# Introduction:

In this Assignment, we had given a antibiotic dataset. And the goal is to use different data visualization methods to better present the data, so that the viewers can understand the information in the dataset more easiler and clear. We hope after seeing several graphs, the viewer can decide which antibotic is better than ohthers, which means that antibictic need less does to prevent virus growth.

# Outline:

1. Input the dataset and relevant packages/environment
2. Manipulate data format so we can better plot the data
3. Using different kind of graph and methods to present the data
4. Get the insight of the the graphs and decide which antibiotic is bette

# Details:

## 1. Input the dataset and relevant packages/environment:

We use jupyter nootbook, python3 kernel.

For input dataset, we can use dataFrame.read_csv funciton. However, in this assignment, we decide to use dictionary to create Pandas dataFrame object, just because the dataset is pretty small.

```
antibotic_dict = {'Bacteria':['Aerobacter aerogenes','Brucella abortus','Brucella anthracis','Diplococcus pneumoniae
                  'Penicilin':[870.0,1.0,0.001,0.005,100.0,850.0,800.0,3.0,850.0,1.0,10.0,0.007,0.03,1.0,0.001,0.005]
                  'Streptomycin':[1.0,2.0,0.01,11.0,0.4,1.2,5.0,0.1,2.0,0.4,0.8,0.1,0.03,1.0,14.0,10.0],
                  'Neomycin':[1.6,0.02,0.007,10.0,0.1,1.0,2.0,0.1,0.4,0.008,0.09,0.001,0.001,0.1,10.0,40.0],
                  'Gram_Staining':['negative','negative','positive','positive','negative','negative','negative','nega

# transfer the dictionary to pandas dataFrame
antibiotic = pd.DataFrame(antibotic_dict)
```

Then, we can import relevant packages, numpy for n-dimensionary array, pandas for Panda dataFrame, matplotlib and seaborn to plot the data

```
# import relevant packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

2. Manipulate data format so we can better plot the data:

Original dataSet looks like below, is a wide format, which is easier for human to read.

| Bacteria | Penicilin | Streptomycin | Neomycin | Gram Staining |
|---|---|---|---|---|
| Aerobacter aerogenes | 870 | 1 | 1.6 | negative |
| Brucella abortus | 1 | 2 | 0.02 | negative |
| Brucella anthracis | 0.001 | 0.01 | 0.007 | positive |
| Diplococcus pneumoniae | 0.005 | 11 | 10 | positive |
| Escherichia coli | 100 | 0.4 | 0.1 | negative |
| Klebsiella pneumoniae | 850 | 1.2 | 1 | negative |
| Mycobacterium tuberculosis | 800 | 5 | 2 | negative |
| Proteus vulgaris | 3 | 0.1 | 0.1 | negative |
| Pseudomonas aeruginosa | 850 | 2 | 0.4 | negative |
| Salmonella (Eberthella) typhosa | 1 | 0.4 | 0.008 | negative |
| Salmonella schottmuelleri | 10 | 0.8 | 0.09 | negative |
| Staphylococcus albus | 0.007 | 0.1 | 0.001 | positive |
| Staphylococcus aureus | 0.03 | 0.03 | 0.001 | positive |
| Streptococcus fecalis | 1 | 1 | 0.1 | positive |
| Streptococcus hemolyticus | 0.001 | 14 | 10 | positive |
| Streptococcus viridans | 0.005 | 10 | 40 | positive |

However, the computer prefer the Long Format which is more accessible for computer to work with, we use the dataFrame.melt function to convert the wide format into long fromat and it looks like below

