Assignment 4: Data Science (written - 30% and 10% presentation of overall credit score)
Due: **Written** assignment - Tuesday November 29, 2016 (by 12pm ET, day of the class), **presentation** – Tuesday December 6, 2016 (during/ after class).
Submission method: LMS. Please use the following file naming for electronic submission for any individual documents:
DataScience_2016_A4_YOURGROUPNUMBER_part1.xxx, etc.
Late submission policy: first time with valid reason – no penalty, otherwise 20% of score deducted each late day
Office hours: Monday 3:00-4:00, Winslow 2120 or Lally 207A 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"), including all web resources, texts, 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 students in the group and with other groups in the class.

General assignment: Working with someone else's data; group chooses an investigation, finds, accesses, analyzes and presents/visualizes the (**more than one set** 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 sources of data that can address a particular data science goal: (7%)
    a. Choose, and state, the goal and reasons why the data sets 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 choice 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 sentences (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 presentations/ visualizations 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 meets the goal of the investigation and highlight any value that was gained, min. 3-4 sentences (3%)
4. Describe your overall data management plan for the results of questions 1, 2 and 3 using the 9 categories from assignment 1, min. 1-2 sentences for each category (5%)
5. Oral presentation by the group of the results for questions 1, 2, 3, and 4. Plan to present for NO MORE THAN 12 minutes, followed by 3-5 minutes for questions. Your participation grade will be assessed on demonstration of learned skills in data management. All members of the team should (~equally) present part of the presentation. Please submit your presentation (ppt, pdf or similar), using the same naming scheme for the written assignment, after the presentation. (10%)