Students in this class will work on a semester-long project analyzing a chosen data set. The data sets all come from real life research projects sponsored by external faculty collaborators. Students who choose the same data set will form a team. All techniques learned during the semester will be applied to the data, likely to varying degrees of success. It is also expected that additional techniques will be researched and applied under the team’s discretion, and in consultation with the sponsor/professor. A research paper is expected from each student at the end of the semester detailing the most successful of these analyses and results, as well as clear descriptions of progression of the project, including methods employed, decisions made, explanations of findings and individual contributions. In addition, there are two check points during the semester where interim reports are expected from each team and a final presentation. All team members will collaborate on check-point reports as well the presentations.

Data Sets

This is a list of the data sets and associated projects you may choose from. You may also provide your own alternative data set and project, in consultation with the professor. These choices must be made within the first 10 days of the semester (see deadlines section).

- MCMi ImPACT - Clinical research data set containing ImPACT scores for approximately 5,400 Maine athletes who have taken the test twice in the time period of 2009-2013. This is the largest ImPACT database on high school athletes in the world available to researchers by 1-2 orders of magnitude and there are a lot of open questions. The ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) is the most-widely used and most scientifically validated computerized concussion evaluation system. Please visit https://www.impacttest.com/about/ for more information on the ImPACT test. Provided by Professor Bruce Maxwell from Colby College. Professor Maxwell will visit Bryn Mawr and present on Friday 9/19.

- Dark Reactions - chemistry data set containing materials chemistry experiments and associated experimental results. Provided by Professor Sorelle Friedler from Haverford College. Professor Friedler will visit class to present on 9/4.

- Digital Du Chemin - the lost voices project. Music data set provided by Professor Richard Freedman from Haverford College. Professor Freedman will visit class to present on 9/4.
• Western Forest - biology data set comprised of Forest Inventory and Analysis data collected by the United States Forest Service and has a very large sample size (150,000 plots spread across the country). Provided by Professor Sydne Record from Bryn Mawr College. Professor Record will visit class to present on 9/9.

**Deadlines**

1. Thursday 9/11: choose a project. A 2-page write-up is required from each student that outlines the amount of background research conducted, the project as he/she understands it, the expected outcomes, and/or intended goals (what questions are you looking to answer?) along with plans on how they may be accomplished.

2. Thursday 10/9: Check Point 1 interim report

3. Tuesday 11/25: Check Point 2 interim report

4. Thursday 12/11: Project presentations

5. Friday 12/19 (12:30pm): Final research paper

**Additional details**

• All projects require no specialized knowledge other than the prerequisites for this class: intermediate programming skills and mathematical maturity. But the projects do need either domain knowledge or the willingness to assimilate it.

• The projects are open ended. It will take time and communication with the sponsor/professor to scope the project and identify first steps. Be prepared for confusion and frustration, but learning to handle this phase of a project is one of the most important skills you can develop!

• Communicating with collaborators and sponsors

  It is not easy to communicate across disciplinary boundaries, which you will all be required to do, as all data sets come from research projects in various application areas. Here are some advices to hopefully smooth the process:

  1. Do your homework.
  2. During conversations, distinguish between things you need to ask about synchronously, and things you know you can look up later.
  3. Before asking questions, try to answer them yourselves. Then if you have to ask, you can include your best guess and ask for confirmation/denial.
  4. Think of your liaison’s time as a resource you need to use sparingly (but not so sparingly that you waste time).
5. Take the questions you are asked and translate them into terms relevant to the project.

**Grading**

The project, which counts for 40% of each student’s final grade, will be graded in the following parts:

- Interim reports (25%) - team grade
- Presentation (25%) - team grade
- Final paper (50%) - individual grade