In this worksheet, you'll use the file *midwest.csv*, a subset of U.S. COVID-19 data originally collected by the COVID-19 Tracking Project. This smaller dataset includes data from 12 Midwestern states:

Date	**State**	**Positive Cases**
20201206	IA	213,390
20201206	IL	787,573
20201206	IN	381,617
20201206	KS	168,295
20201206	MI	426,576
20201206	MN	350,862
20201206	MO	322,298
20201206	ND	82,981
20201206	NE	138,568
20201206	OH	475,024
20201206	SD	85,991
20201206	WI	441,067

1. Briefly describe the dataset. What are the column headers? How many rows are there? Is there anything else you notice?

2. Go to the link posted on our class website: *The State of COVID-19 Race and Ethnicity Data*. Read through the article and take notes on your thoughts on the following prompts:

   (a) Why is it harmful when demographic data (like race, ethnicity, or age) are missing or inconsistently collected in public health datasets?

   (b) How does this connect to the idea that data has limits?

Keep these ideas in mind as you move into the coding activities, we will come back to them later in the week.

3. The code below, from midwest_covid.py, is one way to read in a csv file in Python. Comment the code and write the first item that prints from the list 'data'.

```
data = []

with open('midwest.csv') as file:
    next(file)
    for line in file:
        line = line.strip()
        fields = line.split(",")
        data.append(fields)

print(data)
```

4. Oh no! You ran the code an got this error: 'No such file or directory: 'midwest.csv' '.

This is your file structure:

```
cs111
|----- labs
       |----- midwest.csv
       |----- covid_lab
              |----- midwest_covid.py
```

Where is the file that you are executing? Where is the .csv file you are using? After answering these two questions, explain what caused the error and how you can fix it (hint: there are two ways).

5. Explain what the following lines of code will print:

```
print("Row 1:", data[0])
```

```
for row in data:
    print(row[1])
```

```
print("Row 3, Column 2:", data[2][1])
```

6. Now, write your own code to do the following:
   (a) Print the first 5 rows in the states column.
   (b) Access and print the number of positive cases in Minnesota.