

## Byrne Seminar Fall 2015

### Data! What is it good for? Absolutely Something!

#### Course Syllabus

**Instructors:** Prof. Waheed Bajwa ([waheed.bajwa@rutgers.edu](mailto:waheed.bajwa@rutgers.edu)) and Prof. Anand D. Sarwate ([anand.sarwate@rutgers.edu](mailto:anand.sarwate@rutgers.edu)) - Department of Electrical and Computer Engineering

**Logistics:** We will have 10 sessions of this course. It meets on Tuesdays from 10:20-11:40 in SEC 220 on the Busch Campus. Course dates are:

- 9/1, 9/15, 9/22, 9/29, 10/6, 10/13, 10/20, 10/27, 11/3, 11/10

**What this course is about:** There's no question that the way we view the world is profoundly connected to the collection, processing, and understanding of data. But what is data? How is it collected? What needs to happen between measuring or collecting data and then making some decision or conclusion based on the data? Can all of this be automated? Should we trust the resulting systems? "Big Data" is all the rage these days and the goal of this class is to dig a little deeper and learn about the promises and pitfalls of massive data gathering. We want to think a bit more skeptically and critically about the benefits and harms from data. To do this we will learn how data processing systems work, what they can and can't do, the different people involved, and the kinds of tradeoffs that have to be managed.

**Course objectives:** By the end of the course, we hope you will be able to:

- understand and describe the benefits and drawbacks about relying on "data"
- think critically about new applications that claim to use data to solve problems
- understand the steps involved in developing data-driven solutions for some real-world problems

**Structure of the course:** This is a seminar course and it will be interactive in nature. The time in each class will be divided between class discussions, instructors' presentations, and brainstorming sessions. We'll have you work in groups for the first part of the course. You'll be randomly assigned into an "A group" for the first 3 classes (9/15, 9/22, and 9/29) and a second random "B group" for the next three classes (10/6, 10/13, and 10/20). During each class, two groups will be responsible for presenting/discussing some extra material.

For the "A group" classes, groups will have a little extra reading and should help facilitate discussion with the rest of the class. For the "B group" classes, groups will present "case studies" of individuals/organizations/companies using data to solve real problems.

Finally, for the last part of the class, students can form their own groups (of 2-4 people) to do a group project. The project will involve taking a real world problem (your choice) and then using the framework/ideas from the class to propose how you might use data to solve this problem. We'll give more details as the course goes on. For the project you will have to write a short report and give a presentation during the last two class sessions (11/3 and 11/10).

**Course credit:** In order to obtain credit for this course, you should attend at least 80% of the classes, participate in class discussions, and complete the course project (both report and presentation).

**Case Studies:** The following is a list of some of the case studies that will come up during "B group" classes, which will involve groups discussing real-world cases related to data science. Below are some examples:

#### Web 2.0 and technology case studies

- Recommender systems: Nielsen, Amazon, Netflix
- Page Rank and search algorithms (e.g. Google)
- Hadoop and Map Reduce
- Cloud services: Amazon EC2, Azure, etc.
- Amazon Mechanical Turk

#### Data and society case studies

- Netflix prize and other privacy violations
- Google street view
- Health research in reality vs. media: vaccines and autism
- Reproducible research: the case for austerity and errors in spreadsheets
- Data and politics: polling and elections (c.f. the UK)

#### Data science and modeling

- Census, sampling (extrapolating from samples to populations)
- Models for climate change
- New York parking related (big snow, weather models)

**Description of the final project:** For the final project you will work in teams (you can form your own teams) to find a real-world problem that you think could be solved/addressed using data. This is a *thought experiment*: we want you to play "what if" and think through the steps of how you would get and use the data and what sort of challenges/obstacles/biases you might encounter along the way.

In terms of what you have to write up for the final project, we will want three things. We will give more details later. For now though:

1. *A proposal*: this is 1-page document (12 point font, Times New Roman or similar typeface) that answers some basic questions like: what problem you are trying to solve? What has been tried before? Why was that was not enough? Who would be the best group to solve this problem: a government, a company, a non-profit, or...? What will the benefit to society be?
2. *A final project report*: this is a 3 to 4 page document (12 point font, Times New Roman or similar typeface) containing appropriate figures which goes into the previous questions in more detail (so you can reuse your proposal). We'll give you a structure/format, but it will cover the team you would need to address this problem, what kind of data collection you will need, the data processing, algorithms you might use, how you will interpret the data, what actions/outcomes might result, and most importantly what obstacles or challenges you would need to overcome.
3. *A final presentation*: this will be a group presentation (e.g. using PowerPoint or Keynote) describing the topics in your report. The class will get a chance to ask questions and give you feedback on your presentation!