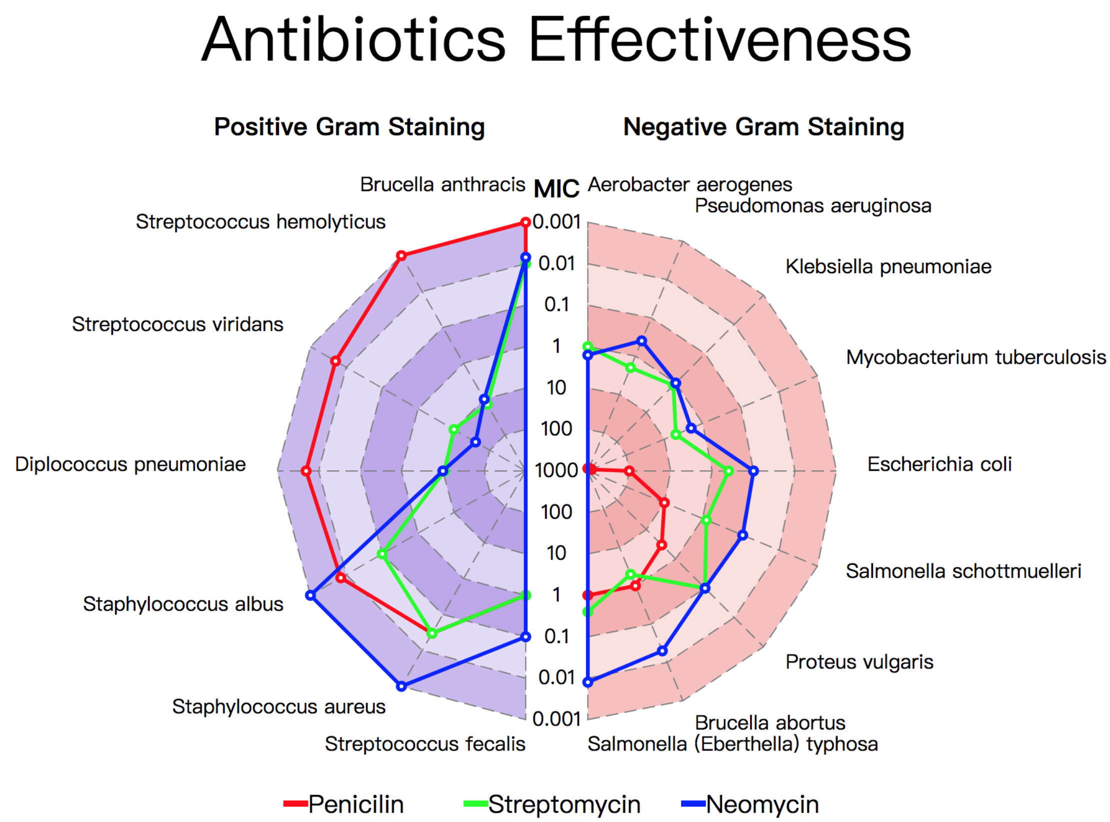
Write-up for Assignment 1

Yiqun Sun

1 Oct 2019



There are 4 data variables: Antibiotic, Bacteria, Gram Staining and MIC. Antibiotic, Bacteria, Gram Staining are nominal variable and MIC is quantitative ratio variable. Each bacteria has a Gram Staining result which is positive or negative, and MIC data for 3 antibiotics. The purpose of this study is to see which antibiotics works for which bacteria, and possibly to see how Gram Staining result affect the effectiveness of antibiotics. That reminds me of the radar chart, which is usually used in games to compare strength and weakness between racing cars or weapons. In a radar chart, there are a few available visual encoding: position (angle and distance from center), hue, length and so on. And I decided to map Bacteria to angles (spokes), map MIC to distance/length from the center, map antibiotic to hue by assigning different color lines to them, and map Gram Staining to position and hue by grouping positive and negative bacteria together and assigning different color background to the positives and negatives.

I made a few data transformation and arrangement to suit the need of visualization. Firstly, the MIC data is mapped to position/length logarithmically instead of linearly. This transformation is based on the fact that the range of the data is pretty large, and the largest data (870) is almost 10^6 times larger than the smallest data (0.001). Directly mapping MIC linearly to length makes it hard to read and compare the small data as they will look like 0. Mapping them logarithmically will reduce the difference from 10^6 times to around 6 times (because $log_{10}1000 - log_{10}0.001 = 6$). Secondly, I applied a pre-plot sorting step. The keys are Gram Staining, MIC for Penicilin, MIC for Streptomycin and MIC for Neomycin, ordered from primary to least significant. The consideration to do this sorting is to make the graph looks a bit more "smooth" after connecting the dots with lines. Thirdly, the positive bacteria and negative bacteria is separately mapped to different halves of the radar chart, although there are 9 negative bacteria but only 7 positive bacteria. This is a trade off between keeping the number of bacteria in each Gram Staining class proportional to the real number, or emphasize the comparison between the two Gram Staining result classes. Another reason to separate them into halves is it will be difficult to color them as it is bad to just color the spikes but not the background, and combining them together will make the coloring of the background confusing, as we need to introduce some unnecessary boundary lines between spikes when the color changes at the boundary line of two groups. That is also the reason why I split the two halves instead of plotting them on one circle. Lastly, I plotted the MIC in inverse way, which the outer circle means smaller MIC (0.001), and the inner circle means larger MIC (1000). This is based on the consideration that usual radar charts is represented in a "bigger-better" way, and to avoid confusing, the better (smaller MIC) the antibiotic, the longer (outer) the length.

The message this chart tries to convey is that which antibiotic is the best (or better) for which kind of bacteria. We can clearly make comparison of the effectiveness of antibiotics on different type of bacteria, and that is the most important message of this chart. We can see Neomycin works well for negative bacteria, Penicilin works well for positive bacteria, and Streptomycin is somewhere between both. Overall, Neomycin seems to work better on most of the bacteria. That is maybe just the simple conclusion seen by non-professionals (like me). Professionals can make further decisions based on knowledge from their domain, like what is the normal used concentration and etc. But, the plot also degraded the number of bacteria in two Gram Staining result groups, making a advantage for antibiotics more effective in the positive than those which is more effective in the negative group. Because the positive group has fewer bacteria, thus each of the bacteria has a bigger area than those in the negative group, and doing well in positive bacteria means occupying more areas, making a visual effect of doing much better. For example, a antibiotic which do the best in 4 positive bacteria will have a even stronger visual impression than those which do the best in 5 negative bacteria. This might introduce a bias when someone who try to find the only best antibiotics among the three, that he or she may be misled to pick the antibiotics which do better for positive bacteria.

The tool I use to generate the graph is D3.JS, and the code is pubilished here: https://github.com/dukesun99/Information-Visualization, and the online view is here: https://dukesun99.github.io/Information-Visualization/. The referenced code is cited in the source code.