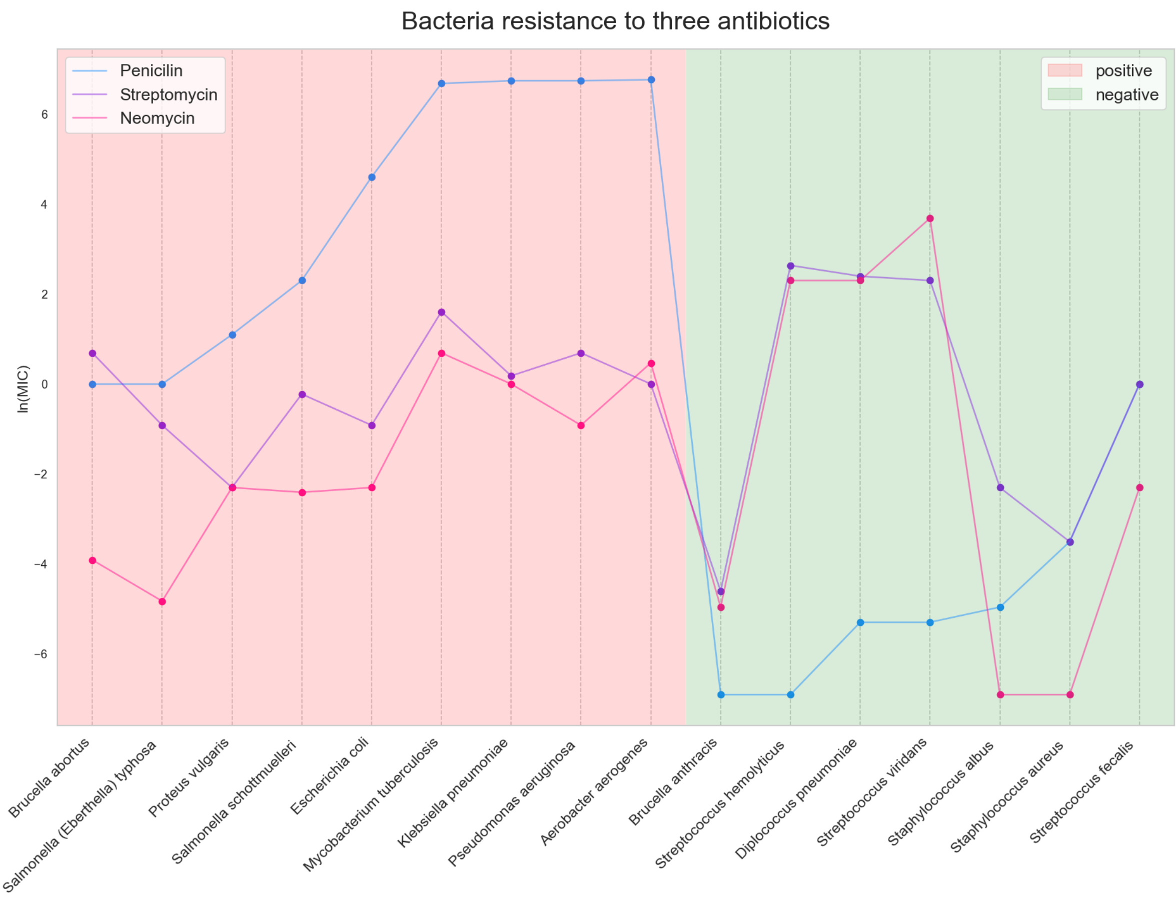
CS 503 – Data Visualization Assignment 1

Yi Ren (002269013), Wentao Lu (002276355)

Visualization 1



In this graph, we want to visualize and emphasize bacteria's resistance to the three antibiotics.

