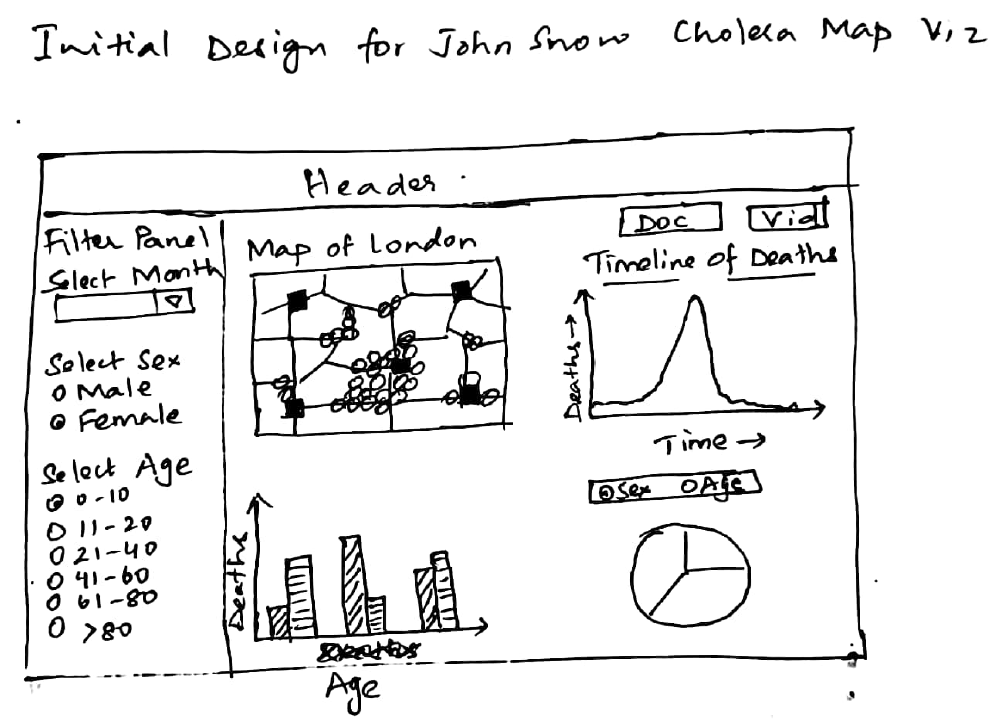
**John Snow’s Cholera Epidemic Viz in D3**

**Design Process:**

The below sketch is how I initially planned to develop the visualization:



Apart from the map showing the locations of the deaths, I planned to make use of a line chart to show the timeline of deaths since we can very easily see trends from a line chart. A bar chart was drawn to show the distribution of deaths by age buckets and sex for easier comparison between genders in each age bucket. Pie charts were used to show the overall effect of the disease on age or sex individually. The age and sex pie charts were to be controlled with the help of a radio button to toggle the view.

The left of the visualization was to have a filter panel using which the entire visualization can be filtered based on month of death, sex or age of the deceased.

This layout could not be entirely achieved due to time restrictions and difficulties in coding the visualization. The completed layout does not contain the filter panel or the radio buttons. A legend was added to understand the colors used on the map and charts.

**Design Choices:**

Layout:

I chose to place the timeline graph on the right of the map for easy visibility during interaction – mouse-over on a particular date on the timeline only shows those deaths on the map which occurred prior to or on the selected date on the line chart.

The bar and pie charts offer additional data and need not be viewed along with the map. Therefore, I chose to place them below the map.

There was difficulty implementing a radio button to control the toggling of the pie charts showing deaths by sex and age. Hence, I decided to control their visibility using buttons placed at the top of the visualization. I also use these buttons to color code the deaths according to age and sex on the map by toggling between them.

Color:

Since most of the data is categorical, I have chosen categorical or diverging-sequential colors. I chose blue and pink colors to display data by sex (using the stereotypical blue color for male and pink for female for easier visual understanding). I chose a set of diverging orange-purple diverging colors to code deaths by age buckets. The same color schemes are used to code data in the map as well in the bar and pie charts so that the data can be easily related between charts.

