Introduction to the course, Introduction to digital data, Op-ed instructions, Presentation model
Announcements 1/20

• Welcome!

• Please sign attendance sheet **each time** you are here (your participation grade depends partly on attendance).
  – If you are on the waiting list and trying to get in the class, please do the same.

• If you decide to drop the class, please let me know ([bermaf@rpi.edu](mailto:bermaf@rpi.edu)) and I will let in someone on the waiting list.

• No Wednesday class January 25.

• Discussion article for next week (Friday, January 27). **Please read:**
Welcome to Data and Society, Low 3039
CSCI 6370 (Grads) / 4370 (Undergrads)
ITWS 6960 (Grads) / 4960 (Undergrads)

• Professor: Dr. Fran Berman
• Office: AE 218, 276-3794
• Office Hours: Friday 1-2 or by appointment (send email to bermaf@rpi.edu)
• Course website (linked off Fran’s RPI web page): http://www.cs.rpi.edu/~bermaf/Data%20Course%202017/Data%20Course%20-%202017.html
Today (1/20/17)

• Why Data and Society?

• Intro – about this course
  – Syllabus and grading expectations
  – Learning objectives and expectations
  – Why are you here?

• Lecture 1

• Break

• Op-Ed instructions, draft due February 10

• Data Presentation (Fran)
Data-driven innovation is a priority nationally, internationally, and in all sectors.
There’s more to the data story than technology.

- **Policy and regulatory issues**
- **Workforce evolution**
- **New possibilities for innovation / new challenges for infrastructure**
- **Privacy and rights**
- **New modes of social and community interaction, organization**

VR image: https://www.flickr.com/photos/nanpalmero/16237219524
Fran Berman, Data and Society, CSCI 4370/6370

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 topics:**

• Data Applications
• Data Infrastructure and Stewardship
• Data and Society

**Guest Speaker this Semester:**

• Hilary Mason, Fast Forward Labs

---

Fran Berman, Data and Society, CSCI 4370/6370
<table>
<thead>
<tr>
<th>Wednesday Section</th>
<th>Friday lecture</th>
<th>First Half of Class</th>
<th>Second Half of Class</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 25: NO class</td>
<td>January 27</td>
<td>L2: Big data applications / Data and the election; Data and Target; Discussion</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>February 1: 6 presentations</td>
<td>February 3</td>
<td>L3: Data and Health / PDB, Precision Medicine; Discussion</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>February 8: NO class</td>
<td>February 10</td>
<td>L4: Data and Science / Earthquakes, LHC; Paper Instructions</td>
<td>4 Presentations</td>
<td>Op-Ed Draft Due</td>
</tr>
<tr>
<td>February 15: 6 presentations</td>
<td>February 17</td>
<td>L5: Data Cyberinfrastructure; Discussion</td>
<td>4 Presentations</td>
<td>Op-Ed Draft Back</td>
</tr>
<tr>
<td>February 22: 6 presentations</td>
<td>February 24</td>
<td>L6: Data Stewardship and Data Preservation; Discussion</td>
<td>4 presentations</td>
<td>Op-Ed Final Due</td>
</tr>
<tr>
<td>March 1: NO class</td>
<td>March 3</td>
<td>NO class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 8: 6 presentations</td>
<td>March 10</td>
<td>L7: Data Futures – Internet of Things; Discussion</td>
<td>4 presentations</td>
<td>Paper Draft Due</td>
</tr>
<tr>
<td>March 15: Spring Break</td>
<td>March 17</td>
<td>Spring Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 22: NO class</td>
<td>March 24</td>
<td>L8: Data rights and policy / U.S. and EU; Discussion</td>
<td>4 presentations</td>
<td></td>
</tr>
<tr>
<td>March 29: 6 presentations</td>
<td>March 31</td>
<td>Op-Ed Pecha-Kucha</td>
<td></td>
<td>Paper Draft Back</td>
</tr>
<tr>
<td>April 5: NO class</td>
<td>April 7</td>
<td>NO class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 12: 4 presentations</td>
<td>April 14</td>
<td>Hilary Mason Guest Lecture</td>
<td>4 presentations</td>
<td>Final Paper Due</td>
</tr>
<tr>
<td>April 19: 4 presentations</td>
<td>April 21</td>
<td>L9: Data and Ethics; Discussion</td>
<td>4 presentations</td>
<td></td>
</tr>
<tr>
<td>April 26: 6 presentations</td>
<td>April 28</td>
<td>Paper Pecha-Kucha</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is a PechaKucha?

