

# MOTION Lab Summer 2024 Code Club

## Session III: Data Visualization in R

### Part 1

September 06, 2024

#### **Presenters:**

- Laura Cowden ([laura.cowden@pennmedicine.upenn.edu](mailto:laura.cowden@pennmedicine.upenn.edu)) and
- Leigh Cressman ([crel@pennmedicine.upenn.edu](mailto:crel@pennmedicine.upenn.edu))

#### **Learning objectives:**

- 1) Become familiar with R Markdown
- 2) Become familiar with ggplot2 syntax
- 3) Understand which plot is best suited for your data
- 4) Understand how to generate graphs in R

#### **Why do we need to visualize our data?**

- Understand distribution of data (this will inform appropriate statistical test)
- Check for problems with data
- Present findings to group members
- Generate plots for manuscripts

#### **Why should we use R to visualize data?**

- Reproducibility and transparency
- Flexibility
- Standard in scientific community for manuscripts

#### **Further reading on data visualization:**

[Edward Tufte's work on displaying quantitative information](#)

## ggplot2

We will focus on this package today. Graphs can also be created in other packages (e.g., base R, lattice). ggplot2 was developed by Hadley Wickham (tidyverse) and his team.

“**gg**” stands for “grammar of graphics,” which is a general scheme for breaking up graphs into semantic components such as scales and layers.

Part of tidyverse: only need to run **library(tidyverse)**

[Additional reading on ggplot2 package](#)

## R Markdown

We are teaching this session using an R Markdown file (.Rmd) instead of the .R script files which we used in previous sessions. R Markdown is a literate programming tool: you have a source file with two representations:

- 1) Code which is understood by a compiler or interpreter
- 2) Formatted documentation which is “woven” (or “knitted”) from the literate source

R Markdown allows you to combine commentary with code to create an output document with results and graphs in various output formats (e.g., PDF, Word, HTML). This tool may be useful for you to summarize lab data and create a report for your PI.

You must surround chunks of R code with “`````”. You can include arguments to suppress the code which is run from showing up in the output document. Finally, you can use the **knitr** package to knit the document. This executes the code chunks and creates the output document.

[Additional reading on R markdown](#)

[knitr documentation](#)

