

The question to be answered with this visualization is this: ***What concentration of a given type of antibiotic is required to suppress a given type of bacteria to an acceptable threshold?***

The following considerations are necessary:

- There are three types of antibiotics given as nominal variable
- There are sixteen types of bacteria given as a nominal variable
- The pairing of antibiotic and bacteria is a floating point quantitative variable between the range of $10e-3$ and $10e+3$, and this is referred to as the **Minimum Inhibitory Concentration**, or **MIC**, for the given antibiotic and bacteria
- Each bacteria is flagged as either gram-stained positive, or gram-stained negative

Because of the large range of MIC values, it makes sense to visually represent each value on a logarithmic scale. The power of the scale is a design decision to be made, but it makes sense to use a common power such as “e” or 10.

A cursory examination of the table provided for the Minimum Inhibitory Concentration (MIC) values shows that it is a relation of the two nominal variables mentioned above. The first variable is the antibiotic type. It is given along the horizontal axis of the table. The second nominal variable is the bacteria type, and it is given along the vertical axis of the table. The intersection of these two relations is a quantitative variable, the MIC. A somewhat unrelated column of the relation is the “gram staining” column, which is positive or negative for a given bacteria.

Because the intersection of antibiotic and bacteria forms a natural 2-dimensional matrix, I decided for my initial submission to retain this essential geometry of a table. I let the nominal variables for antibiotic and bacteria be the axes of position. I chose to create my submission with the R programming language after taking into consideration the recommendation that I “***...may find it most instructive to create the chart from scratch using a graphics API of your choice.***” To construct an effective representation of these data using R, I consulted the descriptions available to me at the website entitled “*The R Graph Gallery*” (<https://www.r-graph-gallery.com/>). The website allows the user to drill down to a particular type of visualization from a library, with types based upon level of abstraction. The two types that best seemed to fit these data were “heatmap,” and “correlogram.” Of the two, “heatmap” looked more promising. A heatmap is described thusly:

*“A **heat map** (or **heatmap**) is a graphical representation of data where the individual values contained in a matrix are represented as colors. It is a bit like looking at a data table from above. It is really useful to display a general view of numerical data, not to extract a specific data point.”* (<https://www.r-graph-gallery.com/heatmap/>)

In short, this looks like a good fit for the data. Of the available examples, I chose *levelmap* as the **most promising**. The *levelmap* allows the observer to quickly determine the magnitude of a value from the antibiotic/bacteria matrix using a color **value**, and the color-coded scale for the magnitude of the value is given immediately to the right of the map. The values are automatically **bucketed** to a specific hue. One shortcoming of this approach

is that the ability to distinguish between small differences within the bucketed hues is lost. However, this is made up for by the ease with which the observer is able to distinguish between differences of magnitude. For example, one can see with the bright red encoding that penicillin is a relatively ineffective antibiotic for six types of bacteria, and is only marginally effective for one other.

Note: While Mackinlay's ranking of perception tells us that hue, or color is a poor choice for effective presentation of a quantitative variable, the bucketed hues essential turn the quantitative MIC into an ordinal variable. In addition, since the hues are on a transitory spectrum from one extreme color to another, I am effectively encoding this "ordinal" MIC value with saturation. According to Mackinlay, saturation is indeed an effective visual cue for an ordinal variable.

The effective encoding of gram-staining (positive or negative) had no example code in the aforementioned library, so I was forced to ad-lib. It might have been possible to create another column in the *levelmap* for this information. However this additional column would have been logically unrelated to the values provided in the other columns, both by color and by the type of data encoded. I noted that since gram-staining is a binary value associated with bacteria type, it can be color encoded with the label color along the vertical axis. Since gram-staining is associated with a dark blue, I chose it as the font color for bacteria with that positive characteristic, and I also made the font bold. For bacteria with negative gram-staining characteristic, I left the font black and unbolded. I noted the proper interpretation of the font in the labeling of the vertical axis.

The feedback I received for the my initial submission to this assignment noted that I had neglected to provide the more logical logarithmic scale for color palette of the levelmap. In the revised submission for this assignment, I have given the color palette the preferred logarithmic scale from $10e-3$ to $10e+3$ with ticks at each power of 10, and a color change between each tick mark as well as on the tick mark. Additionally, I have drawn the viewers attention to the logarithmic nature of the color palette with a comment in the title of the visualization. The logarithmic scale brings more contrast to the different values in the table. **And note this important point: The bright red encoding for high MIC values draws attention to the fact that an indicated antibiotic is a relatively ineffective suppressor of a corresponding bacteria.** White means "good," bacteria free, and red means "bad," lots of bacteria remaining.

Using Tableau, I also sought for the revised submission to create an alternative visualization for these data. This is also now included as both a JPEG and PDF. For this new visualization, I sought to let the quantitative MIC take a more natural position value. The MIC can now be represented as a bar chart, but I now had to encode all of antibiotic, bacteria, and gram staining nominal variables along one axis. For good, comparative visualization between antibiotics, I chose to use a side-by-side bar chart. Although antibiotic is encoded along the bottom of the horizontal axis, I also chose to encode the same with color using the "Miller Stone" palette available in Tableau. Bacteria are encoded along the top of the horizontal axis, grouped by gram staining positive or negative. MIC along the vertical axis has been set on a reversed, base-10 logarithmic scale. I have used the reversed scale to indicate that lower values are better, and have shorter bars.