Data and Society

Syllabus and Course Information

2/16/15
About this course

• COURSE ELIGIBILITY
  The course is open to graduate and undergraduate students in computer science and ITWS. Other majors will be considered by approval of the instructor as space allows.

• COURSE DESCRIPTION
  Digital data has transformed the world, impacting virtually every area of science and society. Data and Society provides a broad overview of how society is leveraging and responding to the social, organizational, policy, and technical opportunities and challenges of a data-driven world. Course themes focus on various aspects of the data ecosystem, data and innovation, and data and the broader community. Course resources include articles, reports, outside speakers, and technical literature on course themes.

  Data and Society is a communication-intensive course. Graded work for the course will focus on helping students develop critical communication and assessment skills needed for professional success. Assignments include written reviews and oral presentations, participation in "data roundtables", section exams, and a paper. Course prerequisites are CSCI/ITWS 4350/6350 or consent of the instructor.
Instructor

• Professor: Dr. Fran Berman
• Office: AE 218, 518-276-3794
• Office Hours: Friday 1-2 or by appointment (send email to bermaff@rpi.edu)
• Course website (linked off Fran’s RPI web page): http://www.cs.rpi.edu/~bermaff/Data%20Course.html
Data and Society – about this course

• This course will provide a broad snapshot of the data-driven world
  – We’ll skim the sea of interesting data stuff, but we won’t / can’t include everything
  – We’ll focus more on societal issues than technical issues
  – The course should provide a complement to the material in the ITWS Data Science, Web Science and other courses

• The course will be structured to
  – Increase your engagement with material
  – Evolve your professional communication and assessment skills
  – Help you develop as a “data-literate” professional

Course structure:
• Section 1: The Data Ecosystem -- Fundamentals
• Section 2: Data and Innovation – How data has transformed science, commerce, and life
• Section 3: Data and Community – Social infrastructure for a data-driven world

Guest Speakers this Semester:
• Colin Bodel, CTO, Time Inc.
• Mike Schroepfer, CTO, Facebook
• Bulent Yener, RPI CS Professor
<table>
<thead>
<tr>
<th>Section 1: The Data Ecosystem -- Fundamentals</th>
<th>January 30</th>
<th>Class introduction; Digital data in the 21st Century (L1)</th>
<th>Data Roundtable / Fran</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 6</td>
<td>Data Stewardship and Preservation (L2)</td>
<td>L1 Data Roundtable / 5 students</td>
<td></td>
</tr>
<tr>
<td>February 13</td>
<td>Data and Computing (L3)</td>
<td>L2 Data Roundtable / 6 students</td>
<td></td>
</tr>
<tr>
<td>February 20</td>
<td>Colin Bodel, Time Inc. CTO Guest Lecture and Q&amp;A</td>
<td>L3 Data Roundtable / 5 students</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 2: Data and Innovation – How has data transformed science and society?</th>
<th>February 27</th>
<th>Section 1 Exam</th>
<th>Data and the Health Sciences (L4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 6</td>
<td>Paper preparation / no class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 13</td>
<td>Data and Entertainment (L5)</td>
<td>L4 Data Roundtable / 6 students</td>
<td></td>
</tr>
<tr>
<td>March 20</td>
<td>Big Data Applications (L6)</td>
<td>L5 Data Roundtable / 5 students</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 3: Data and Community – Social infrastructure for a data-driven world</th>
<th>April 3</th>
<th>Data in the Global Landscape (L7) Section 2 paper due</th>
<th>L6 Data Roundtable / 6 students</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 10</td>
<td>Bulent Yener Guest Lecture, Data Privacy / Bad guys on the Internet (L8)</td>
<td>L7 Data Roundtable / 5 students</td>
<td></td>
</tr>
<tr>
<td>April 17</td>
<td>Data and the Workforce (L9)</td>
<td>L8 Data Roundtable / 6 students</td>
<td></td>
</tr>
<tr>
<td>April 24</td>
<td>Mike Schroepfer, Facebook CTO Guest Lecture and Q&amp;A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 1</td>
<td>Data Futures (L10)</td>
<td>L9 Data Roundtable / 5 students</td>
<td></td>
</tr>
<tr>
<td>May 8</td>
<td>Section 3 Exam</td>
<td>L10 Data Roundtable / 5 students</td>
<td></td>
</tr>
</tbody>
</table>
Course Information
http://www.cs.rpi.edu/~bermaf/Data%20Course.html

• Course website (above) will have all up-to-date information and materials.
  – Syllabus may evolve slightly

• Course Readings
  – Provide background and information for the lectures and issues
  – Will be given every week. Most source materials and readings for each week provided on the course web page
  – Readings eligible for Data Roundtable reviews/presentations will be specified in class and on the web

• Source and reading materials may be tested on the Section exams – relevant material will be indicated
How you’ll be graded

Student grades are computed from:

- **2 section exams** (20 points each)
- **1 section paper** (20 points):
  - Undergrads: 6-8 page research paper on an approved Section 2 topic
  - Grads: NSF-style 10 page mini-proposal on an approved Section 2 topic
- **Class participation** (10 points)
- **3 Data Roundtable reviews / presentations** (roughly 1 per section, 10 points each)

Students can obtain up to 5 points extra credit by doing an Op-Ed (due any time up to May 1)

Fran Berman, Data and Society, CSCI 4967/6963
More about grading
(additional grading specifics for each component later today)

• **Data Roundtables:**
  – Students are responsible for scheduling their Data Roundtables and ensuring that all 3 are done.
  – Students are responsible for self-organizing so that each student presents a **different** Roundtable article.