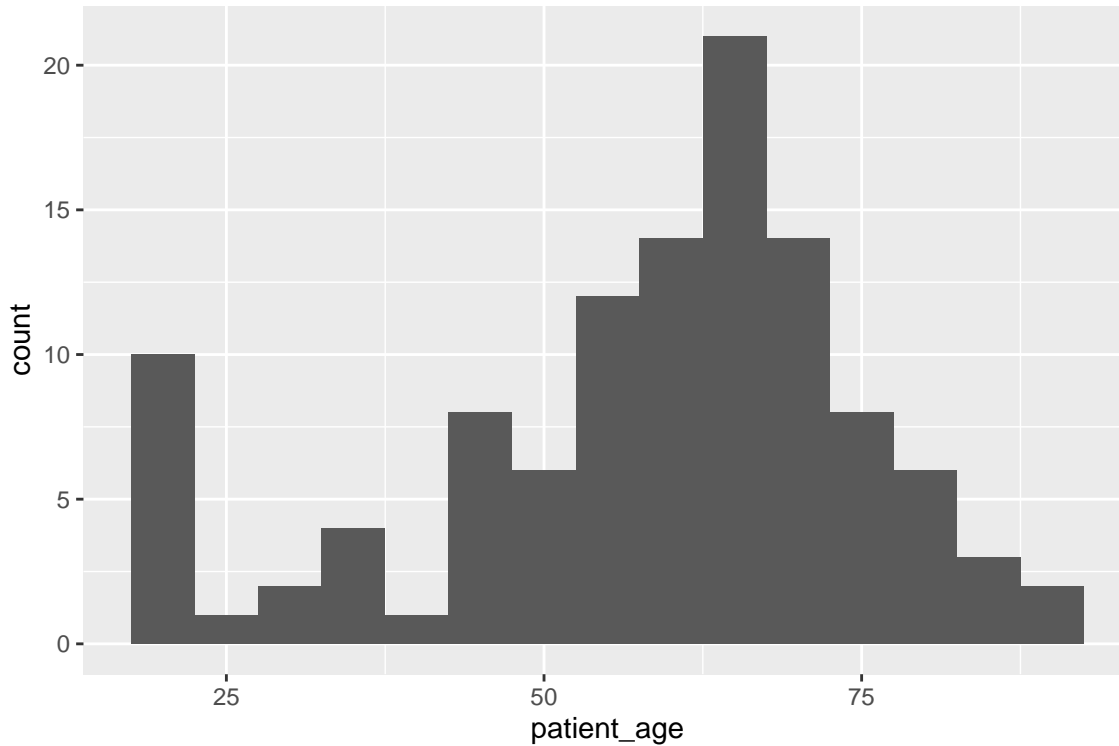
## LAURA’S SECTIONS

\*The data in this section were masked and partially fabricated for presentation purposes. The variables were originally collected for our MERLIN study.

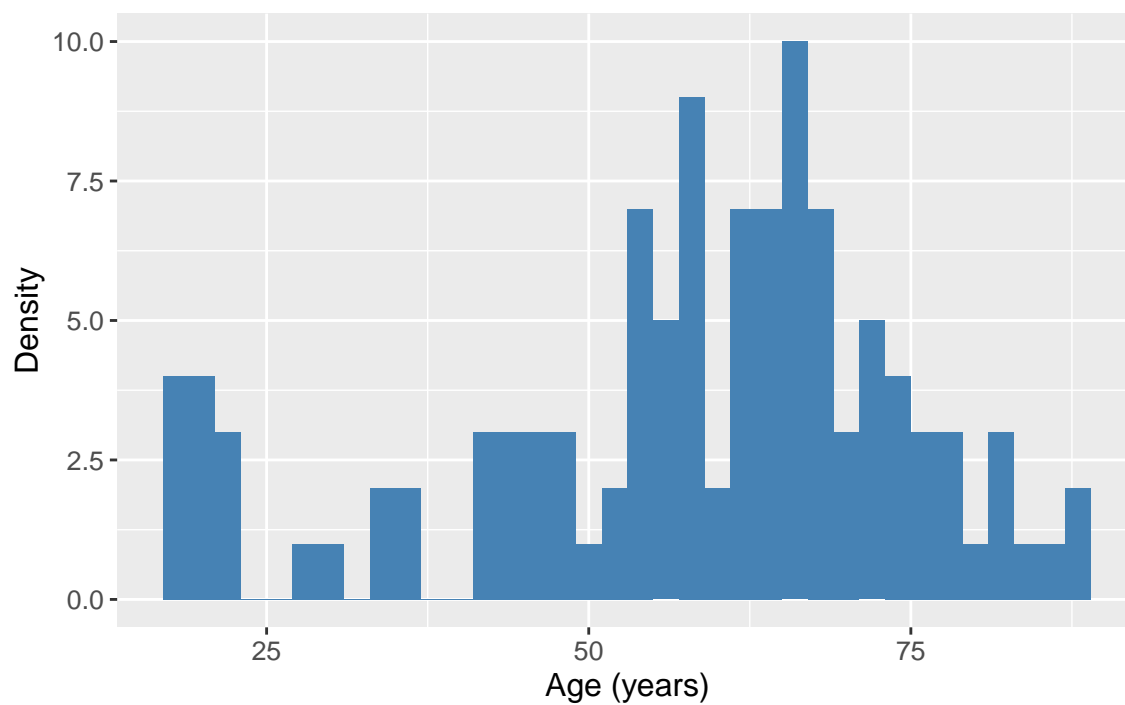
## Histograms and Density Plots

These plots are useful for examining the distribution of numeric data. Histograms split a continuous measure into bins or groups. Data points are grouped into bins.

