**COVID-19 Fact Sheet / Data Brief Instructions (40 points)**

As the culminating assignment for this course, students will create a data brief to disseminate COVID-19 related information to the general public using state-level information on COVID-19 (1) tests, (2) confirmed cases, (3) hospitalizations, and (4) deaths for June 1-7, 2020 (“study period”) from the Center for Systems Science and Engineering at Johns Hopkins (<https://github.com/CSSEGISandData>).

Please note that the data presented for tests, confirmed cases, etc. for each day *include those that occurred on that day*. Information should be presented as bullet points and data visualizations to illustrate key information clearly and concisely to a population potentially unfamiliar with epidemiologic and statistical language. All datasets needed for this assignment are provided by the instructors on NYU Classes and include:

* Eight (8) SAS datasets with daily COVID-19 information for the study period (“cd0601”- “cd0607”) and for the day prior to the start of the study period (“cd0531”)
* One (1) SAS dataset indicating region for each U.S. state/territory, based on U.S. Census classification (“region”) (<https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf>)

**Step 1**. Use SAS to answer the following questions:

First, let’s examine the data for the entire United States, i.e., all states and territories:

1. What was the total number of (a) tests, (b) confirmed cases, (c) hospitalizations, and (d) deaths recorded in the U.S. at the beginning of the study period (on June 1)?

What was the total number of (a) tests, (b) confirmed cases, (c) hospitalizations, and (d) deaths recorded in the U.S. at the end of the study period (on June 7)? **1 point**

Next, merge all datasets provided into one combined dataset. Compute the number of new tests, new confirmed cases, and new deaths for each day of the study period (June 1-7, 2020) by subtracting the total number recorded the day before from the total number reported that day.

1. How did the total daily number of new tests, new confirmed cases, and new deaths change over time in the U.S. during the study period (June 1-7, 2020)? Did these numbers increase, decrease, or pretty much stay the same? **6 points**

Second, let’s investigate the data by region.

1. At the beginning of the study period (on June 1, 2020), how did the average hospitalization rate compare across the five regions of the U.S.? In other words, which region had the highest and which region had the lowest average (mean) hospitalization rate? Please answer the same question for the last day of the study period (June 7, 2020). **4 points**

Finally, let’s look at the data at the state level:

1. At the beginning of the study period (on June 1, 2020), which state or territory had the highest and which state or territory had the lowest number of (a) new tests, (b) new confirmed cases, and (c) new deaths?

At the end of the study period (June 7, 2020), which state or territory had the highest and which state or territory had the lowest number of (a) new tests, (b) new confirmed cases, and (c) new deaths?

Which state or territory had the highest and which state or territory had the lowest total number of (a) new tests, (b) new confirmed cases, and (c) new deaths recorded during the study period? **3 points**

1. Select one state/territory from each of the five regions. (NOTE: your selections are completely up to you.) Compare trends in daily (a) new tests, (b) new confirmed cases, and (c) new deaths across the study period for the selected states and/or territories. Do you see different or similar patterns across the five selected locations? **6 points**

**Step 2**: Using the examples provided as a guide, present the information you obtained in Step 1 in a data brief using bullet points and data visualizations. Be sure to:

* Include answers to all questions posed in Step 1. NOTE: You do not need to include responses to each of the questions separately from what you present in the data brief. Just be sure all information you obtained for Step 1 is included in the brief.
* Fact sheet/data brief should be 1-2 pages (single spaced) and include 1-2 visualizations (short tables or figures), 11-point font, submitted as a Word document (*not* a Google doc, .pages document, or PDF)
* Provide basic information about the data source
* Provide information in language that is understandable to a diverse audience without epidemiologic and statistical background
* Include the SAS code you used for this assignment in your submission. Add comments in your SAS code to improve readability of your code. Submit SAS code in a Notepad (.txt) or Word (.docx) file.

**Grading**:

* SAS code (10 points) - correct syntax, includes commenting
* Correct answers to questions in Step 1 (20 points)
* Presentation of information in the data brief (10 points) - correct interpretation, clarity