• “PechaKucha (Japanese: ペチャクチャ, IPA: [petɕa kʊtɕa], chit-chat) is a presentation style in which 20 slides are shown for 20 seconds each (6 minutes and 40 seconds in total). The format, which keeps presentations concise and fast-paced, powers multiple-speaker events called PechaKucha Nights (PKNs).” [Wikipedia]

• Later in the semester, we’ll do a pechakucha class for op-ed presentations (March 24) and a pechakucha class for research papers (April 28).
  – These are optional for speakers
  – If you are a speaker, your pechakucha grade can be used to replace one of your presentation grades if it is higher

• TK Woodstock will be “PechaKucha TA”. More later.
Course Information

ww.cs.rpi.edu/~bermaf/Data%20Course%202017/Data%20Course%20-%20202017.html

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

• Discussion articles and presentation articles will be on the web
  – Embedded reference materials in the lecture will be given by URL. Lectures will be on the web.

• Discussion articles will be assigned prior to the Friday class in which they will be discussed. You are responsible for reading them and for participating in the discussion.
How You’ll be Graded

Research Paper Draft: 16 points
Research Paper Final: 16 points
Op-Ed Draft: 14 points
Op-Ed Final: 14 points
Participation: 10 points
Presentation 1: 15 points
Presentation 2: 15 points

Pecha-Kucha (optional): Score is 15 points, can be used to replace your lowest presentation score

No late work.
More about grading
(additional grading specifics for Presentations and Op-eds later today)

• **Presentations:**
  – Students are responsible for scheduling their Presentations and ensuring that both are done. Information about Presentations will be given at the end of today’s class.

• **Op-Eds, Papers:**
  – Information about Op-Eds will be given today. Information about the Paper Assignment will be given on February 10 (Lecture 4).

• **Class participation / attendance:**
  – **Students are expected to attend 18+ out of 20 Wednesday and Friday class meetings.** Attendance will be taken in class.
  – Participation grade: 5% attendance, 5% class participation (Discussions, Questions for speakers, etc.)

• **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)
# 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.
Why are you here?

1. Name, major, school, grad or undergrad?

2. What do you expect to be doing after you finish your degree?

3. Why did you take this course?

4. What do you hope to get out of this course?

5. What areas / topics in the data landscape are of most interest to you?

6. What is the coolest recent thing you’ve heard about digital data?
Lecture 1: Introduction to Data
Lecture 1 Outline

• Some fundamentals
  – Data basics
  – How much data is there and where does it come from? What does the “data universe” look like?

• Data Transformation -- The Information Age
What is Digital Data?

- *Wikipedia*: “Digital data, in information theory and information systems, are discrete, discontinuous representations of information or works, as contrasted with continuous, or analog signals which behave in a continuous manner, or represent information using a continuous function.

- Although digital representations are the subject matter of discrete mathematics, the information represented can be either discrete, such as numbers and letters, or it can be continuous, such as sounds, images, and other measurements.”
Digital data comes from everywhere

- Entertainment
- Education
- Health
- Commerce
- Physical Infrastructure and Smart Systems
- Research
- Communication / Community
All data are not alike

- **Volume**: amount of data, number of files
- **Velocity**: Rate at which data flows into an organization as well as speed of the “feedback loop” (can the data be where you want it when you want it)
- **Variety**: Diversity of data types and sources (“messiness” of using, combining, managing data)
- **Value**: importance of the data
- **Volatility**: how quickly data changes, how long the data is useful for

- **Validity**: legitimacy / accuracy of sources
- **Viscosity**: resistance to flow in the volume of data (improved infrastructure, management, and technologies can reduce viscosity)
- **Virality**: how quickly the data is dispersed and shared
- **Variability**: Extent to which data points differ from each other. (commonly used measures of variability: range, mean, variance and standard deviation)

Sources:  
http://www.theblueballroom.com/articles/do-you-know-how-to-use-big-data-in-your-business,  
http://strata.oreilly.com/2012/01/what-is-big-data.html;  
How we access, manage, use, store and preserve data also varies widely

- **RETENTION TIMEFRAME:**
  - Short-term (few minutes, months, years) to long-term (decades, centuries, ...)