### Basic histogram

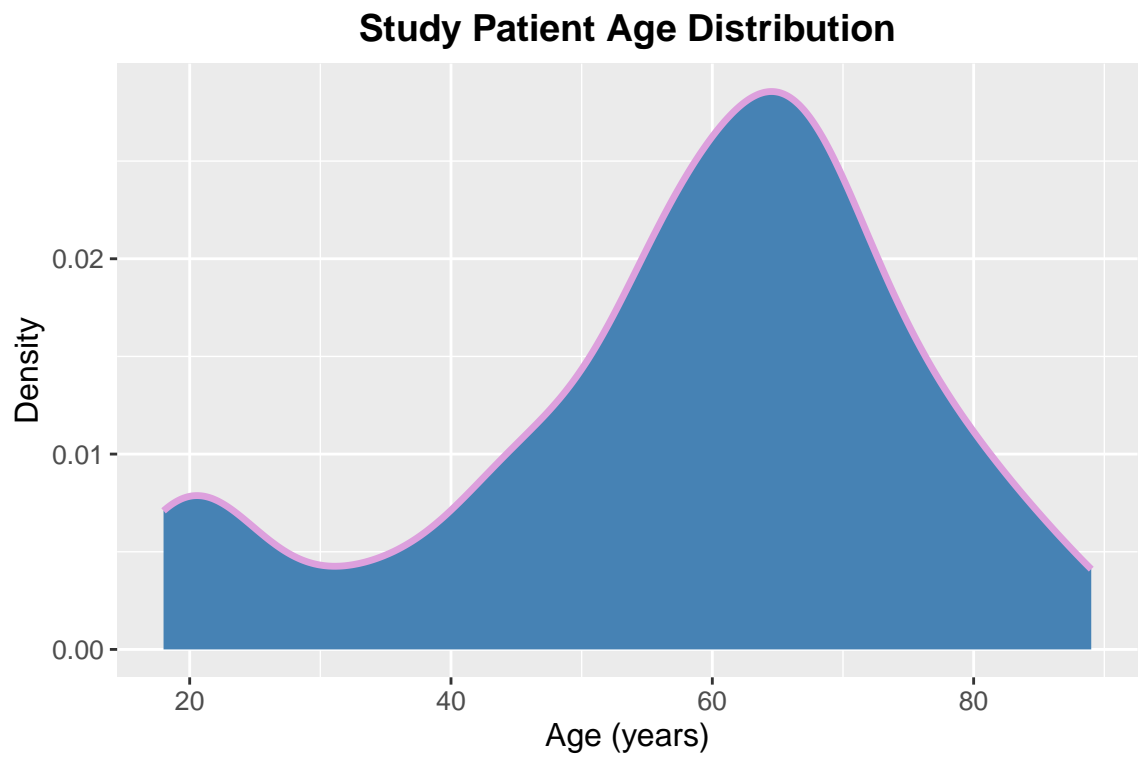


## Study Patient Age Distribution

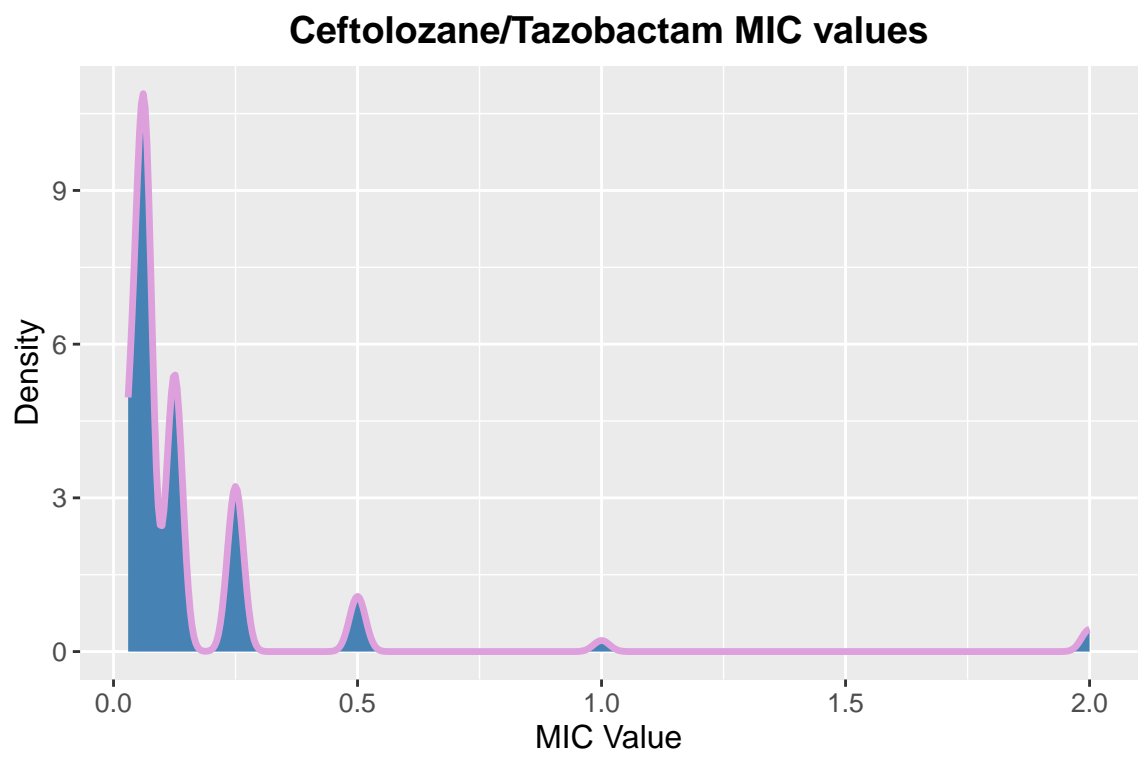


Is patient age normally distributed?

