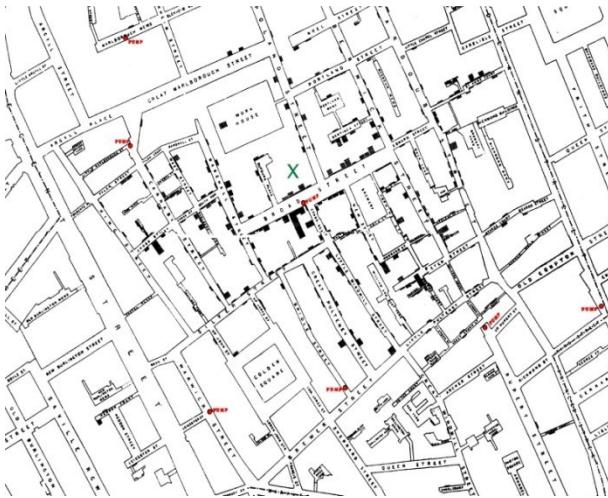




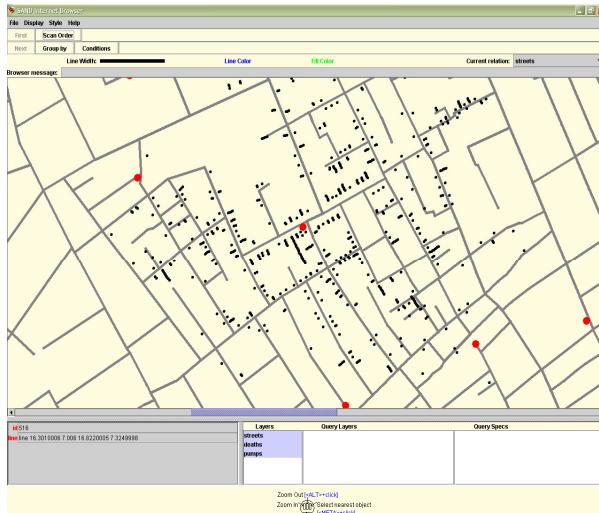
Knowledge Discovery using the SAND Spatial Browser

Hanan Samet, Adam Phillippe and Jagan Sankaranarayanan
University of Maryland at College Park

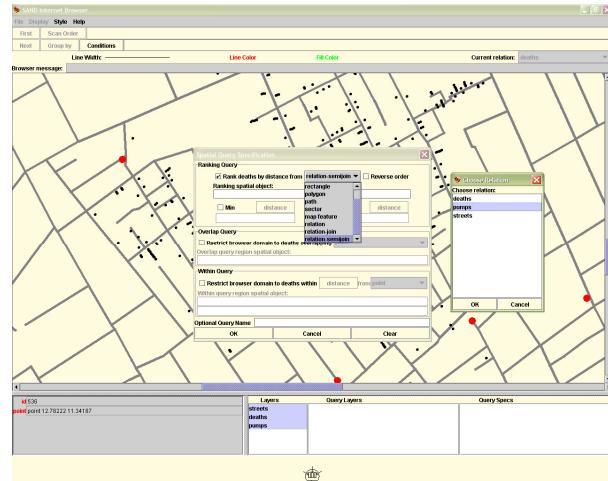
In 1854 Dr. John Snow, a pioneer in epidemiological research, showed that the cholera outbreak in Soho, London was a result of the contamination of a water pump on Broad Street [4]. He used a spot map that showed the number of deaths in each household, which he then overlayed on a map of Soho, that also contains the locations of the water pumps, as shown below. Upon looking at the map, it is immediately clear that the water pump, labeled X on the map, is responsible for the cholera deaths in Soho as the majority of the deaths due to cholera are clustered around this pump. The result was revolutionary in many ways. First of all, it established water as the carrier medium of the cholera disease. This was counter to the widely, yet mistaken, belief at that time that "bad air" was the culprit responsible for the spread of the cholera disease. Secondly, it laid the foundation for the field of epidemiological research which today makes heavy use of GIS technologies.



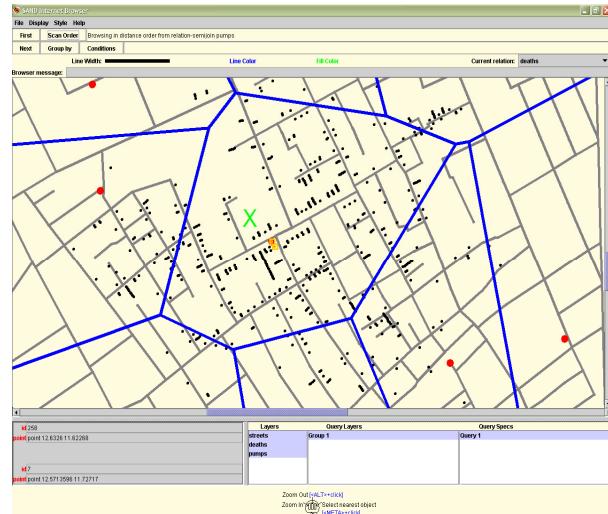
In order to demonstrate the utility of some of the operations that we have developed in the course of our research on spatial spreadsheets and browser, the figure below illustrates how we recreated the 1854 experiment of Dr. John Snow using the SAND Internet Browser [1,3]. We first created the road layer from a map image of Soho from around the same time period. Next, we overlaid the position of the water pumps on the map. Each death occurring in a Soho household at that time was recorded as a unique point. For example, if a household h recorded c deaths due to cholera, we placed c unique points at the position corresponding to the location of h on the map.



We can make this cause/effect deduction using the SAND Internet Browser by computing the distance semi-join [2] of the points in the death relation with the points in the pump relation as shown below. The distance semi-join operator uniquely associates each incidence of deaths with the nearest pump on the map



The thick polygonal subdivision lines below illustrate the result of drawing the discrete Voronoi diagram for the locations of the water pumps. The result of the distance semi-join operation yields the explicit association of the incidents of deaths with the closest water pump from which it is easy to see that the Voronoi cell that contains pump X as its Voronoi site has the most incidents of deaths.



- [1] F. Brabec and H. Samet. Client-based spatial browsing on the world wide web. IEEE Internet Computing, 11(1):52-59, January/February 2007.
- [2] G. R. Hjaltason, H. Samet. Incremental distance join algorithms for spatial databases. In Proceedings of the ACM SIGMOD Conference, pages 237-248, Seattle, WA, June
- [3] H. Samet, H. Alborzi, F. Brabec, C. Esperança, G. R. Hjaltason, F. Morgan, and E. Tanin. Use of the SAND spatial browser for digital government applications. Communications of the ACM, 46(1):63-66, Jan. 2003.
- [4] J. Snow. On the Mode of Communication of Cholera. John Churchill, London, England, second edition, 1855.

SAND Internet Browser is available from <http://www.cs.umd.edu/~brabec/sandjava/>