

Source: <https://www.washingtonpost.com/graphics/2018/local/school-shootings-database/>

underlying data: <https://github.com/washingtonpost/data-school-shootings>

This visualization consists of a helix along which red dots are placed to indicate incidents of school shootings in chronological order proceeding counter-clockwise. Beginnings of years are labeled on the helix. Four well known incidents appear as fixed text inside the helix.

At either side of the helix are clusters of black dots. As the viewer hovers over a cluster, a pop up displays the school name, the date of the shooting, and the number of children in school. Each dot in a cluster represents ten children in school. The placement of each cluster appears dictated by where it can fit near the incident on the helix.

Although the visualization is striking and its use of pop ups is not intrusive, I confess I do not like it. It uses a lot of real estate to say not much. Its impact is more numbing than illustrative.

I think the visualization could be greatly improved. Since the data contain the geolocation of each school, overlaying each incident on a US. map would provide another meaningful dimension of data with no additional effort or cost. The size of the plotting symbol could represent the number of children in school. The color level filling each symbol could indicate the date of its corresponding incident. Spreading the incidents across the map could possibly allow enough room to annotate each incident with the school name. This would eliminate the need to use pop ups. Noting the median number of children in school and the median days between incidents would be illuminating.

I think that some perspective is necessary. Although school shootings are horrific, their number is minuscule compared to the total incidents of gun violence. This proportion should be noted on the visualization. The Post has collected a homicide data set which includes geolocation. From census data and school enrollments, a count and distribution of children nearby could be estimated.

I think there are problems with the interpretation of the underlying data:

* The visualization indicates “children in school” whereas the data record each school’s enrollment. It is easy to misinterpret “children in school” to mean children present at the time of the incident. Why not use “children enrolled” if that is what is available?
* The data almost certainly underestimate the number of children affected by incidents of gun violence at schools. Do the children have siblings? Couldn’t this be imputed from census data? The data record only incidents occurring in schools. What about gun violence occurring on adjacent streets during school hours?