- **SIZE / SCALE:**
  - Small-scale (KBs, GBs, MBs) to large-scale / “big” (TBs, PBs, EBs)

- **PREPARATION:**
  - Well-tended (curated, sufficient metadata, cleaned and filtered) to poorly tended (flat files, insufficient metadata)

- **POLICY / REGULATION RESTRICTIONS:**
  - Subject to more restrictive policy and regulation (e.g. HIPAA) vs. subject to less restrictive policy and regulation

- **LIFE CYCLE PLANNING:**
  - Has a data management and / or sustainability plan vs. ad hoc approach

- **COMMUNITY ACCESSIBILITY:**
  - Shared with others in the community vs. kept private; Curated and organized using community standards vs. ad hoc or home-grown approaches
Meaning and context increase the impact of data: Data, Information, Knowledge, Wisdom (DIKW)

- **Data** = Qualitative or quantitative values at the lowest level of abstraction
- **Information** = Data and its associated meaning
- **Knowledge** = Theoretical or practical understanding of information
- **Wisdom** = The quality of having experience, knowledge and good judgment

Considerable overlap and many definitions ...

*Many articles use digital data and digital information interchangeably. For the most part, we will too.*

... Where is the life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information? ...

First recorded instance of DIKW in 1934 poem “Choruses from the Rock” by T.S. Eliot.
How Much Data is There?

- There won’t be an exaflop supercomputer until the end of the decade+
- We have had exabytes of data for at least 10 years and hit a zettabyte in 2009-2010

![The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020](chart.png)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilo</td>
<td>$10^3$</td>
</tr>
<tr>
<td>Mega</td>
<td>$10^6$</td>
</tr>
<tr>
<td>Giga</td>
<td>$10^9$</td>
</tr>
<tr>
<td>Tera</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>Peta</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>Exa</td>
<td>$10^{18}$</td>
</tr>
<tr>
<td>Zetta</td>
<td>$10^{21}$</td>
</tr>
<tr>
<td>Yotta</td>
<td>$10^{24}$</td>
</tr>
</tbody>
</table>

Digital universe doubling every two years

How big is ... (roughly)

- **A byte**: ~ 1 character
- **A megabyte**: ~ 1 small novel
- **A gigabyte**: 50 gigabytes → ~ 1 floor of books
- **A terabyte**: 10 Terabytes → ~ printed collection of the Library of Congress
- **A petabyte**: 2 Petabytes → ~ All US academic research libraries
- **An Exabyte**: 5 exabytes ~ all words ever spoken by human beings recorded in text
- **A Zettabyte**: 42 zettabytes → ~ all words ever spoken by human beings digitized as audio

Source: http://highscalability.com/blog/2012/9/11/how-big-is-a-petabyte-exabyte-zettabyte-or-a-yottabyte.html
Where does all the digital data come from?

- Digital information comes from many sources: computers, RFIDs and sensors, scientific and other instruments, imaging devices and surveillance, cell phones, etc.
- Most of the digital universe is transient – unsaved Netflix streams, temporary routing information in networks, sensor signals discarded when no alarms go off, etc.
- 2014: Digital universe = 1.7 MB/minute for every person on earth
- Areas experiencing increasing data analysis and use:
  - Surveillance footage
  - Embedded and medical devices
  - Entertainment and social media
  - Consumer images
  - Enterprise transactional data; data processing
  - Internet of Things and smart applications
Huge growth in data from the “Internet of Things”

- Major growth spurts in the digital universe:
  - Film $\rightarrow$ digital technology
  - Analog functions monitoring and managing physical world $\rightarrow$ digital functions involving communications and software telemetry
  - Analogue TV $\rightarrow$ digital TV
  - Increasing data from embedded systems

- IDC estimates that
  - In 2014, things in the digital universe approaching 200 billion, 10% (20 billion) of those wired and communicating with the Internet
  - In 2020, roughly 30 billion connected devices in the digital universe

Huge growth in mobile “things” – connected and otherwise

Finally, a good portion of the digital universe will be generated by mobile devices and people – from 17% in 2013 to 27% in 2020 – but the percentage of mobile “things” in the IoT will be more than 75% by 2020.

More growth trends in the Digital Universe

- Growth of the Internet (> 1 billion users) and broadband availability
- Conversion of formerly analog information to digital
- Falling prices and increased performance for digital devices; ability to store more information and share it in standard formats
- Rise of automation, data-intensive, graphics-intensive, and “smart” applications
- Rise of data centers, cloud computing, social networks
- Regulations mandating new archiving and privacy protection rules
- Increased computerization of business, education, entertainment, etc.
- Growth in access, markets, prevalence of digital technologies
- Growth of the Internet of Things
Which data is useful?

