

The objective of this project was to use D3 to create an interactive version of Dr. John Snow's map of the 1854 London cholera outbreak using D3. The data used to generate these visualizations is mostly designed to mimic Dr. Snow's map, however it is worth noting that the gender and age of the victims was randomized based on data from the Naples cholera outbreak in 1884. All design requirements, as well as descriptions of the datasets used for this project are outlined on this [page](#).

The design of this project began with planning the layout of the required map and charts. The first two graphical elements that were planned were the map showing the location of the pumps and the deaths and the timeline graph showing the number of deaths per day. In order to ensure the viewer could clearly see the map in its entirety and clearly see the locations of the deaths and pumps at a perceptible size on a screen, I chose to make the map rather large. As this is the main feature of the project it seemed appropriate to highlight it front and center on the page. Below you can see an initial drawing of the planned layout (Figure 1). This was adjusted as other design elements were considered and implemented.

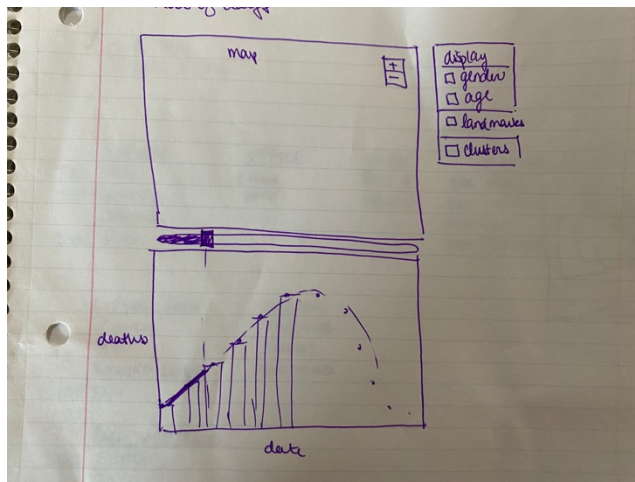


Figure 1

When creating the map, it was important that three key elements were clearly visible: the streets, pumps, and locations of the deaths. The overall color pallet of the entire page was selected based on a suggested colorblind safe scheme from [ColorBrewer](#). In Figure 2, you can see the final webpage displayed in normal color as well as if viewed by an individual with Protantopia (red-green color blindness). The streets were drawn as black lines to make them distinct, clear, and easy to see. The pumps are represented by blue squares with a white outline to make them distinct from the roads and deaths

around them. For the initial render of the visualization, each death is represented by a small orange circle, but this can be changed based upon options selected by the viewer.



Figure 2

To create the chart of deaths per day, for ease of visibility and to implement some mouseover methods with color changes, I chose to represent this data as a bar chart. I believe this makes it

easier for the viewer to see which date they are hovering over, as well as creates a larger clickable area for when the viewer would like to set the date filter on the data.

To keep the visualization interactive, I wanted the viewer to be able to choose which of the attributes of the deaths (gender, age) to display on the map. In order to do this, all deaths are initially plotted as similar orange circles. If the viewer would like to display the gender, they can select the associated checkbox and this will cause the male deaths to be displayed by open circles while females are represented by filled in circles. Likewise, to display the age, if the viewer selects the corresponding checkbox, the colors of the circles will change to match those shown on the age graph. This color change will apply to the circle regardless of if the gender box is checked, so if gender is showing the male deaths will continue to be represented by open circles while the female deaths will be filled in circles of the corresponding color of the age graph. This allows the viewer to see just the gender, just the age, or both displayed on the map. While changing the shape of each death to represent one of the attributes was considered, ultimately each of the deaths is the same type of datapoint and I chose to maintain that consistency of keeping all the deaths represented as some sort of circle.

Once the filtering was implemented, the viewer can select different days to see how the outbreak has progressed from the start up until that point. This filtering updates the deaths, as well as the charts showing the distribution of the deaths by both gender and age group from the start of the outbreak until and including the selected date. This allows the viewer to explore how the different groups were impacted over time throughout the outbreak. One interesting observation from exploring the data with that functionality was that the split between deaths of men and women remains mostly equal throughout the entire outbreak. However, the age groups typically considered most at risk (0-10, <80) were most effected from days 13 through the end of the outbreak (Figure 3).

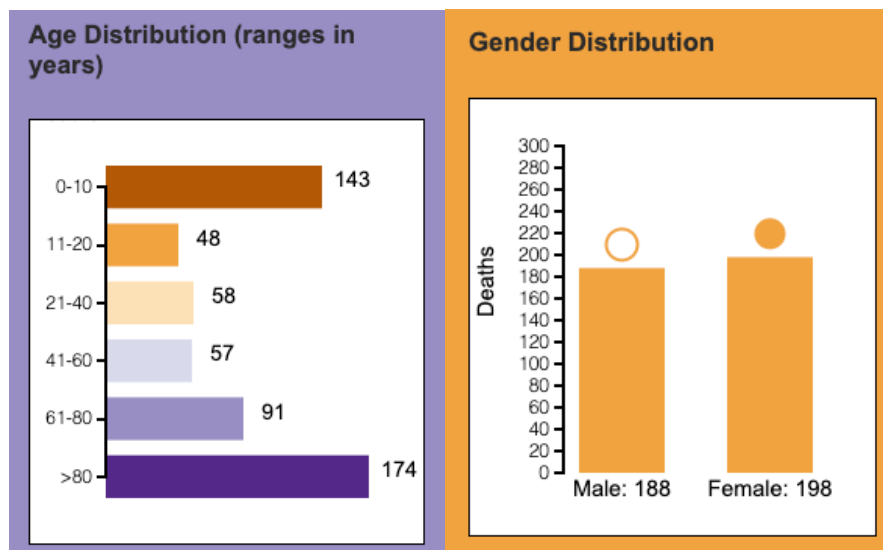


Figure 3