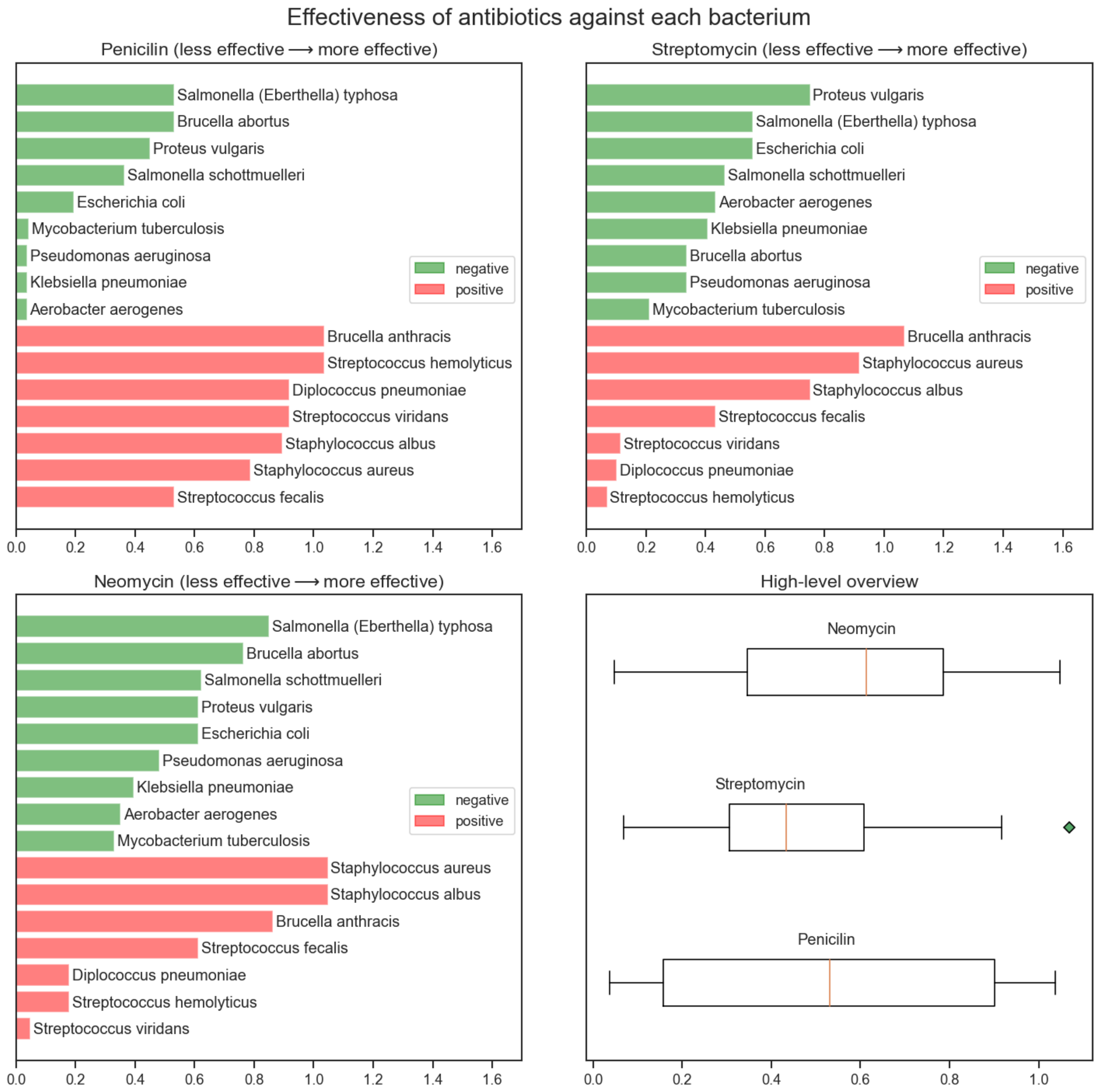
First, we sort all bacteria by staining states, and further sort them by their Penicilin MIC values, both in ascending order. This way, the data points are separated into two regions (positive or negative), sorted by Penicilin MIC to exhibit some patterns. Besides, to eliminate the large disparity in data, we take the natural logrithm of the raw MIC values so that data is relatively on the same scale.