• **Data is useful when we know something about it** – what it represents, where it was collected, what units are being used, etc.? **Metadata** a critical part of the data universe that makes data useful.

• IDC estimates tremendous growth in useful data from 2014 to 2020
All digital data cannot be stored

- 2007 was the “crossover year”: Began to generate more digital data than storage to keep it.
- In 2013, current storage capacity could hold just 33% of the digital universe.
2010 Update on the Storage Gap:
By 2020, more than twice as much information will be created as storage available

Figure 5: The Emerging Gap
Information Creation > Storage Available

Source: IDC Digital Universe Study, sponsored by EMC, May 2010

Fran Berman, Data and Society, CSCI 4370/6370
The “Digital Shadow”

- Less than half of your digital footprint is related to individual actions – taking pictures, making VoIP calls, uploading files, etc.

- The rest of your digital footprint is “ambient” content and metadata related to you: surveillance images, banking records, medical records, information about your web searches and behavior in social networks, etc.
Who can collect, has rights to, and can use information about you, and under what conditions is the subject of national discussions world-wide.

The information about you is much greater than the information you create yourself. This is called your...

Digital Shadow

... and it's growing continuously.

Need for data and IT-savvy professionals having tremendous impact on the workforce

From McKinsey Report on Big Data:
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

140,000–190,000 more deep analytical talent positions, and 1.5 million more data-savvy managers needed to take full advantage of big data in the United States.
How did IDC calculate their estimations?

- Forecasts developed for 40+ classes of devices and/or applications that can capture or create digital information.
- Estimate annual usage and number of times a unit of information is replicated, either to share or store.
- Analysis based on previous IDC research, information capture and workload characteristics, surveys, studies, etc.
- Data adjusted for geographic region, kind of device, kind of information, etc.
Data Transformation: The Information Age
The Information Age

• “The Information Age (also known as the Computer Age, Digital Age, or New Media Age) is a period in human history characterized by the shift from traditional industry that the industrial revolution brought through industrialization, to an economy based on information computerization. The onset of the Information Age is associated with the Digital Revolution, just as the Industrial Revolution marked the onset of the Industrial Age.” Wikipedia
How did the Industrial Revolution Transform the World?

• Transition to new manufacturing processes in late 18th / early 19th century.
  – Hand production → machines, new chemical manufacturing, new iron production processes
  – Improved efficiency of water power and the increased use of steam power
  – Wood and bio-fuels → coal
  – England → Western Europe, US

• Major turning point, almost every aspect of daily life influenced in some way
Technological Innovation during the Industrial Revolution

- New technological capabilities and emerging needs had a transformative effect on:
  - Work opportunities and workforce needs
  - National and international priorities
  - Economic, cultural, social, and political structures
  - Leading sectors (manufacturing, health, energy) and new areas for innovation and impact. Broad ripple effect from both.

Jacquard Loom – Precursor to the Programmable Computer

• **Jacquard loom** invented by Joseph Marie Jacquard and first demonstrated in 1801.

• Loom controlled by punch cards for the purpose of manufacturing textiles with complex patterns.
  
  – Rows of holes were punched on each card corresponded to one row of the design.

• Loom serves as an important conceptual precursor in the development of computer programming

Social Innovation during the Industrial Revolution

- **Economic transformation**
  - Better standard of living
  - Better agricultural practices, housing, food supplies
  - Less expensive clothing and consumer goods

- **Urbanization**
  - Rise of factories and modern cities
  - Change in employment options

- **Social policy**
  - Child Labor laws
  - Growth in trade unions

*Cottonopolis* is a name given to the city of Manchester, in England. It denotes a metropolis of cotton and cotton mills, as inspired by Manchester's status as the international centre of the cotton and textile processing industries during this time.

Engraving by Edward Goodall (1795-1870), original title *Manchester, from Kersal Moor* after a painting of W. Wylde. Wikipedia (cropped from original)
Fast forward to the Information Age

• We are experiencing a transformation analogous to the Industrial revolution

• New technological capabilities and emerging needs again having a transformative effect on
  – Work opportunities and workforce needs
  – National and international priorities
  – Economic, cultural, social, and political structures
  – Leading sectors and new areas for innovation and impact. Broad ripple effect from both.
Transformative Potential of Data: Emerging Technologies

Exascale computing →
more compute and data at all tiers in the Branscomb Pyramid.

New breakthroughs in power and computer architectures required.