```
vertical_antibiotic = antibiotic.melt(id_vars=['Bacteria','Gram_Staining'])
print(vertical_antibiotic)
```

executed in 14ms, finished 21:10:30 2022-10-07

| | Bacteria | Gram_Staining | variable | value |
|---|---|---|---|---|
| 0 | Aerobacter aerogenes | negative | Penicilin | 870.000 |
| 1 | Brucella abortus | negative | Penicilin | 1.000 |
| 2 | Brucella anthracis | positive | Penicilin | 0.001 |
| 3 | Diplococcus pneumoniae | positive | Penicilin | 0.005 |
| 4 | Escherichia coli | negative | Penicilin | 100.000 |
| 5 | Klebsiella pneumoniae | negative | Penicilin | 850.000 |
| 6 | Mycobacterium tuberculosis | negative | Penicilin | 800.000 |
| 7 | Proteus vulgaris | negative | Penicilin | 3.000 |
| 8 | Pseudomonas aeruginosa | negative | Penicilin | 850.000 |
| 9 | Salmonella (Eberthella) typhosa | negative | Penicilin | 1.000 |
| 10 | Salmonella schottmuelleri | negative | Penicilin | 10.000 |
| 11 | Staphylococcus albus | positive | Penicilin | 0.007 |
| 12 | Staphylococcus aureus | positive | Penicilin | 0.030 |
| 13 | Streptococcus fecalis | positive | Penicilin | 1.000 |
| 14 | Streptococcus hemolyticus | positive | Penicilin | 0.001 |
| 15 | Streptococcus viridans | positive | Penicilin | 0.005 |
| 16 | Aerobacter aerogenes | negative | Streptomycin | 1.000 |
| 17 | Brucella abortus | negative | Streptomycin | 2.000 |
| 18 | Brucella anthracis | positive | Streptomycin | 0.010 |
| 19 | Diplococcus pneumoniae | positive | Streptomycin | 11.000 |
| 20 | Escherichia coli | negative | Streptomycin | 0.400 |
| 21 | Klebsiella pneumoniae | negative | Streptomycin | 1.200 |
| 22 | Mycobacterium tuberculosis | negative | Streptomycin | 5.000 |
