**PROBLEM**

Monkeypox is a viral infection that spreads from human to human and can lead to medical complications particularly in new-born babies, children and people with underlying conditions. A multi-country outbreak of monkeypox has been reported in places where the virus has not been typically found before. Our objective is to understand how case rates differ by month for different regions within the EU and to determine if there is a relationship between certain demographics and MPX rates in order to prepare appropriate responses to avert further spread.

**METHODS**

We will utilize four datasets and R statistical software for analysis. First, the monkeypox case data file from that contains information on the number of monkeypox cases reported by EU/EEA countries or collected throughout epidemiologic intelligence at European Centre for Disease Prevention and Control. Second, the population denominator from Eurostat that indicates the number of persons having their usual residence in a country on 1 January of the respective year. Third, world region which includes country names linked to region, subregion, country code and more. Lastly, European statistical system census data file from 2011 European census which includes country code, age, education, economy and population.

**Data cleaning**

We import our datasets into RStudio and utilize select () and filter () functions to remove unnecessary fields, change casing using rename\_with () and correct missing variables.

**Variables retained**

MPX data set: (country name, subregion, country code, time period, obs\_value, month)

Population denominator dataset: (country code, time period, obs\_value)

World region dataset: (country name, region, sub region)

Census dataset: (country code, CAS “economically active”, res\_population, population)

**Creating new variables**

With the MPX data file we create 3 new variables to include **aggregated MPX** cases by month using the group\_by () summarize () and mutate () functions, **rate** by taking monthly total MPX cases divided by the obs\_value per 100000, rounded to 3 decimal places and the sum\_rate given by sum monthly rate of monkeypox cases in each sub-region.

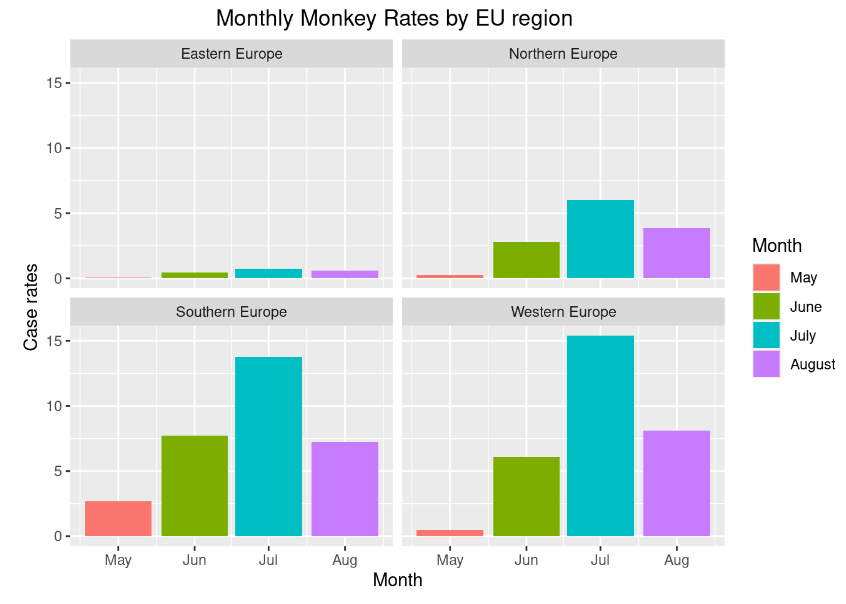
With the census data file, we group and summarize our data into a country level demographic metric (population size) using case\_when () function to create logical conditions (small, medium, large). We then narrow our focus to the small category and calculate the percent of population living in small cities. We do the same grouping and summary for CAS (economically active which we categorize as employed versus unemployed. We narrow our focus on the employed using filter ().

With the population denominator data file, we filter observations by country code (EU region) and time period (2022).

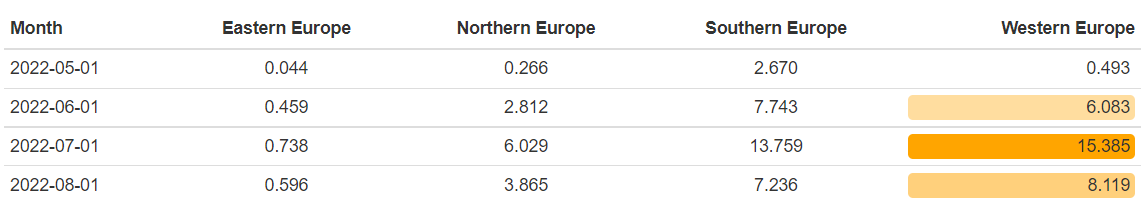
Next, we match the clean MPX data file, population denominator file and census data file by country code variable using left\_join () function.