Smart Devices, Sensor Networks →
More data-enabled devices and approaches drive crowd-sourced, real-time, and other aggregation applications

Information-Driven Analysis →
X-informatics and X-analytics enable new targets for data-driven research and decision-making models

Fran Berman, Data and Society
Transformative Potential of Data:
Massive-scale coordination, inclusion, access

Greater access →
Greater participation, “democratization” possible

High quality, on-line education → On-line / on-site education solutions have the potential to transform higher education

Greater transparency, management, monitoring →
More measurement, transparency, monitoring possible
Social Impacts – adequate legal, regulatory, and policy underpinnings for data needed

• How do you maintain personal freedom and sufficient privacy / control over your information?

• What are your rights?
  – What do you own?
  – What can you distribute?
  – What can you charge for?

• What / whom do you trust?
  – Your data?
  – Your respondent?
  – Your hardware?
  – Your system / software?
What happens when digital data becomes the vehicle for progress in the Information Age?

• How is digital data being used to drive new innovation? -- **How do we make the most out of data?**

• What kind of technological and human infrastructure is needed to support the access, management, use and re-use of digital data today and tomorrow? -- **How do we create a useful data ecosystem?**

• What social and community constructs are needed to realize data’s potential? – **What’s needed for a digitally responsible society?**

Fran Berman, Data and Society, CSCI 4370/6370
Towards a sustainable ecosystem

*Sustainable development:* "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."

*Our Common Future, U.N. Brundtland Commission*

- **Key components**
  - Ecological sustainability
  - Cultural / institutional sustainability
  - Economic sustainability
  - Political sustainability

Towards a sustainable data ecosystem

*Sustainable development*: "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."

*Our Common Future, U.N. Brundtland Commission*

- **Key components**
  - Ecological sustainability
  - Cultural / institutional sustainability
  - Economic sustainability
  - Political sustainability
Next time: Big Data and Data Applications

- Read for next Friday’s discussion (article on the website):
  - “Eight (no Nine) Problems with Big Data”, New York Times,
Lecture 1 Sources (not already in text)

• “The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things”, IDC report

• Previous IDC Reports: http://www.emc.com/leadership/digital-universe/index.htm#Archive


Break
How You’ll be Graded

Research Paper Draft: 16 points

Research Paper Final: 16 points

Op-Ed Draft: 14 points

Op-Ed Final: 14 points

Participation: 10 points

Presentation 1: 15 points

Presentation 2: 15 points

Pecha-Kucha (optional): Score is 15 points, can be used to replace your lowest presentation score
Op-Eds
Why is it good to know how to write an op-ed?

• Op-ed is a good example of a brief, persuasive communication to advance your point of view
  – Op-eds can have tremendous influence on community and stakeholders
  – Can establish you as an expert
  – Can be useful to your company, project or community
  – Can get your point of view into a more public discourse

• **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 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* (i.e. send it in to a publication rather than do it for class):
  – 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.

Fran Berman, Data and Society, CSCI 4370/6370
Grading Detail – Op-Ed
(Draft and Final each 14 points)

• Grade distribution for draft / final:
  – 7 points on editorial content: ideas, thesis, and support
  – 7 points on writing: does it work as an op-ed, is it compelling, does it make sense?

• Draft op-eds due February 10 at the beginning of class. Bring a hard copy to class and turn it in at the beginning. Also turn in a .pdf file using Submitty: https://submitty.cs.rpi.edu/index.php?semester=s17&course=csci4370
  – Op-ed drafts will be returned with comments on February 15.
  – Final op-eds due on February 22.
  – Op-ed grades: Draft grade (14 points) + Final grade (14 points)
  – Important note: If your draft op-ed is strong, you may choose to not turn in the final op-ed and double your draft grade (op-ed = 2 X draft grade). You must make this decision and let Fran know before February 22.

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

Do 2 of these, 15 points each

- You are responsible for ensuring that you sign up for 2 during the semester
- Note that you may do these on Friday Sessions or designated Wednesday Sessions
- Presentation articles will be given in class and can be found on the class website
- A print-out of your slides (please include your name) should be handed in by the beginning of the class in which you will present.

Point Distribution

- Content 4 pts
- Talk 5 pts
- Visuals 4 pts
- Q&A 2 pts

Fran Berman, Data and Society, CSCI 4370/6370
Grading Detail – Presentations

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

1. Summary:
   - What is the article about?
   - What are the main points of the article?

2. What are the data issues?
   - How is data used to support the article’s point of view?
   - What data infrastructure, policy, practice, etc. is needed to be there for the data to play its role?

