Data Science Fall 2020

**Assignment 4**: Data Science (written – 30% and 10% presentation of overall credit score)

**Due: Both written part of assignment and the poster – 17th December 2019 by 11:30 am ET on LMS.**

Group poster presentations during the finals week Thursday, 12/17/2020 from 11:30 am – 2:30 pm ET via online.

Submission method: LMS. Please use the following file naming for electronic submission for documents: DataScience_A4_YOURGROUPNUMBER_written_part.xxx ect.

Late submission policy: **This assignment is due at the end of term. If your group submission is more than 2 days late, it is likely that you will not have your grade for this assignment included in your final grade before they need to be submitted. 20% of score deducted each late day.**

Instructor Office Hours: Tue/Fri 12:30 pm – 1:30 pm, via WebEx or by appointment.

Note: Your report for this assignment should be the result of group work. All members of the group will receive the same grade. Take care to avoid plagiarism ("copying"), include all web resources, text, and class presentations. You are expected to work within a group setting contributing equally and with complementary skills and are encouraged to discuss your ideas and the tasks for this assignment with other student the group with other groups in the class.

General Assignment: Working with someone else's data. The group chooses a data science investigation, finds, accesses, analyzes and presents/visualize the (more than one set/type of) data and manages the resulting products. The weighting score for each question is included below. Please use the question numbering (1-4) below for your written assignment.

1. Choose an investigation and identify pre-existing source of data that can address a particular data science goal (7%)
   a) Choose, and state, the goal and reasons why the datasets were chosen and how they were found and managed, Min 3-4 sentences.
   b) Document and discuss the data formats and any metadata standards/ conventions in use, and the method(s) of discovery and access and how they helped or hindered the process, Min 3-4 sentences.
2. Data Analysis (10%)
   a) Develop and state two particular questions/hypotheses related to the goal of the investigation and that can be answered using the datasets under consideration. Design an analysis study (preliminary, full and post) to answer these questions and document the analysis design, Min 3-4 sentences (3%)
   b) Provide a description of the choices of tools/methods used or a description of any code or scripts written, and describe how your results were stored and managed, Min 3-4 sentence. Submit your code to course GitHub repository for evaluation (3%)
   c) Perform the analysis in a form that can be validated and describe the steps and results your group took to ensure this validation, Min 3-4 sentences (4%)

3. Presentation/Visualization (8%)
    a) Prepare presentation / visualization of both the data (and any metadata, information) and the results of the analysis and describe them, Min 2-3 sentences. (3%)
    b) Document the management of the presentation / visualization products and any associated metadata, etc. Min 2-3 sentences (2%)
    c) Describe how your presentation/visualization supports the goal of the data science investigation and highlight any value that was gained, Min 3-4 sentences (3%)
4. Describe your overall data management plan for the results for questions 1,2, and 3 using the 9 categories from assignment 2, Min 1-2 sentences for each category (5%)
5. Create a poster (poster templates are available on LMS). Please submit your poster on LMS using the same naming scheme mentioned above. Mandatory peer-evaluation form must be submitted within 12 hours of the final project submission on LMS in order to receive the presentation and participation grade. (10%)