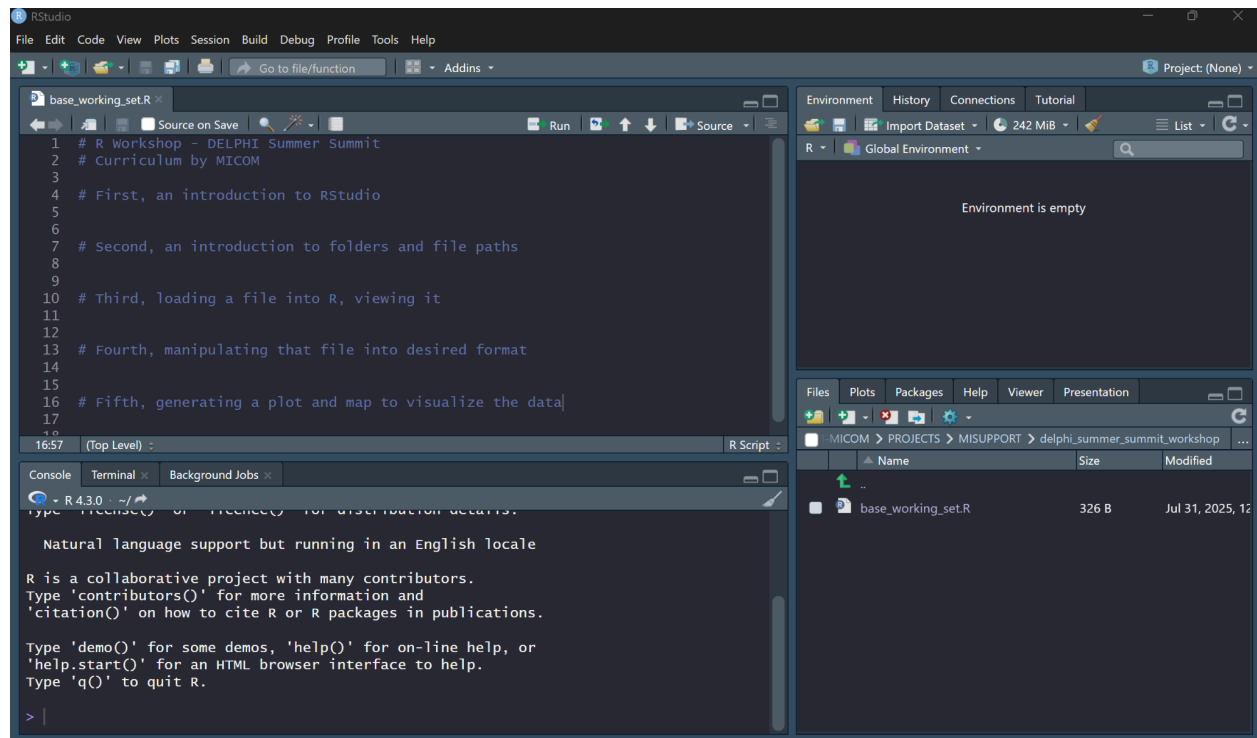


```
# R Workshop - DELPHI Summer Summit  
# Curriculum by MICOM / MI-Support  
# R Worksheet
```

First, an introduction to RStudio



Upper left-hand corner pane:

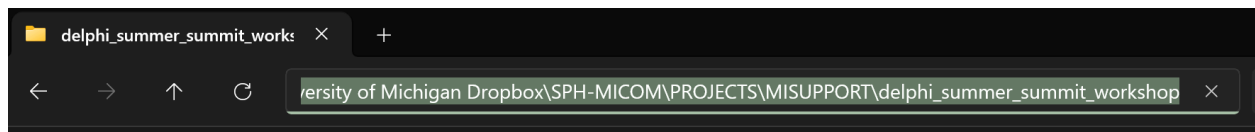
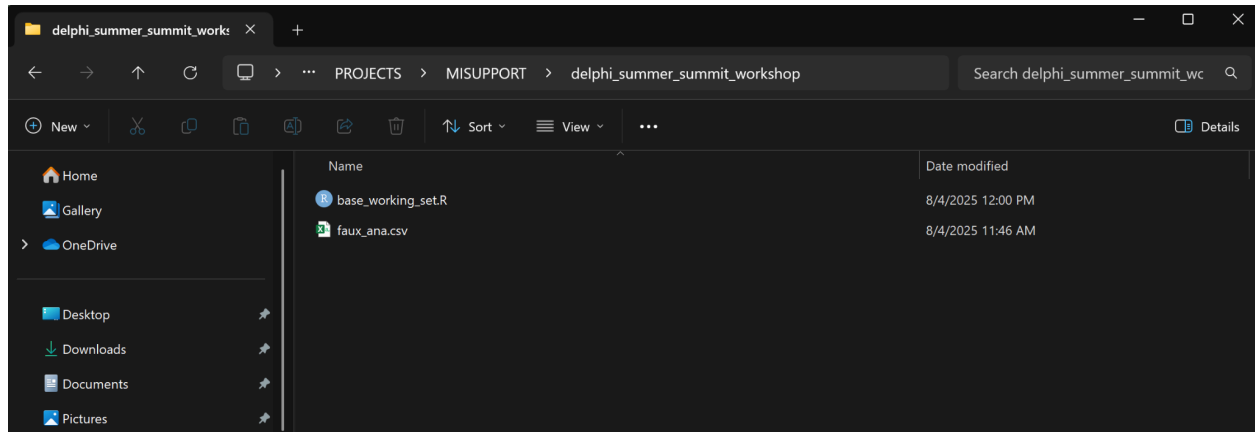
Lower left-hand corner pane:

Upper right-hand corner pane:

Lower right-hand corner pane:



Second, an introduction to folders and file paths



```
getwd()  
setwd()
```



Third, loading a file into R, viewing it

1) Create a new variable, and assign it the character string of the file path where your data file is located at.

2) Create a new variable that you want to store your data in. Use the functions in the below box to read the data file into R into the variable you just created.

```
read.csv()  
paste0()
```

Why might this method of reading in a file make sharing this code set easier in the future?

3) Click on your variable name in the Environment pane. Look through the data. In the console pane, use the functions in the below box on HOSP, AGE, and SEX. Which function is better for which type of data?