**Density Plot:** (smoothed-out version of histogram)



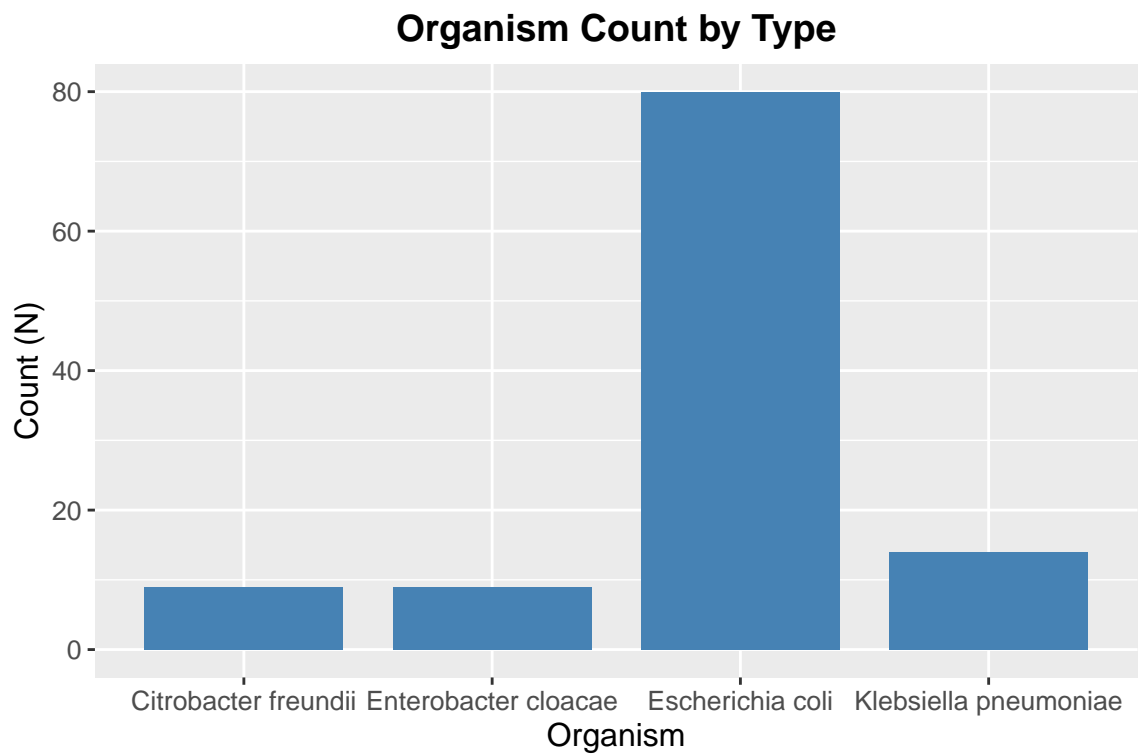
**Density plot of ceftolozane/tazobactam MIC values:**



