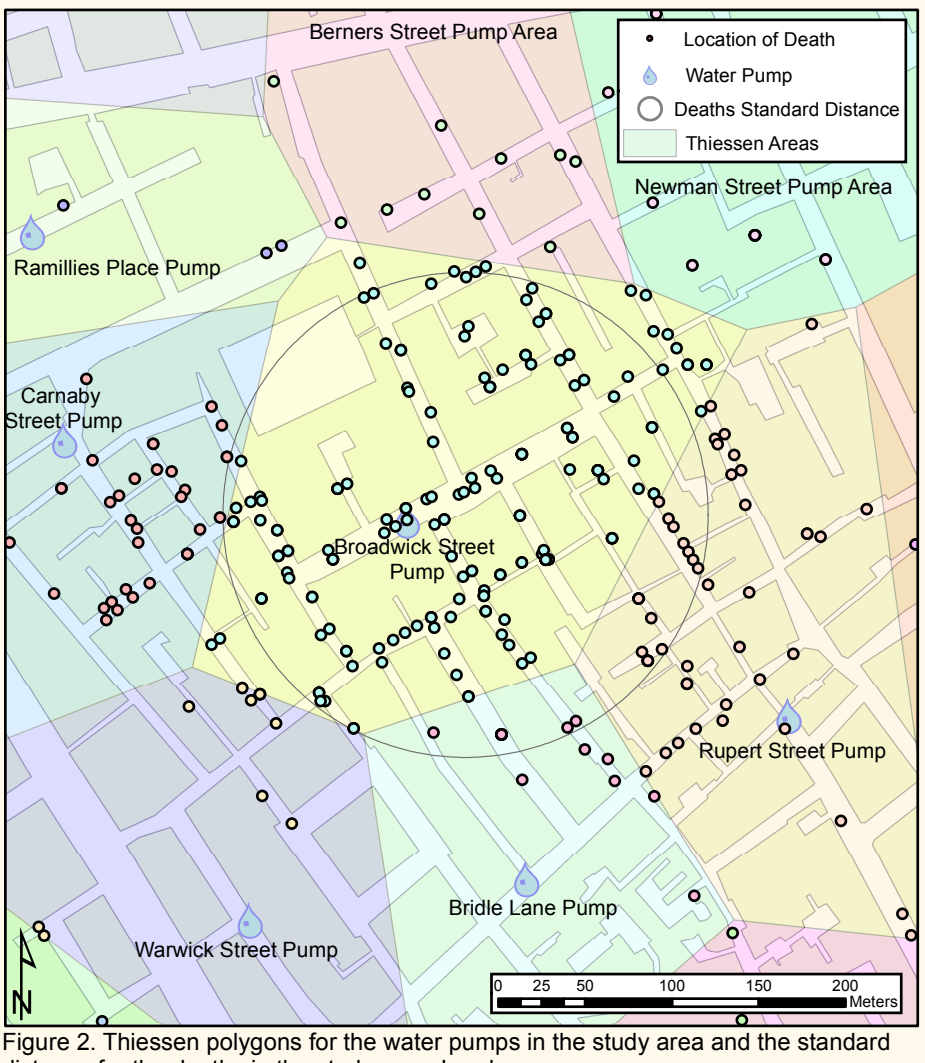

Figure 2. Thiessen polygons for the water pumps in the study area and the standard distance for the deaths in the study area, London.