• **Class engagement / attendance:**
  – Students are expected to attend 90+% of the class meetings (12/13).
    Attendance will be taken in class
  – Engagement grade: 5% attendance, 5% class participation

• **Exams** will be primarily in essay format. You’re responsible for anything covered in class and in the relevant readings.

• **Extra credit:** Up to 5 points extra credit can be earned by doing an Op-Ed by May 1.

• **There will be a slightly different workload for grad students and undergrads**
  – Section 2 paper / mini-proposal requirements are different.
  – In writing and presentations, each student will be assessed at a level appropriate to their educational level (undergrad or grad)
## Learning Objectives and Outcomes

<table>
<thead>
<tr>
<th>Learning Objective</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop greater data literacy</td>
<td>Be able to understand and explain the role that data plays as well as its limitations in various areas of research, commerce and modern life.</td>
</tr>
<tr>
<td>Develop critical thinking skills around data</td>
<td>Be able to read, understand, assess, and discuss data-oriented professional and popular publications and articles.</td>
</tr>
<tr>
<td>Develop communication skills around data</td>
<td>Be able to advance an evidence-based argument about data, data cyberinfrastructure and data-oriented efforts to both knowledgeable specialists within the field as well as non-specialists.</td>
</tr>
</tbody>
</table>
Academic Integrity

• Student-teacher relationships are built on trust. For example, students must trust that teachers have made appropriate decisions about the structure and content of the courses they teach, and teachers must trust that the assignments that students turn in are their own. Acts, which violate this trust, undermine the educational process. The Rensselaer Handbook of Student Rights and Responsibilities defines various forms of Academic Dishonesty and you should make yourself familiar with these.

• In this class, all assignments that are turned in for a grade must represent the student’s own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration. If references or other materials are used, they should be cited. Submission of any assignment that is in violation of this policy will result in a penalty.

• If found in violation of the academic dishonesty policy, students may be subject to two types of penalties. The instructor administers an academic (grade) penalty, and the student may also enter the Institute judicial process and be subject to such additional sanctions as: warning, probation, suspension, expulsion, and alternative actions as defined in the current Handbook of Student Rights and Responsibilities. If you have any question concerning this policy before submitting an assignment, please ask for clarification.
Grading Detail – Data Roundtables

Do 3 of these, 10 points each, one in each Section.

• Grade distribution:
  – Written review: 3 points on content of review, 2 points on writing. Reviews should be 3-4 typed pages (12 pt. font).
  – Oral presentation: 3 points on presentation slides, 2 points on presentation style.

• Roundtable sources should come from the designated materials on the class website

• Students are responsible for self-organizing so that each student presents a different Roundtable article.

• All written reviews must be turned in at the beginning of the class during which you do your oral presentation. Please send a copy of the presentation slides and a .pdf of the review to bermaf@rpi.edu.
Grading Detail – Data Roundtable Written Review

Each **written review** (3-4 pages) should include:

1. **Succinct summary of the points of the article**
   - What is the article about? What issues does it focus on?
   - Why are these issues interesting/important?
   - What is newsworthy in the article and why?

2. **Data Issues:**
   - How does digital data play a role in the article?
   - What is the “data backstory”, i.e. what data infrastructure, policy, practice, etc. is needed to be there for the data to play its role?

3. **Next steps for exploration**
   - Where you would go next to find out more about these issues. What would you read?
   - What would be appropriate/interesting areas for exploration based on this article?

4. **Your thoughts:**
   - If the article is about a report, article, paper or other source material, did they do an accurate job of representing it?
   - Did you like the article? Why or why not?

**Written review Grading Metrics:**

**Content (3%):**
- Does the review content demonstrate a clear understanding of the material?
- Are the main points and issues clearly described?

**Writing (2%):**
- Is the review well-organized and readable by non-specialists?
- Does the review “tell a story”?
- Are the “next steps for explanation” and “thoughts” section thought-provoking and interesting?
Grading Detail – Data Roundtable Oral Presentation

Oral presentation components (10 minutes presentation + 5 minutes Q&A):

1. **Summary:**
   - What is the article about?
   - What is the point of view presented in the article?

2. **What are the data issues?**
   - How is data used to support the article’s point of view? Does it succeed in doing this?
   - What is the “data backstory”, i.e. what data infrastructure, policy, practice, etc. needed to be there for the data to play its role?

