Data and Society
Syllabus and Course Information

Spring 2016
Data and Society
CSCI 6370 (Grads) / 4370 (Undergrads)

• Professor: Dr. Fran Berman
• Office: AE 218
• Office Hours: Friday 1-2 or by appointment (send email to bermaf@rpi.edu)
• Course website: http://www.cs.rpi.edu/~bermaf/Data%20Course%20-%202016.html
Data and Society
CSCI 6370 (Grads) / 4370 (Undergrads)

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
Data and Society focuses on how data is changing how we work and live. The course provides an overview of the data-driven world touching on big data, the Internet of Things, data rights and privacy, data and health, and other topics. Course meetings consist of lectures, student presentations and guest speakers. Guest speakers this semester include Dr. Phil Bourne, Associate Director for Data Science at the National Institutes of Health, and Professor Bulent Yener from RPI.

Data and Society is designed to help develop the sophisticated data literacy needed to navigate today's world. It is also a communication-intensive course. Assignments will help build writing, presentation, and critical thinking and assessment skills, all of which are important for professional success. Both graduates and undergraduates are welcome. For additional information, contact Professor Fran Berman at bermaf@rpi.edu.

Graded work for the course will focus on helping students develop critical communication and assessment skills needed for professional success. Students' grades will be based on two written reviews and oral presentations from class "data roundtables", two midterms, an op-ed, and class participation. For more information, please contact the instructor.
Syllabus

January 29: L1 -- Intro: The Data Ecosystem / Data Roundtable
February 12: L3 -- Data-driven Science / L2 Data Roundtable
February 19: L4 -- Future Infrastructure -- Internet of Things / L3 Data Roundtable
February 26: Section 1 Exam / L4 Data Roundtable
March 4: Paper assignment description / Section 1 Data Roundtable
March 11: L5 -- Data and Health -- Phil Bourne Guest Lecture / Section 2 Data Roundtable
March 18: Spring Break / no class
March 25: L6 -- Data and Entertainment / L5 Data Roundtable
April 1: L7 -- Big Data Applications / L6 Data Roundtable
April 8: L8 -- Data in the Global Landscape / L7 Data Roundtable
April 15: L9 -- Digital Rights / L8 Data Roundtable
April 22: L10 -- Data Security and Privacy -- Bulent Yener Guest Lecture / L9 Data Roundtable
April 29: L11 -- Digital Governance and Ethics / L10 Data Roundtable
May 6: Section 3 Exam / L11 Data Roundtable
<table>
<thead>
<tr>
<th>Section Theme</th>
<th>Date</th>
<th>First “half”</th>
<th>Second “half”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section 1:</strong> The Data Ecosystem - - Fundamentals</td>
<td>January 29</td>
<td>Class introduction; Digital data in the 21st Century (L1)</td>
<td>Data Roundtable / Fran</td>
</tr>
<tr>
<td></td>
<td>February 5</td>
<td>Data Stewardship and Preservation (L2)</td>
<td>L1 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>February 12</td>
<td>Data-driven Science (L3)</td>
<td>L2 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>February 19</td>
<td>Future infrastructure – Internet of Things (L4)</td>
<td>L3 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>February 26</td>
<td>Section 1 Exam</td>
<td>L4 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>March 4</td>
<td>Paper assignment description</td>
<td>Section 1 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>March 11</td>
<td>Data and Health: Phil Bourne guest lecture (L5)</td>
<td>Section 1+2 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>March 18</td>
<td>Spring Break / no class</td>
<td></td>
</tr>
<tr>
<td></td>
<td>March 25</td>
<td>Data and Entertainment (L6)</td>
<td>L5 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>April 1</td>
<td>Big Data Applications (L7)</td>
<td>L6 Data Roundtable / 5 students</td>
</tr>
<tr>
<td><strong>Section 2:</strong> Data and Innovation – How has data transformed science and society?</td>
<td>April 8</td>
<td>Data in the Global Landscape (L8) Section 2 paper due</td>
<td>L7 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>April 15</td>
<td>Digital Rights (L9)</td>
<td>L8 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>April 22</td>
<td>Bulent Yener Guest Lecture, Data Security (L10)</td>
<td>L9 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>April 29</td>
<td>Digital Governance and Ethics (L11)</td>
<td>L10 Data Roundtable / 5 students</td>
</tr>
<tr>
<td></td>
<td>May 6</td>
<td>Section 3 Exam</td>
<td>L11 Data Roundtable / 5 students</td>
</tr>
</tbody>
</table>
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, Data Analytics 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 and infrastructure
• 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:
• Phil Bourne, NIH
• Bulent Yener, RPI CS Professor
Course Information
http://www.cs.rpi.edu/~bermaf/Data%20Course%202016/Data%20Course%20-%202016.html

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

• Reference and Roundtable materials will be on the web
  – Embedded reference materials in the lecture will be given by URL.
    Lectures will be on the web.