| 23 | Proteus vulgaris | negative | Streptomycin | 0.100 |
| 24 | Pseudomonas aeruginosa | negative | Streptomycin | 2.000 |
| 25 | Salmonella (Eberthella) typhosa | negative | Streptomycin | 0.400 |
| 26 | Salmonella schottmuelleri | negative | Streptomycin | 0.800 |
| 27 | Staphylococcus albus | positive | Streptomycin | 0.100 |
| 28 | Staphylococcus aureus | positive | Streptomycin | 0.030 |
| 29 | Streptococcus fecalis | positive | Streptomycin | 1.000 |
| 30 | Streptococcus hemolyticus | positive | Streptomycin | 14.000 |
| 31 | Streptococcus viridans | positive | Streptomycin | 10.000 |
| 32 | Aerobacter aerogenes | negative | Neomycin | 1.600 |
| 33 | Brucella abortus | negative | Neomycin | 0.020 |
| 34 | Brucella anthracis | positive | Neomycin | 0.007 |
| 35 | Diplococcus pneumoniae | positive | Neomycin | 10.000 |
| 36 | Escherichia coli | negative | Neomycin | 0.100 |
| 37 | Klebsiella pneumoniae | negative | Neomycin | 1.000 |
| 38 | Mycobacterium tuberculosis | negative | Neomycin | 2.000 |
| 39 | Proteus vulgaris | negative | Neomycin | 0.100 |
| 40 | Pseudomonas aeruginosa | negative | Neomycin | 0.400 |
| 41 | Salmonella (Eberthella) typhosa | negative | Neomycin | 0.008 |
| 42 | Salmonella schottmuelleri | negative | Neomycin | 0.090 |
| 43 | Staphylococcus albus | positive | Neomycin | 0.001 |
| 44 | Staphylococcus aureus | positive | Neomycin | 0.001 |
| 45 | Streptococcus fecalis | positive | Neomycin | 0.100 |
| 46 | Streptococcus hemolyticus | positive | Neomycin | 10.000 |
| 47 | Streptococcus viridans | positive | Neomycin | 40.000 |

## 3. Using different kind of graph and methods to present the data:

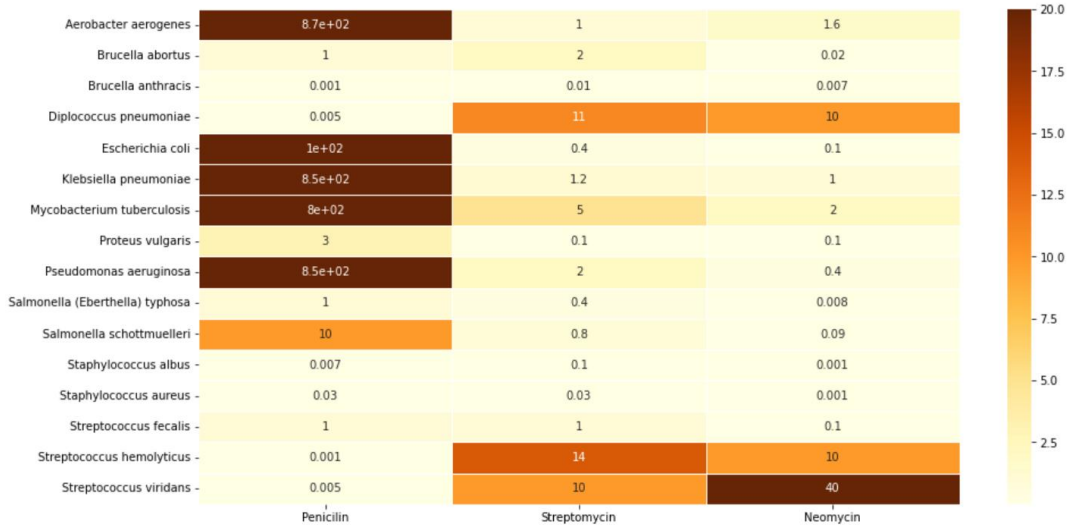
