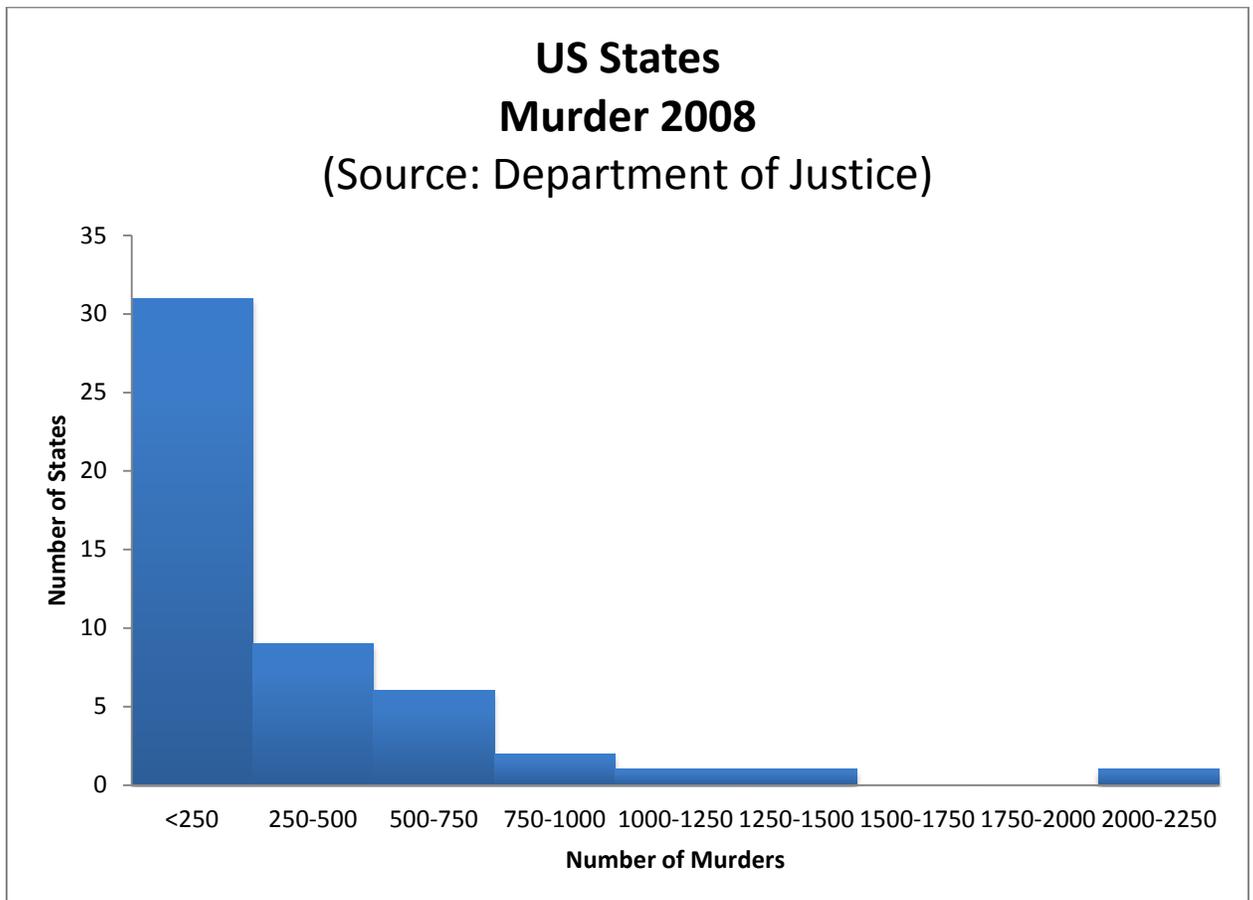


## Histograms and data manipulation

- 1) The population distribution is heavily population with values under 6 million (0-6), and a few under 12 million (6-12). There are some outliers from 12 million up to over 30 million (12-30+). The data is positively skewed, which will shift the mean higher than the median, and ranges from 0 to and even past 30 million. A typical state population will most likely fall in the 0-6 million range, which has a median of 3 million.
- 2) Based on the histogram is it safe to say the median will lie closer to Q1 than the middle due to its positive skew. The median is probably somewhere close to 5 million.
- 3) The mean state population is expected to be higher due to the few, but still important, outliers in the 12-30 million range. The median is lower due to the more frequent low numbers than high because it is based on the number of states recorded rather than the actual values that the mean is calculated with.
- 4)



- 5) For this histogram I felt an interval of 500 would nicely divide up the range of values that all fell between 0 and 2500 but there were so many that fell under 500 that the table didn't show that much information. A smaller interval like 250 helped break up the one huge section into two new ones; one is still large but not nearly as large as it was before. I suppose either interval would have worked though because with this one there are still similar extreme outliers when looking at it as a whole.
- 6) When comparing the Murder and population histograms the same shape appears on both. Both have a large proportion of states with lower numbers. They are similar because logically a state with a lower population is expected to not have as many murders as one with a larger population. This means the population of a state can determine the outcome of the number of murders in that state.
- 7) After calculating the Rates Per 100k we see DC having a rate higher than other states for a few of the categories. This may be considered an unfair comparison as it can be considered an outlier. It is important in terms of an overall scale, but on a state or local scale it is nothing like the others.
- 8) Using data from DC is much different than those of other states due to its urban setting. States that contain larger cities are bound to have more murders than those with smaller cities. When we look at DC, which is made up of mostly a city itself it is expected to have higher crime rates than the states who might contain a city similar to DC but also count the population in the spread out rural area where crime is much lower, lowering the overall crime densities for other states. This would be like comparing St Louis' rate to Missouri's rate, they will be much different.
- 9) The first step in calculating the rates compared to the national rate is to first calculate the national rate. This would include all of the state's crimes added up to produce a nation's crime count and divided by the total of all the state's populations (national population), finally multiplied by 100,000 to come up with the rate per 100k. A state's rate compared to the national rate can be determined by using the specific state's rate per 100k divided by the nation's rate for that crime, finally multiplied by 100.