Assignment 1: Protests
The past few years in the United States, there has been a surge in protests in support of Black Lives Matter, gender equity, and other social issues. In this assignment, you'll work with data from [CountLove](https://countlove.org/) -- the same data often [cited](https://www.nytimes.com/2020/08/28/us/black-lives-matter-protest.html) by the New York Times -- to learn more about demonstrations over the past few years.

By completing the assignment, you will demonstrate the following skills:

- Use of **version control** for managing your code
- Declaring document rendering using **markdown** syntax
- Foundational programming skills in R.

Background Research
Before diving into this (or any) dataset, it's important to have _domain familiarity_ (i.e., to know something about the topic). As preparation, I'm asking that you read **three articles** about protests in the U.S., and provide a brief 1 - 2 sentence summary or takeaway from each one.

In the section below, create an **unordered list** of the three articles you found. Make sure to provide an appropriate markdown link (_not_ just the URL) to the article in addition to your 1 - 2 sentence summary.

1. George Floyd Protest (2020-2021): The George Floyd Protest were comprised of many protests against police brutality and racism. This originated in Minneapolis on May 26, 2020 after the brute police shooting of George Floyd.
[George Floyd Protests](https://www.nytimes.com/article/george-floyd-protests-timeline.html)

3. The Women's March (2017): The Women's March Protest was mainly centered around the message that women's rights are human's rights. The main protest took place in Washington D.C.
[The Women's March](https://en.wikipedia.org/wiki/2017_Women%27s_March)

5. The March on Washington for Jobs and Freedom (1963): This protest was targeted to send the message about the inequalities and struggles African Americans were facing during this time. This is also where Martin Luther King Jr. delivered his famous "I Have a Dream Speech.
[The March on Washington](https://www.history.com/topics/black-history/march-on-washington#:~:text=The%20March%20on%20Washington%20was,challenges%20and%20inequalities%20faced%20by")

Accompanying Image
In this section, please **display one image** to accompany your text, and describe _why_ you included it (~2 - 3 sentences). This will require that you download an image into your project

folder. In your description, use **bold** and _italics_ (at least once, for practice) to emphasize some of your points.

I used this photo because this is a big and **recent** movement in American history. I chose this photo because it shows how many people attended this event, and how these protests fueled the Black Lives Matter campaigns, which _continues_ to spread awareness until this day!

Analysis
At this point, you should open up your `analysis.R` script to begin working with the data. The script will guide you through an initial analysis of the data. Throughout the script, there are prompts labeled **Reflection**. Please write 1 - 2 sentences for each of these reflections below:

- What does the difference between the mean and the median tell you about the *distribution* of the data?
- The mean was greater than the median, so the data was positively skewed to the right.

- Does the number of protests in Washington surprise you? Why or why not?
- The number of protests in Washington does surprise me because I didn't know the history of WA had that many protests in the span of 4 years!

- Looking at the `state_table` variable, what data quality issues do you notice, and how would you use that to change your analysis (no need to actually change your analysis)?
- Some data quality issues I noticed was that the table wasn congested and a but disorganized, making it hard to understand the information. Something that would make it more clear could be a function that allow the numbers to be displayed horizontally, beside the location. I also noticed that the values printed out in the table were all locations that only started with the letter 'B'. Perhaps there is another way for me to analyze it so that all the location names come up? I wonder why my specific code made it so that only specific locations were displayed.

- Does the change in the number of protests from 2019 to 2020 surprise you? Why or why not?
The change in the number of protests from 2019 compared to 2020 does surprise me because it increased by a lot in just one year!

- Do a bit of research. Find at least *two specific policies* that have been changed as a result of protests in 2020. These may be at the city, state, or University level. Please provide a basic summary, as well as a link to each article.
-One policy change that some states have enforced is about police reform. Since the protests if 2020, specifically the George Floyd protests, there is now a ban on chokeholds or neck holds. Police are only allowed to use strangleholds, which is meant to simply make a subject unconcious; the choke hold restricts breathing and can result in death (as it did).

[2020 Protests Impact - Police Force](https://www.usatoday.com/in-depth/news/2020/06/18/2020-protests-impact-city-and-state-changes-policing/5337751002/)

-Another policy that has changed was decreasing police funding. Two statea that enforced this were New York City and California. Instead of funding money to the police, they boost funding for health care, jobs and peace centers.
[2020 Protests Impact - Police Defunding](https://www.usatoday.com/in-depth/news/2020/06/18/2020-protests-impact-city-and-state-changes-policing/5337751002/)

- Take a look (`View()`) your `high_level_table` variable. What picture does this paint of the U.S.?
- The table shows how the biggest problem in the U.S. is surrounding Racian Injustice, which recently increased significantly due to the uproar of the Georg Floyd incident. It also has somewhat of an equal spread among other purposes, showing how one doesn't trump over others.

Critical Reflection
Now that you have had time to work with the data and visualize it, I want you to practice thinking critically about the dataset we provided. An important part of data analysis is reflecting on the assumptions, limitations, and gaps in your datasets. Remember that all datasets have to make assumptions and not all assumptions are bad or unreasonable! Practicing these skill will help you in your career to be self critical and ethically alert.

For this section, please write 2-3 sentences for each of the reflections below:

- How was the dataset collected and who collected the data?
- The dataset was collected by reading local news reports and television sites daily and compiling them into one big dataframe. The people who collected the data were Tommy Leung and Nathan Perkins, who are engineers and scientists that have a passion for public policy and civic responsibility.

- What assumptions does the dataset make? List atleast two assumptions you identified. For inspiration, check out this [blog post](https://towardsdatascience.com/check-your-assumptions-about-your-data-20be250c143) that describes one method for identifying data assumptions.
1. One assumption the dataset makes could be the exact amount of attendees that actually attended the protests.
2. Another assumption the dataset makes about the attendees is whether or not they were actually there for the purpose of the event or were just people among the crowd. How exactly would they know the purpose behind the individuals who attended?

- What data is missing from the dataset? Think about if there are any data points you would have liked to have seen in the dataset.
- There were lots of protests that did not have a number of attendees who were present at the protest.

Final Thoughts
When you are finished, with your analysis, please answer the following questions in 1-2 sentences each.

- What about the analysis surprised you?
- What surprised me was how convenient many of the functions were, since we re-use many of the same functions for different questions throughout the assignment. I was also surprised how efficient R was able to make tables and vectors for the different tasks we assigned to the different variables using the different objects.

- What parts of this analysis did you find challenging?
- What I found challenging was having to use different methods to analyze the various parts of the data frame, since there were multiple components I needed to consider before answering some questions. For example, the question regarding a specific parameter and returning a sentence with the specific location was tricky at first because I needed to code multiple lines, using the str package!

- What types of analysis do you wish you were able to do with the dataset, but currently don't have the technical skills to do?
One specific thing I wish I was able to do better would probably learn how to make a clearer box plot for my data since the one I coded was very unclear. Similarily I want to learn how to make table clearer as well, such as the one we had to make in state_table.