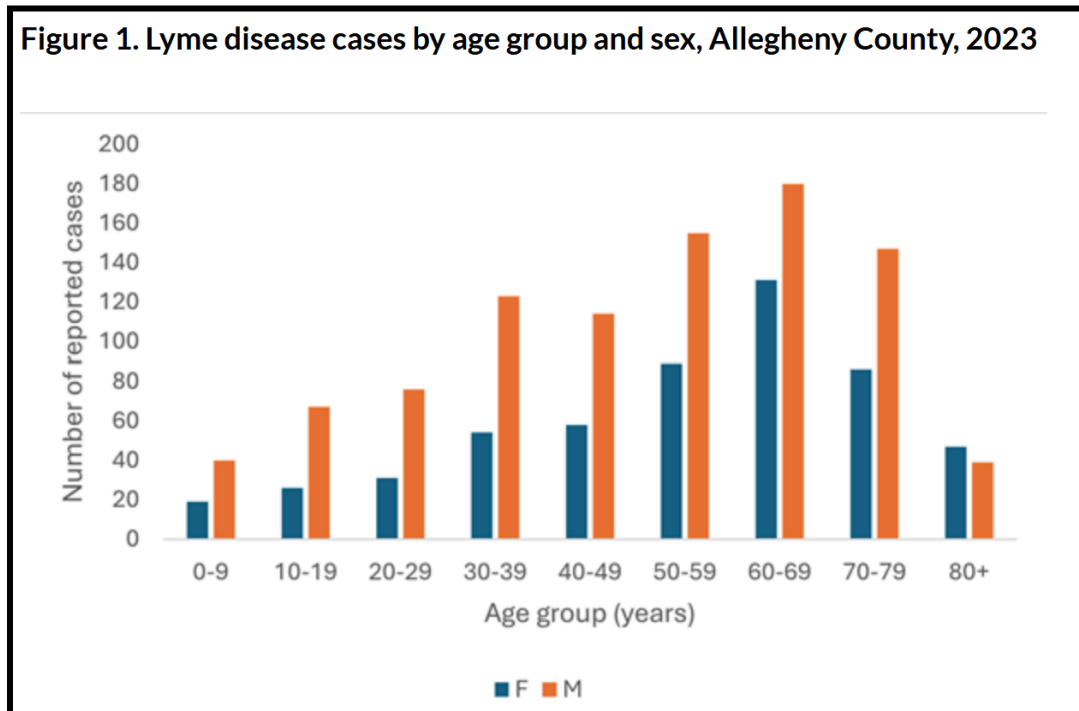
```
table()  
summary()
```

Extra) Fill in the blank: Of 109 cases reported in 2023, 69 (63%) were male and ____ were > 50 years old.



Fourth, manipulating that file into desired format

We're going to use R to make this chart (but with non-suspect Anaplasmosis data):



This will also give us a chance to start practicing one of the most useful R packages—the tidyverse!

1) In order to make this chart, what variables do we need to use? What are we grouping by, and what are we counting? Do we need to remove anyone from consideration?

2) Are there any groupings that we need for the chart that are not present in the dataset? If so, use the functions in the below box to make it / them.

```
mutate()
```



3) Use your answer to (1), (2), and the functions in the below box to make a new dataframe with the data for your chart. Assign that dataframe to a new variable.

```
filter()  
group_by()  
summarize()  
length()  
unique()
```

What do you notice about the difference in end result between mutate() and summarize()?



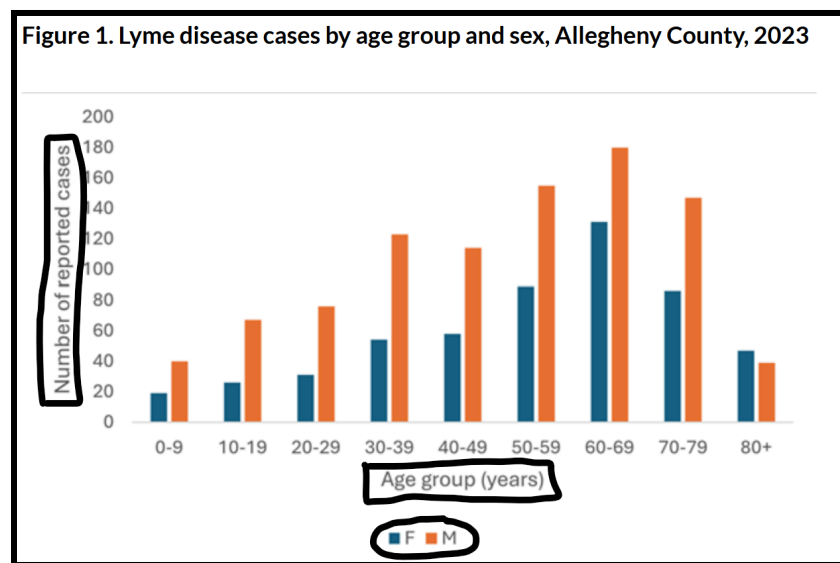
Fifth, generating a plot to visualize the data

1) In order to make this chart, what variable is assigned to each piece or “encoding” of the chart?

X-axis:

Y-axis:

Color:



2) Use the functions in the below box to make a “draft” bar plot.

```
ggplot()  
geom_col()  
labs()
```

3) Use the functions in the below box to save the image as a 4in x 6in .png image in the same folder as your code and data.

```
ggsave()
```

Extra) Format the bar plot more nicely, including color, label, and alignment, to match the website formatting more closely.



Extra) Save your code file, then email it to someone else at the workshop (perhaps the person sitting next to you). Can they download and run your code file?

