**Analysis of the AIDs mortality rate in the 80’s, 90’s and early 2000’s**

**By Jacob Letterman**

**Data: AIDS mortality**

       When thinking about raw data, many will think of votes being counted in states or the amount of money that a company spent in the past few decades. When asked to look for a raw data file, I nearly went the same way that many others did, until I thought about what I truly wanted to look at. Coming from a medical background, as I nearly majored in medicine, I started looking for any medical raw data that was out there. Some were easy to find, number of hospitals built in the past few decades or how much was spent in a certain state for medical purposes, but I wanted something that made me want to think about the questions that could be asked from it. That’s when I found data about the mortality of the AIDS virus. This immediately piqued my interest as this was the type of raw data that I was looking for, one that made me think of the stories and questions this data could answer.

       The data is of exactly how it sounds, a look at the mortality of the AIDS virus from before 1981 to the end of 2001. The choice to lump all the data of 1980 and before is due to how the virus spread in the 80’s, with it being labeled as an epidemic in the middle of the decade. It is also proven when looking at the data, with only 100 people being diagnosed and 30 dying in that span of time and every other data point having over 100% more infections or deaths than that point. This data was compiled by the CDC, looking at the infections and deaths of adults/adolescents and children under 13. This was most likely compiled this way due to how the virus was being spread among these groups in the mid to late 80’s before almost spiraling out of control in the early 90’s.

       The story of this data is quite sad when you think about it. This data isn’t about money lost or gained, or even about votes gathered during an election, this was about people getting infected with a very harmful disease and potentially dying. If you know Freddy Mercury, he is probably one of the most famous people to have the disease, and if you seen photos or videos during his last few years, you’d see how much his body had succumbed to it, become very pale and nearly able to see the bones under his skin. Realizing that, it was hard to envision this data being about people suffering and dying from this disease, but that was the main story this data was telling.

       When first looking at the data, there were a few questions that I wanted to ask. One was when the spike of the epidemic took place, more specifically where is the point that the greatest increase of infections/deaths happened. I lumped infections and deaths together as, at least in this case, they both grew at the same rate that the other increased. This was quite easy to find due to the data being very clean and easy to see every data point, leading to the answer that the spike was in 1992, as that was the biggest increase in infections, while 1994 was the biggest increase for deaths. Another question that I was looking at was when did the fall of the infections/deaths start. This question was easily answered when I put the data on to a graph, putting both the adults and children data points together to see the highs and lows, leading to the answer of 1994 for adults and 1995 for children. The final question that I wanted to ask was when the biggest downfall in cases was. This question was created when answering the previous question, as well as looking at the graph of the data. Looking at the graph, you see a sharp decline in the number of infections/deaths, starting in the mid 1990’s and leading up to 2001, where the cases are steadily going down.

       The visualizations that I wanted were ones where you could clearly see the data points for each group, and clearly see where they each stand against one another. The first visual I chose was a graph comparing the infection and death in adults. A positive of this graph is that it shows the infections vs. deaths really well, putting each infection with the number of deaths that happen in that year. A negative is that it doesn’t have the year stated out, due to how excel would put it on the graph as data, instead of on the x-axis. The next visual I chose was a bar graph. This bar graph puts every point with the respective year and the respective infection/ death number. A positive is that it puts every point with each other, showing how much worse the adult figures are compared to the adolescent figures. A negative is that you can barely see the adolescent numbers, but this is more of a problem due to the numbers, not purely because of the graph. The next graph shows how much bigger the number of infections compared to the number of deaths in adults. The positives and negatives are the same as the last graph, again the choices of visuals for this data are slim when looking at graphs due to the big difference between adults and adolescents when they are compared together. The final visual I chose was a graph for the adolescent set of data. I did this as I wanted to use the data from the adolescent side as it did not get any chance to show its significance to the story of the data. The positive to this graph is that you can clearly see the infections and deaths as the years go on and you can also clearly see the rise and fall of both groups. The problem with this graph, however, is that though you can see how many years this data takes place in, it would have been better to see the years as it takes place.

       The story that this data creates is helped very much by the background knowledge of the AIDS epidemic, but even without it, the data still creates a compelling story, one that will stay with me long after this year has ended.