In concrete, the sorted bacteria is on the x-axis, the logrithm of MIC is on the y-axis, which indicates bacteria's resistance to a certain type of antibiotic. The higher the y value, the more doses of antibiotic are required to inhibit that bacterium, so the bacterium is more resistant. Each data point and connected line segment for a particular antibiotic uses a different color, so that it's easy to distinguish between them. The region with red background are gram-positive bacteria, green background are gram-negative bacteria. To convey information effectively, these elements are concluded in two legends and easy to understand. In addition, the colors are transparent by setting alphas as appropriate, to prevent readers' eyes from getting exhausted.

From this graph, we can easily find which antibiotic is the most effective (lowest y value) for each bacterium on the x axis, which bacterium is very hard to kill (all 3 y values are high), as well as how effective a certain antibiotic is against all bacteria. For example, the blue line shows that Penicilin is not very effective for gram-positive bacteria since its y values are hign in the red region, or in other words, positive bacteria tend to be resistant to Penicilin.

However, given that the line is very low in the green region, we can infer that Penicilin is a panacea for most gram-negative bacteria.

Visualization 2



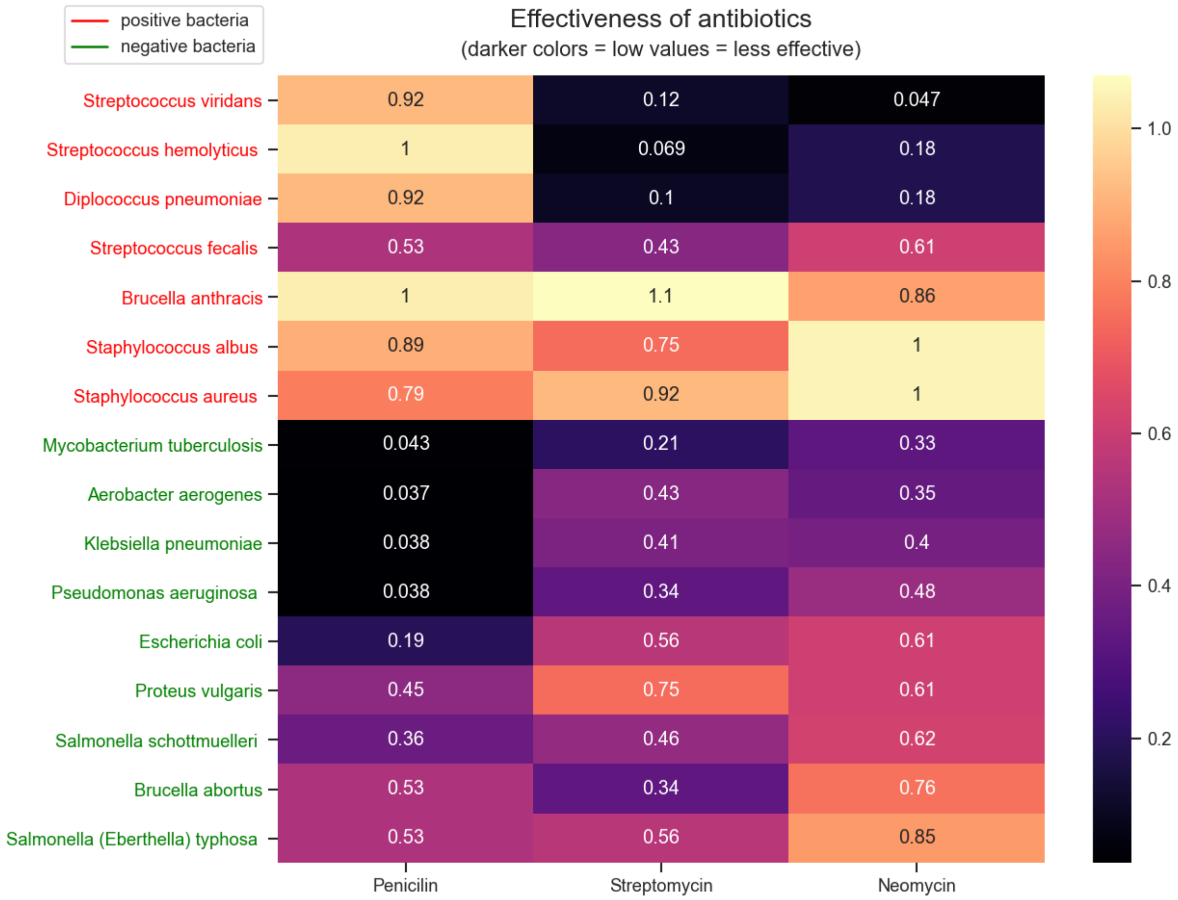
In this graph, we want to emphasize the effectiveness of antibiotics against each bacterium. To do so, we calculate the reciprocal of MIC, take the logrithm of it to reduce magnitude disparity, and then scale the data to range 0~1 via min-max normalization. This gives a rough estimate of the effectiveness. As usual, we first group the data by staining states, and then sort them based on estimated effectiveness in each group.