3. **How is the article useful in a broader context?**
   - What questions arise from reading this article?
   - How can the article “takeaways” be applied in other settings?

**Note:** You may need to read additional publications, websites for your presentations and reviews

Oral Presentation Grading Metrics:

**Talk (3%):**
- Does the speaker understand and communicate well about their topic?
- Does the presentation tell an interesting story?
- Is the speaker well prepared for questions?

**Slides (2%):**
- Are the slides well-organized and informative?
- Do the slides help tell the story?
- Are the slides visually interesting?
Grading Detail – Extra credit Op-Ed

• Grade distribution: 3 points on editorial content (ideas, thesis, and support), 2 points on writing (does it work as an op-ed, is it compelling, does it make sense)

• Op-eds must be turned in before the beginning of class on May 1.

• Op-eds should be in 12 pt. font and between 500 and 1000 words

• FYI: See http://www.nytimes.com/2013/10/14/opinion/op-ed-and-you.html?pagewanted=all&_r=0 (pdf on course website) for an article on writing an op-ed for the NY Times.
Op-Ed Detail -- Structure

Not all Op-Eds are like this, but many good Op-Eds have this structure:

• Lede – Lead-in around a news hook or personal experience

• Thesis – your position (explicit or implied)

• Argument – should be based on evidence (stats, news, reports, expert quotes, scholarship, history, experience). Arguments often presented as a series of points.

• Criticism pre-emption – take the lead in acknowledging the flaws in your argument and address potential counter-arguments

• Conclusion – circle back to lede?

Lede Options
• Current news
• Dramatic or personal anecdote
• Reference to popular culture or twist on conventional wisdom
• Anniversary of an event
• Major new study
Grading Detail – Section 2 Undergrad Paper

Specs

• Paper: 6-8 pages, 1.5 spaces, 12 pt font
• PDF due to bermaf@rpi.edu by 8:45 a.m. on April 3.
• Focus of paper should be an area of science or society that has been transformed by the availability of digital data

• General outline:
  – Description of the area and how data has transformed it
  – Specifics on the kind of innovation in the application of data has made this possible
  – What kind of infrastructure is needed to make this possible
  – Conclusion and thoughts about transformative potential of data in this area in the future

• Paper should include adequate references and bibliography (not included in the page count)

• Send a 1-2 page description of the topic of the paper and a detailed outline in .pdf to bermaf@rpi.edu by 12:00 a.m. on March 9.
UG Section 2 Paper Grading Metrics (20 points total)

Content (10 points):
- Does the paper content provide adequate depth and evidence to describe the transformation of an area by digital data?
- Are the references reasonable and adequate?

Writing (10 points):
- Is the paper well-organized, readable by non-specialists, and credible to specialists?
- Is the writing articulate and compelling?
- Is the paper well-structured with the main points backed up by evidence?
Grading Detail – Section 2 Grad mini-proposal

Specs

• Mini-Proposal: 10 pages, 1.5 spaces, 12 pt font
• PDF due to bermaf@rpi.edu by 8:45 a.m. on April 3.
• Focus of mini-proposal is a research project in an area of science or society that has been transformed by the availability of digital data.
• Target solicitation: NSF Computational and Data-Enabled Science and Engineering (CDS&E) program (http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504813&org=CISE&sel_org=CISE&from=fund)
• Mini-proposal should include:
  – 1 page proposal summary providing Description, Intellectual Merit, and Broader Impacts sections
  – 9 page proposal description including:
    • Introduction (what is the proposal about?)
    • Related work (relevant to proposed project)
    • Project plan (plan / approach for accomplishing the work)
    • Metrics of success
    • Conclusion
• Mini-proposal should include adequate references and bibliography (not included in the page count)
• Send a 1-2 page description of the topic of the paper and a detailed outline in .pdf to bermaf@rpi.edu by 12:00 a.m on March 9.
Grad Mini-proposal Grading Metrics (20 points total)

Content (10 points):

• Does the project have a clear focus and a research plan?

• Do the metrics of success adequately support the goals and approach?

• Is the project a departure from related work and are the references reasonable and adequate?

Writing (10 points):

• Does the proposal follow the guidelines provided? Is it well-organized, readable by non-specialists, and credible to specialists?

• Is the writing articulate and compelling?

• Is the proposal well structured?
Heilmeier’s Catechism (from Wikipedia)

George Heilmeier was former Director of DARPA (Defense Advanced Research Projects Agency), former CTO of Texas Instruments, former President of Bellcore, and former CEO of SAIC.

Heilmeier’s Catechnism is a set of questions credited to Heilmeier that anyone proposing a research project or product development effort should be able to answer:

• What are you trying to do? Articulate your objectives using absolutely no jargon.
• How is it done today, and what are the limits of current practice?
• What's new in your approach and why do you think it will be successful?
• Who cares?
• If you're successful, what difference will it make?
• What are the risks and the payoffs?
• How much will it cost?
• How long will it take?
• What are the midterm and final "exams" to check for success? (metrics of success)