FALL 2018: CAP 5768 – Intro to Data Science
[Final Report]

Submit a final report on your project by noon on Monday, December 3, 2018.
The main purpose of the final report is to document all that was done and to make the work *reproducible*. One way to organize your report is to include the following sections: Introduction, Approach, Data and experiments, Results, Discussion and conclusion. An alternative organization may include the following sections: Introduction, Data, Aim 1 (experiments, results, conclusions), Aim 2 (experiments, results, conclusions), ..., Final discussion and conclusion.

The **introduction** section should contain a brief summary of your project, adequate background and motivation for the work. It should also state the overall aims of the project. The **approach** should contain a summary of the tools and techniques used in the work. Provide adequate references for the techniques used.

The **data** section should provide the source of the data. Provide URLs and/or repository details. If only part of the data set was used (e.g., some years or some cross section), please specify. If queries or code (API) was needed to download the data, provide the information. If the data is constantly changing, provide the date on which the data was downloaded. Provide details on the information available in each of the tables and/or files being used. If the data is very large, provide a summary of the breakdown on the data in your report (for example, how many subjects, how many males/females, how many years, how many in each age group, etc.). Provide overall size information for the data (how many rows, subjects, MB or GB, etc.).

When describing the **experiments**, make sure to provide the scripts or programs you wrote as well as the specific commands and parameter settings used for running the scripts or code. Don't forget that the goal is reproducibility. For every tool used, including Python and R, make sure to include the version number. Provide the machine(s) configurations on which the experiments were run. Provide information on the sizes of the outputs produced and the time taken, if this is relevant to the project.

In the **Results** section, provide a summary of the results, charts, tables, output, etc. This section may be combined with the experiments section. If the results are deposited on a website or on a separate CD, provide information and organization of the results. Even if all the results are deposited elsewhere, it would be wise to include one or two representative (summary) tables or charts or graphs in your report to document your findings. Learn how to summarize and visualize a large body of results. Organize into subsections, if needed.

Besides describing the experiments and results, it is essential to interpret and discuss the results. The **discussion** section is meant for making observations and discussing possible explanations for the observed results. Without this critical section, your work is incomplete. Organize into subsections, if needed.

In the **conclusions** section, you should summarize your results. You could also use this section to point to things that you did not do or questions that were left unanswered and also suggest future work.

There is no page limit for this report.