**Visualizations**

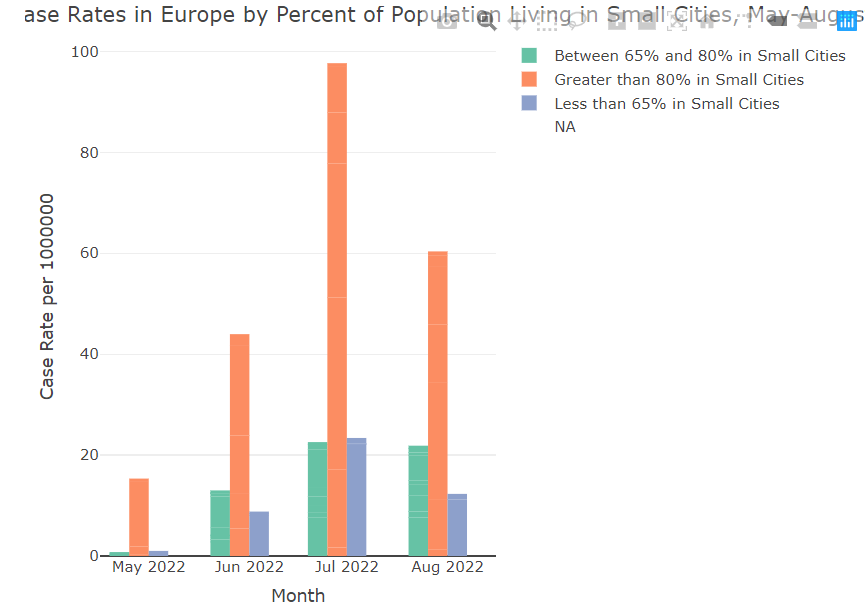
1. We use ggplot to plot the case rates in Europe by region per month in the months May to August 2022. In the month of July, monkeypox case rates were highest across all regions in Europe. In the month of May, monkeypox case rates were lowest across all regions in Europe.



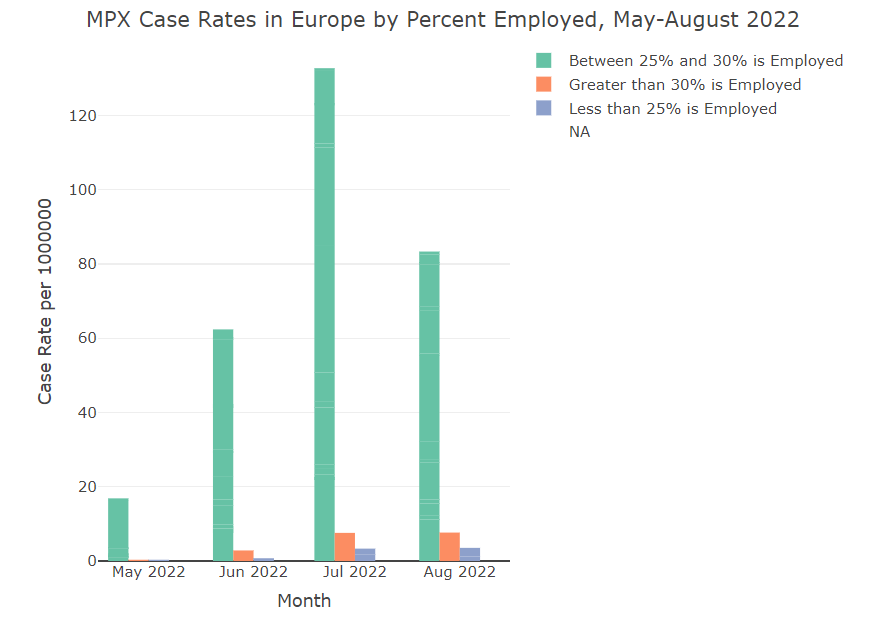
1. We use formattable to display the case rates in different European regions between the months of May and August 2022. Western Europe has the highest case rates in the month of July.



1. We use plotly for an ecology analysis to output a graph showing the case rates in Europe by the percent of the population living in small cities between the months of May to August 2022. The monthly monkeypox case rates peak in July 2022 particularly in countries where greater than 80% of the population lives in small cities.



1. We use plotly for an ecology analysis to show the case rates in Europe by the percent of employed population between the months of May to August 2022.The monthly monkeypox case rates between May 2022 and August 2022 are highest in countries where between 25% and 30% of the population are employed.



**Discussion**

Our findings reveal that highest case rates of monkey pox occurred in the month of July 2022 across all regions of Europe. The worst hit region was Western Europe. Countries in which greater than 80% of population was living in small cities were affected the most. Also, countries where between 25% and 30% of the population are employed registered the highest rates of monkey pox cases.

**Hypothesis**

1. Is there a difference between monkeypox cases by region in the EU?

Yes, our findings reveal that there is a difference between monkeypox cases by region in the EU.

1. Is there a difference between country level monkeypox case rates by certain demographic

Yes, our findings reveal that there are variations between country level monkeypox case rates by population size and economic activity.