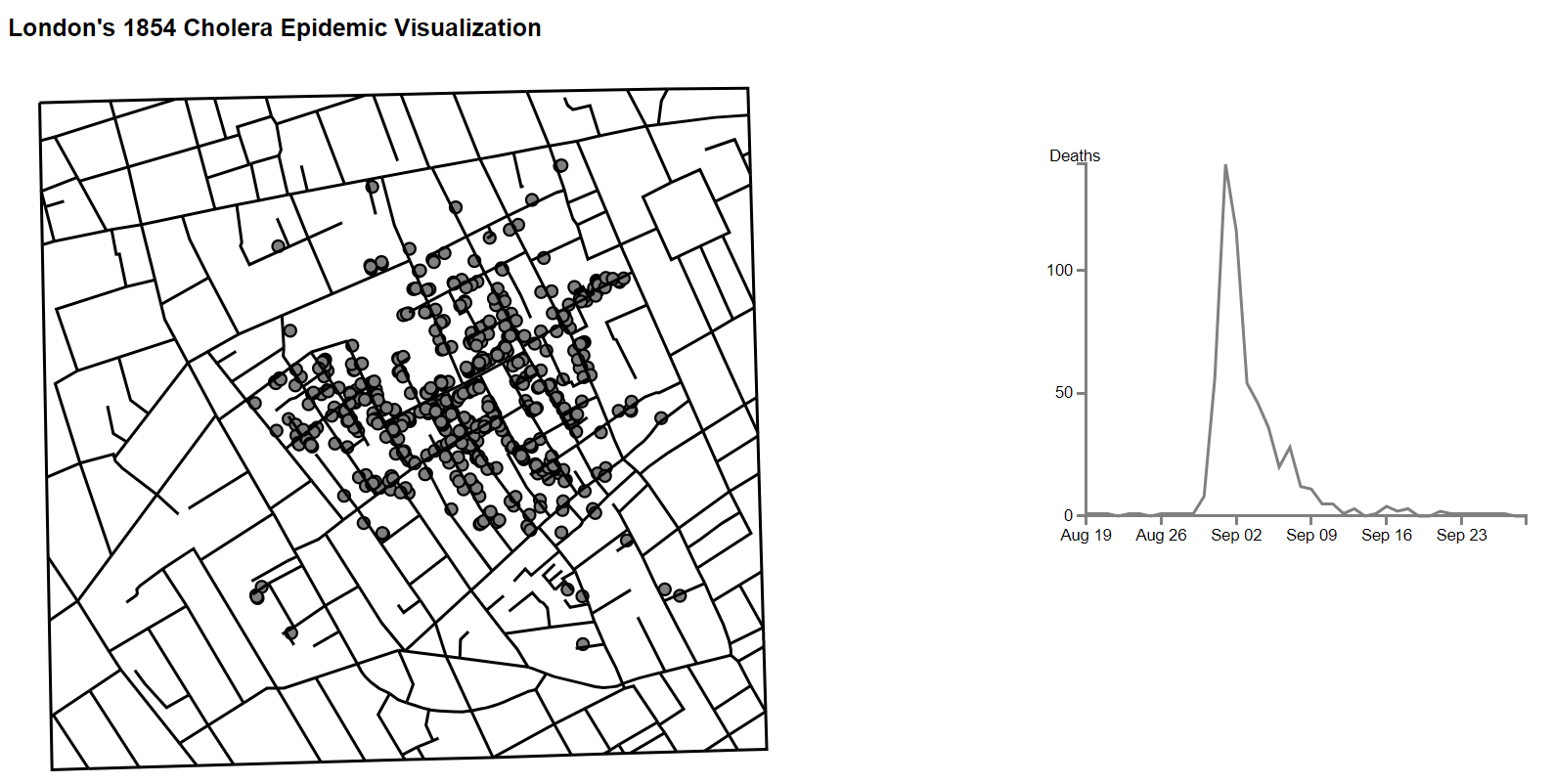
All colors chosen are color blind safe (from [www.colorbrewer2.org](http://www.colorbrewer2.org)).

Data:

This visualization uses a json file to plot the streets in the map and csv files to plot the deaths and location of the pumps on the map, the number of deaths on each day and death classification by age and gender. No modifications were done to the files or to the data stored in them.

Visualization development:

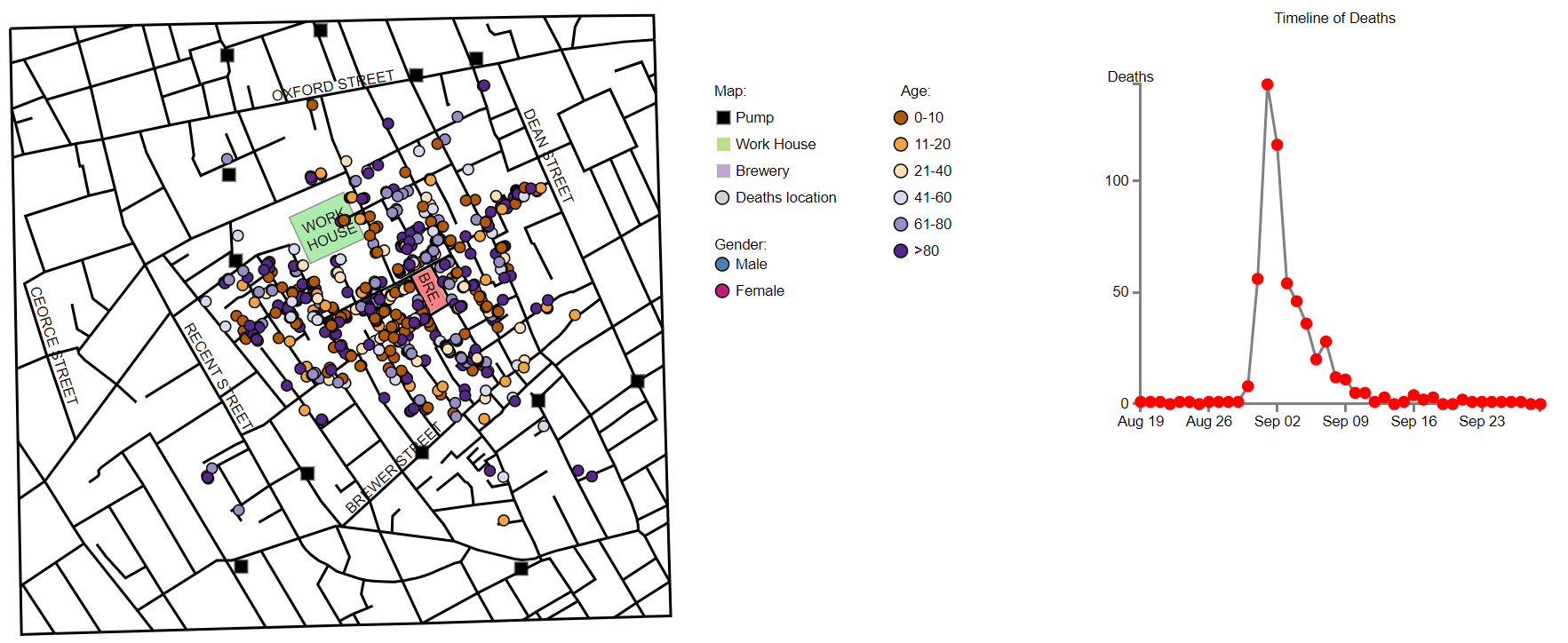
The screenshot below depicts the initial stages of the layout of the visualization with the deaths plotted on the map and the line chart.



The color used to plot the deaths was made lighter so that the plotted points can be more easily perceived against the black streets of the map. Circles were also drawn on the line chart to clearly see the days on which deaths occurred and to enable mouse-over on the timeline.

Mouse over was implemented on the line chart – hovering over any date on the line chart will filter the data on the map to only show deaths that occurred on or before that date.

On the completed viz, without filtering by hovering on the line chart:



After filtering by hovering on 31st August:



The bar chart shows data for each age group separated by sex as well. From this chart we can analyze to see if cholera is more prevalent amongst a particular gender or affects a particular gender more in any given age group.

The pie charts show deaths by sex and age individually. From these charts we can draw conclusions on which gender is more affected overall or about which age groups were more susceptible to contracting the then fatal disease.

**Analysis from the visualization:**

From the visualization, we see that, as concluded by Dr. Snow, there is a higher concentration of deaths around a few particular water pumps, especially the pump right in the middle of the map. The number of deaths also rise quite rapidly until the 1st of September. Beyond that date, there is a steady decline in the number of casualties, probably due to steps taken to replace the water pump or due to better prevention and hygiene measures.

We also see from the bar graph that there are more casualties who are 80 years and older. The second highest age category for deaths is children less than 10 years of age. Understandably, people in these age categories have weaker immune systems and are more susceptible to the disease. In all other age groups, males seem to be more affected than females. Therefore, we may be able to conclude that females could have had better resistance to the disease than men at the time.

Overall however, the number of deaths in males and females is almost equal, as can be seen form the pie chart depicting distribution of deaths according to sex.

**References:**

<http://colorbrewer2.org/#type=diverging&scheme=PuOr&n=6>

https://www.color-blindness.com/coblis-color-blindness-simulator/

<https://github.com/thfield/johnsnow>

http://thfield.github.io/johnsnow/

http://bl.ocks.org/enjalot/1203641

<https://github.com/zhuangke16/zhuangke16.github.io>

<https://www.dashingd3js.com/lessons/basic-chart-pie-chart>