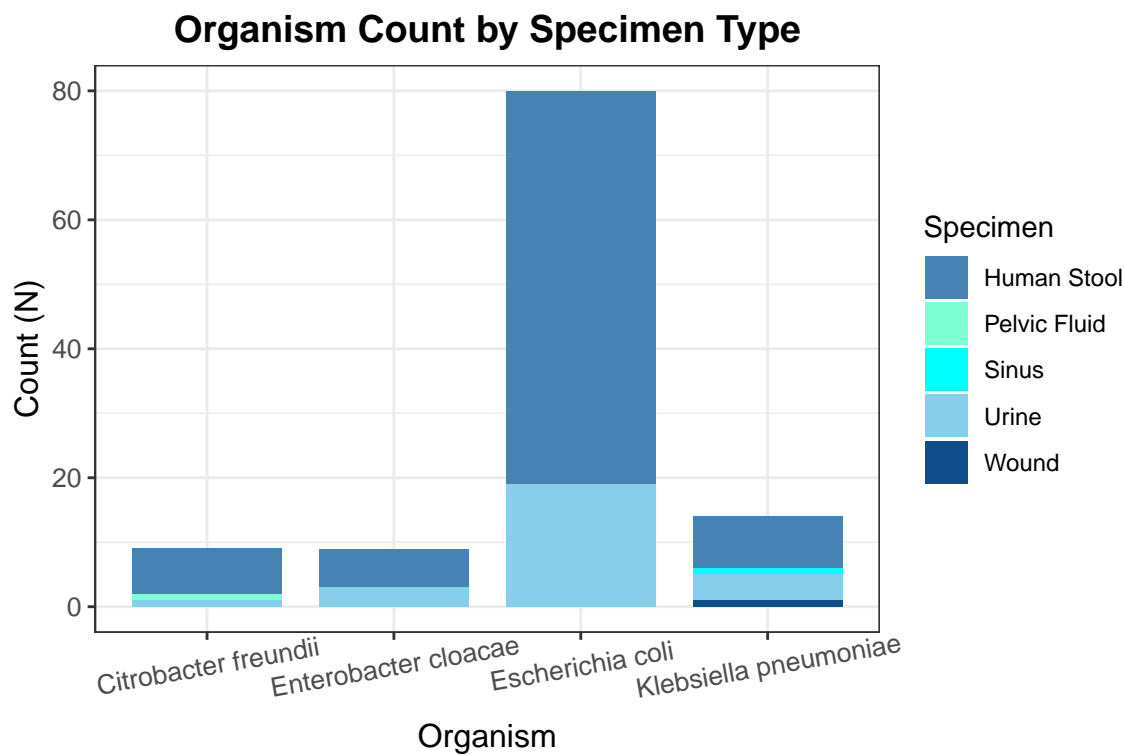
# Barplot

Use to show the relationship between a numeric and a categorical variable.