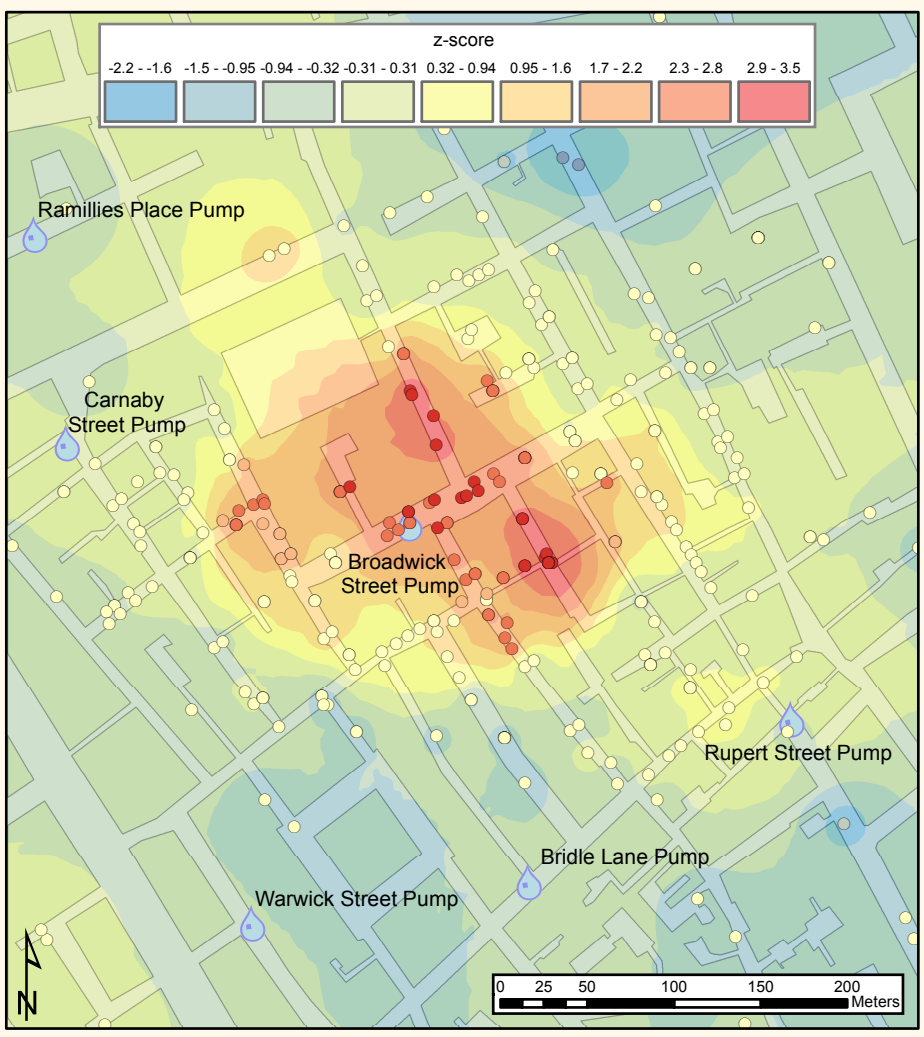

Figure 3. Getis Ord Gi Hot spot analysis of area around the Broad Street Pump, London. Computed using 105 meter bands and displayed using an inverse distance weighted method.

More using more advanced spatial statistic techniques, it is common now to use computers to identify patterns in data that the human eye may not be able to identify. Spatial statistics are also useful to compare patterns that identified by both computers and human eyes.

Advanced spatial statistics in the case of epidemiology are based on each point's relationship to each other, as well as to a completely random distribution. Incremental spatial autocorrelation is used to quantify how much each feature is spatially and numerically similar to the other points nearby and if there is a grouping of similar distances away. This grouping distance is used in the Getis-Ord GI* statistic to identify spatial clusters of high and low values. Using an inverse distance weighted interpolation technique, data is interpolated for areas where data had not been collected, creating a continuous surface called a 'hot spot' map, which is easier to interpret. The z-score refers to the number of standard deviations away the mean (random) distribution. High z-scores represent statistically significant clusters of high values, while low negative z-scores represent statistically significant clusters of low values.

Advanced spatial statistics show that there are statistically high values of cholera infection surrounding the Broadwick Street Pump, and that most of the surrounding area contains no significantly values with the exception of some clustered low values in the north. The Broadwick Pump was the original Pump identified as being contaminated, resulting in its closure and investigation.

Tabor Reedy, WWU, 11/20/2014, Data: WWU, ESRI Projection: WGS 1984 UTM Zone 30N