• Reference and reading materials may be tested on the Section exams
How you’ll be graded

Student grades are computed from:

- **2 section exams** (20 points each)
- **1 section paper** (20 points):
  - Undergrads: 7 page research paper on an approved topic
  - Grads: 10 page research paper on an approved topic
- **Class participation** (10 points)
- **1 Op-Ed** (10 points, due before April 29)
- **2 Data Roundtable reviews / presentations** (roughly 1 per section, 10 points each)

Time permitting, one “do-over” op-ed or data roundtable may be accepted. More information on whether that is possible will be provided after Section 1.
More about grading
(additional grading specifics given in course lectures on website)

• **Data Roundtables:**
  – Students are responsible for scheduling their Data Roundtables and ensuring that both are done. Information about Data Roundtables will be given during Lecture 1.

• **Op-Eds, Papers:**
  – Information about Op-Eds will be given during Lecture 2. Information about the Paper Assignment will be given on March 4.

• **Class engagement / attendance:**
  – Students are expected to attend 14/15 out of the class meetings. 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.

• **There will be a slightly different workload for grad students and undergrads**
  – Section 2 paper lengths are different.
  – In writing and presentations, each student will be assessed at a level appropriate to their educational level (undergrad or grad)
Grading Detail – Data Roundtables

Do 2 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 will be given in class and can be found on the class website.

- All written reviews and pptx slides must be turned in before 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 Backdrop:
   • 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. Your thoughts / assessment:
   • If the article is about a report, article, paper or other source material, did they do an accurate job of representing it?
   • What should be done next to explore this issue?
   • 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”?
   • Is the “thoughts / assessment” 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 / backdrop?
   • 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?
   • Where would you go from here if you were interested in this topic?

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?
• Is the presentation compelling?
• Does the presentation tell an interesting story?
• Did the speaker use the timeframe effectively?
• 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?
Data Roundtable Presentation format

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

2. **What are the data issues / backdrop?**
   - 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?
   - Where would you go from here if you were interested in this topic?
Section 2 Paper Assignment

• **Section 2 paper** (20 points):
  – **Undergrads**: 7 page research paper on an approved topic (does not include references)
  – **Grads**: 10 page research paper on an approved topic (does not include references)
  – **Paper due April 8, 8:45 a.m.**

• **1-page Outline due March 7, 12 noon**
Paper Structure

Specs

• Paper: 7 pages (undergrad) or 10 pages (grad), double-spaced, 12 pt font
• Final paper pdf due to bermaf@rpi.edu by 8:45 a.m. on April 8.
• 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
  – Specifics on the kind of data infrastructure is needed to support this transformation
  – 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).
  – If you use material from a source, reference it in the bibliography.
  – If you copy material from a source, put it in quotes and reference it in the bibliography.
One-page outline (pdf) due to bermaf@rpi.edu by March 7 before 12:00 noon

• Focus of paper should be an area of science or society that has been transformed by the availability of digital data

• Please provide a one-page description of your paper project including the items below. (This will not be graded; its purpose is to provide you with early feedback on your project).

  – Title/Topic of Paper: What are you planning to write about?
    • Examples from 2015: “Sports: The Data Revolution”, “The VA Data Backlog”, “Mobil Banking in Africa”

  – Outline of Paper:
    • Introduction: Major approach you plan to take, what aspect of the topic will you focus on?
    • Innovation: What areas of innovation will you highlight?
    • Infrastructure: What aspects of data infrastructure will you focus on?
    • Summary / Conclusion: What would you like the reader to take away from the paper
    • Bibliography: Initial references that are pertinent
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?
- Does the paper include specifics on both innovation and infrastructure?
- Is the writing articulate and compelling?
- Is the paper well-structured with the main points backed up by evidence?
Grading Detail – Op-Ed

• Grade distribution: 5 points on editorial content (ideas, thesis, and support), 5 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 April 29.

• 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.
Writing op-eds

• Op-eds can have tremendous influence on community and stakeholders
  – Can establish you as an expert
  – Can get your point of view into the public discourse
  – Can be useful to your company, project or community

• Who is your audience: General public

• What is your purpose: Persuasively get your point of view across
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
Op-Ed Tips

• Write it in a way that smart people can relate to, even if they are not in your discipline. Don’t use buzzwords or talk “inside baseball” without explaining things.

• Pay attention to publication word count – op-eds are usually quite short

• *If you do this for real:*
  – The final version may be reviewed and/or edited – what you send in may not be the final draft
  – Do your homework – everyone will read this
  – Be prepared for feedback – blogs, tweets, etc.
# 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.