In this figure, there are 4 subplots. Each of the first 3 subplots shows an antibiotic's effectiveness on different bacteria, which is crystal clear by looking at the length of a bacterium bar. For example, the first subplot clearly implies that Penicilin in general is very good at killing gram-positive bacteria, wheras it's not so useful against some gram-negative bacteria.

The last subplot ignores the details of our data, but focuses on the high-level overview by comparing the box plot of each antibiotic, this gives us some insight into an antibiotic's generalizability. For example, the graph shows that Neomycin has the highest median among the three, and it's box (range between 0.25 and 0.75 quantiles) is on average to the right. Hence, we can infer that Neomycin is on average effective to most bacteria, it has less weakness than the other two, if we don't know the type of bacterium we are fighting against, taking Neomycin may be the best option.

Colors are selected to be consistent with the previous graph. To make the graph clean and neat, I've put the bacteria names on the y axis so that their long labels don't need to rotate, and I've also replaced the y labels with in-graph text annotations to increase visible graph areas.

Visualization 3



In the last graph, we use a heat map to reflect the effectiveness of antibiotics.

As previous, we use the same effectiveness metrics, grouping and sorting.

In this heat map, antibiotic's effectiveness on different bacteria is represented by cells, a cell with lower value or darker color means ineffective, and vice versa. Compared to the previous graph, this is a more vivid presentation with concrete numbers and colors. Data details are given since we want to reflect the precise effectiveness of antibiotics, but as a price, the overall information may be less intuitive.

Basically we can draw similar conclusions as before. For example, the first column is very bright for gram-positive bacteria, which indicates that Penicilin is good at killing gram-positive bacteria, but the rest of the cells tend to be dark, so it's less effective against some gram-negative bacteria. Similarly, the third column is on average bright, without too many dark cells, so that Neomycin is on average effective to most bacteria, it's more robust than the other two.

Here I'm using the magma cmap because it's monotonically increasing from dark to bright colors, this makes it easier to spot an very effective or ineffective cell. The color map is based on all data we have, so for each bacterium (row), we can easily find the most effective antibiotic (the brightest cell). In addition, a color bar is drawn to the right that helps readers interpret the relationship between colors and values.

Conclusion

As stated above, based on our design, all three graphs can provide help to doctors.

- if the specific type of bacterium is known, visualization 1 and 3 can help doctors choose the best antibiotic.
- if we only know whether the bacterium is positive or negative, visualization 1 and 2 are good references.
- if we know nothing about the bacterium, visualization 2 can give us the best guess.