Use sns.heatmap in the oringinal dataset, set the vmin, vmax and colorMap in the order to better distinguish the different. As we can see, Penicillin has more dark part which means it needs more dose. Streptomycin and Neomycin is pretty much the same while Neomycin seems little brighter.

However, it this picture we cannot tell which Bacterial is Gram Staining positive or nagtive. The Gram Stining might be a factor that affect the result. Also, although Penicillin has some dark part, it also has very bright part.

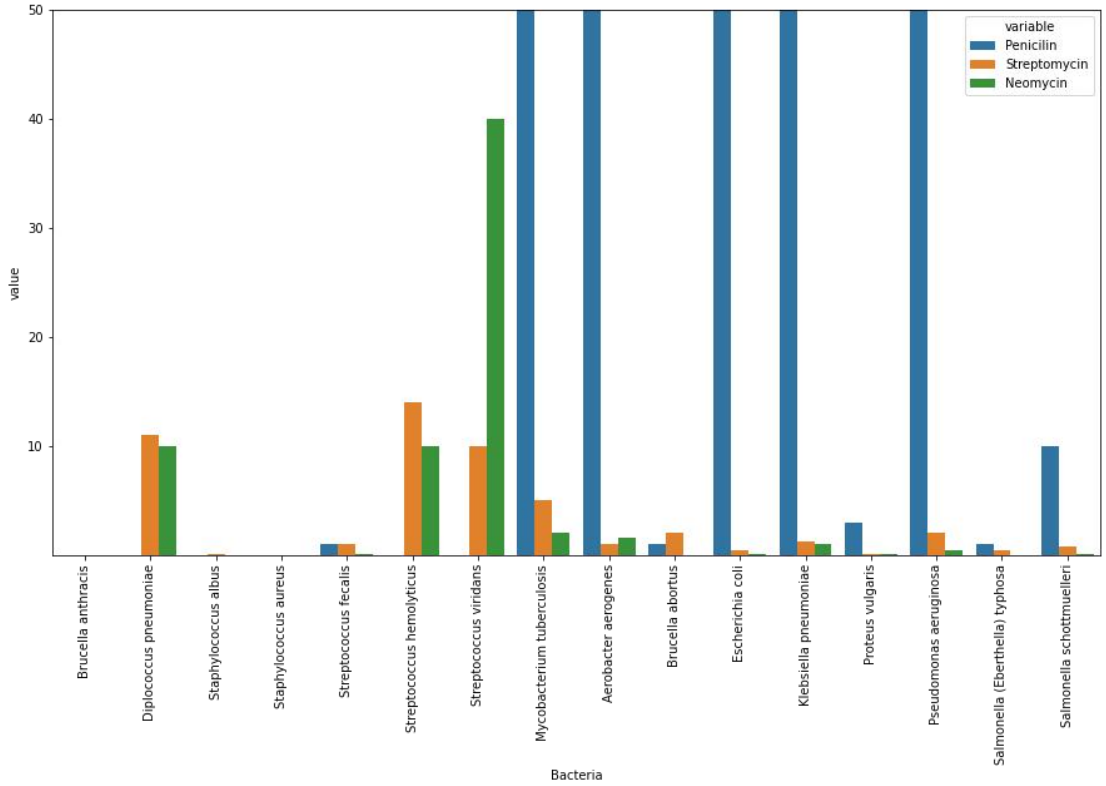
```
min_antibiotic = antibiotic[['Penicilin','Streptomycin','Neomycin']]
plt.figure(figsize=(15,8))
sns.heatmap(min_antibiotic,cmap='YlOrBr',yticklabels=antibiotic['Bacteria'],vmin= 0.001,vmax=20,annot=True,
            linecolor='white',linewidths=0.5)
```

executed in 762ms, finished 22:14:42 2022-10-07

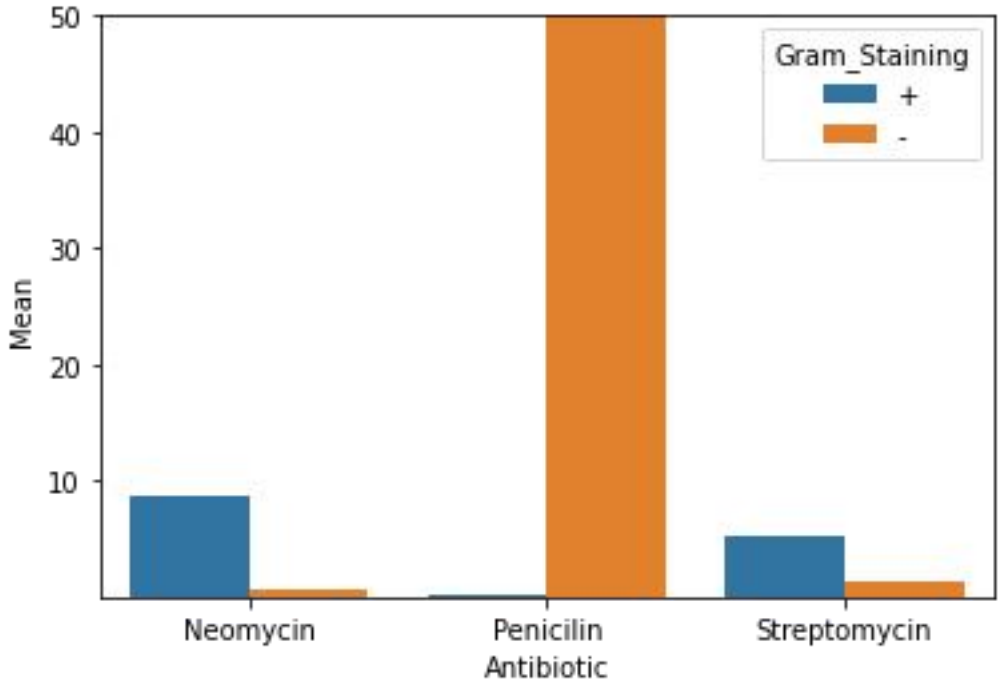
```
<AxesSubplot:>
```



We also can use barplot to show the overall dataset




Also, we can calculate the mean base on different antibiotic and get a more global view



4. Get the insight of the the graphs and decide which antibiotic is bett:

As we can see in the graphs. Overall Neomycin and Streptomycin are perform very good in 16 bacterias. And Neomycin is a little better than Strptomycin.

Also, we can see Pencillin did the best in bacteria which are gram staining positive, however, in gram staining negative, pencillin works very bad.

If we know the bacteria's gram staning is positive, we can definitely choose Pencillin. If we don't or gram staining is negative, we can use Neomycin or Streptomycin, and the first choose would be Neomycin if we don't specific the bacteria.