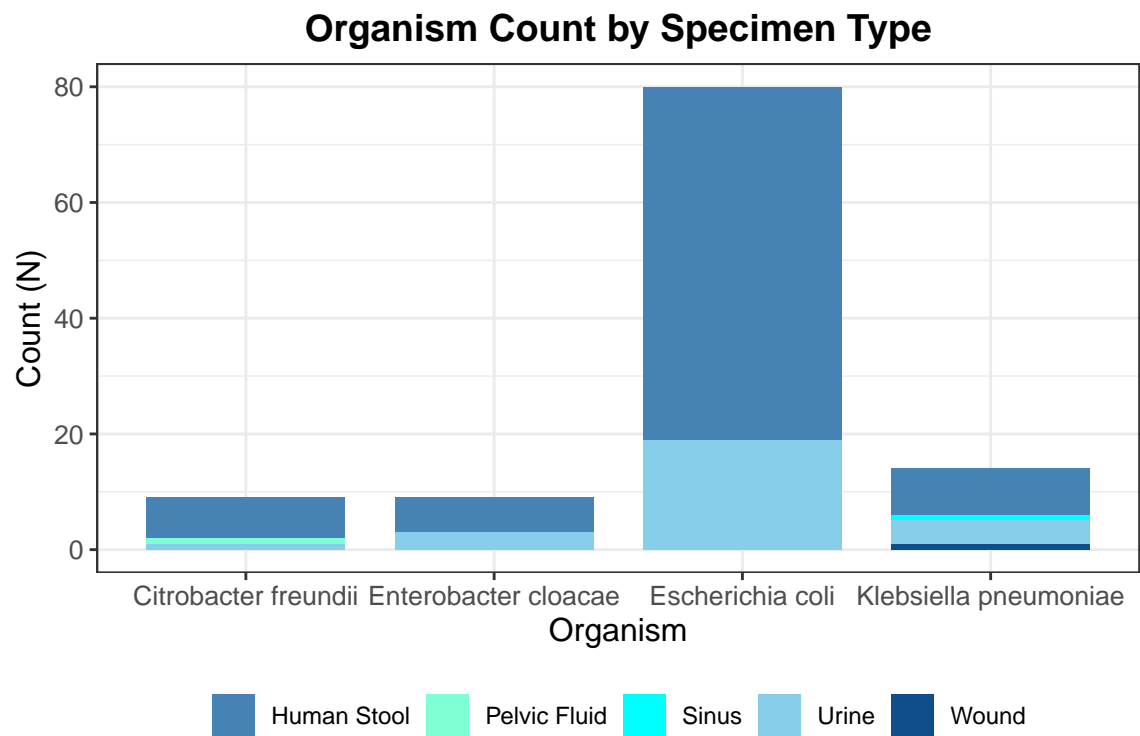
## Barplot of organism



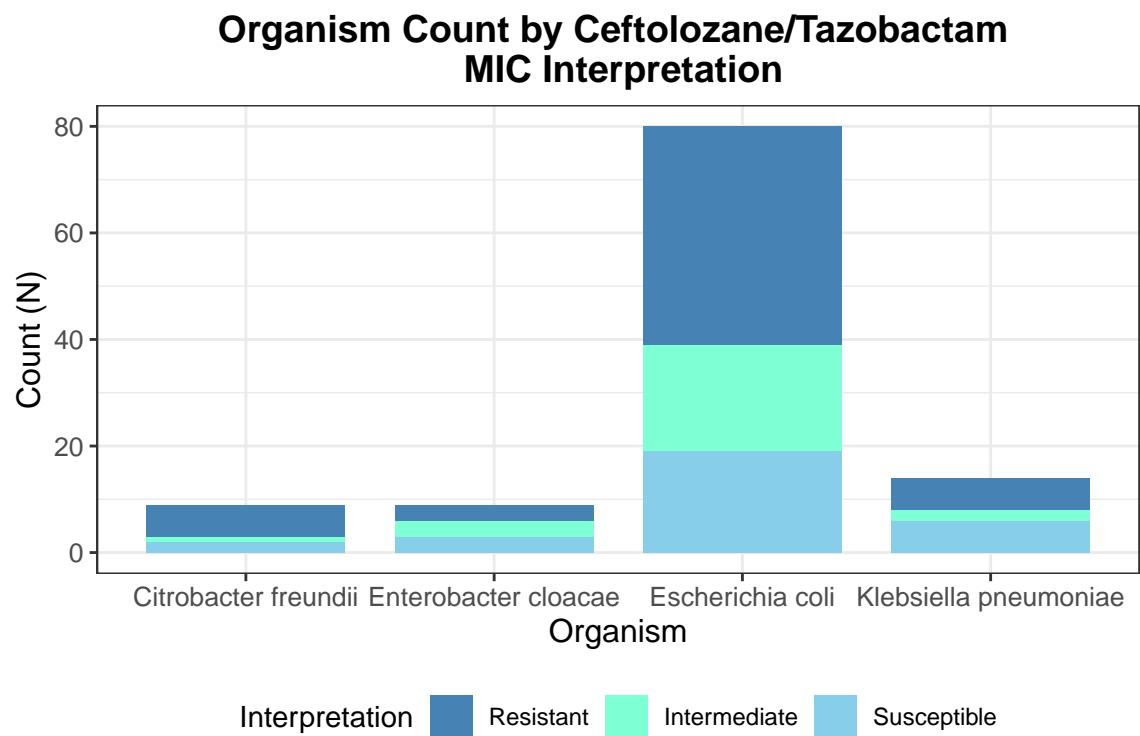
## Stacked barplot



Move legend to bottom

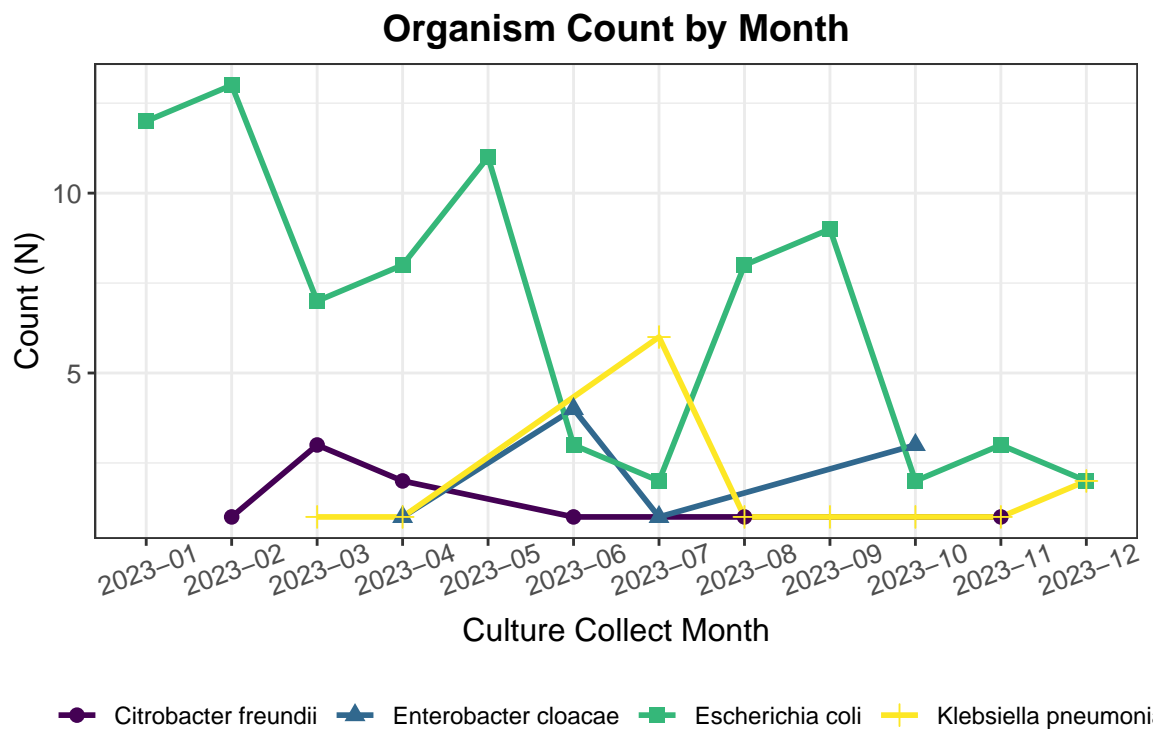


Organism count by ceftolozane/tazobactam MIC interpretation



# Line Graphs

Use with time series data.



**Viridis** scales can be perceived by users with common types of color blindness.  
[Further reading on viridis](#)