3. What is the broader context?
   - What questions arise from reading this article?
   - What is the point of view presented in the article?

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

Presentation Grading Metrics :

Talk (5 pts):
- Is the presentation compelling?
- Does the presentation tell an interesting story?
- Did the speaker use the timeframe effectively?

Visuals (4 pts):
- Are the slides well-organized and informative?
- Do the slides help tell the story?
- Are the slides visually interesting?

Content (4 pts):
- Does the speaker understand the topic?
- Has the speaker leveraged appropriate additional materials as needed to support their presentation?

Q&A (2 pts):
- Is the speaker well prepared for questions? Can they respond to them articulately?
The crack in this Antarctic ice shelf just grew by 11 miles. A break could be imminent.  

Washington Post, 1/6/17

Growing rift can shrink Larsen C ice shelf in Antarctica

- Growing rift on the floating Larsen C ice shelf moving towards break with loss of ~2000 square miles of ice (almost as large as Delaware)
  - Larsen C ice shelf is > 1000 feet thick
  - 4th largest ice shelf in Antarctica
- Only 12 miles of remaining ice now connect the shelf to the “emerging mega-iceberg”.
  - Rift has grown 50 miles since 2011 and has widened to 1000+ feet
- Scientists fear the loss of ice shelves around the frozen continent will allow glaciers inland to slide faster towards the sea as temperatures rise, raising world sea levels.
  - “Calving” of ice shelf could lead to a 10 cm rise in global sea levels.
What happens when the ice shelf breaks ("calves")?

- Larsen C Ice Shelf will lose more than 10% of its area
- Ice front will be at its most "retreated" position ever recorded.
- Landscape of Antarctic Peninsula will change.
  - New configuration will be less stable than it was prior to the rift
  - Larsen C could disintegrate (similar to Larsen B after calving)
What would make it calve?

• British Antarctic Survey: Ice is being thawed both by warmer air above and warmer waters below.

• Scientists believe that jumps of the rift tip occurring in shorter time intervals the longer the rift gets.
  – Due perhaps to longer “lever” for the forces acting to advance the rift – up and down tides, strong winds towards the sea, etc.
  – Scientists believe it will go in the next few months
How do we know all this?

• Project MIDAS team (funded by U.K.) monitoring rift on Larsen C ice shelf.

• Data: Break not visible on optical data (cloud-free Landsat images) but radar analysis has shown a growth of the rift of about 18 km/11 mi).

• Growth expands 13 mile growth in the rift earlier this year.

Map of Antarctica showing the amount of melting of ice shelves from below. Blue shades represent melt rates of greater than 5 meters per year. Arrow points to Larsen C Ice Shelf

Image from Daily Mail

Fran Berman, Data and Society, CSCI 4370/6370
What is the data backstory?

• Models used to predict the rise of sea level validated by reproduction of previous high sea levels of the past.
  – Provides confidence that model will reasonably accurately predict future sea levels

• Data collected from satellites used in predictive models. Satellites equipped with altimeters that measure elevation by shooting a beam of radio waves or light waves to the ice surface. Altimeter records the time it take the waves to bounce off the surface and back to the satellite.

• Researchers can then map how ice elevation changes over time.

• Data must be curated, archived, annotated to provide useful information for models

Image from WP article
What are the repercussions of the rift on the Antarctic?

- Floating ice shelf fed by glaciers that sit above sea level on the Antarctic Peninsula
- When shelf loses mass, glaciers can flow more quickly and sea levels can rise
  - Losses of mass from ice shelf alone do not have the same effect as the shelf is already floating on water
- West Antarctic ice sheet sits in a deep bowl that extends far below sea level
  - If it loses its protective fringes of floating ice, the result is likely to be the formation of vast, sheer cliffs of ice facing the sea
  - Cliff height could make them unstable and ice sheet could experience rapid shrinkage as unstable cliffs collapse into the water
Does this have anything to do with Climate Change?

• For the last half century, climate scientists have thought that the West Antarctic ice sheet is potentially vulnerable to disintegration from a relatively small amount of global warming, and capable of raising the sea level by 12+ ft should it break up.
  – Researchers believed that this scenario would take 100’s or 1000’s of years to occur.
  – High emissions of heat-trapping gases could launch a disintegration of the ice sheet within decades, heaving enough water into the ocean to raise the sea level as much as 3 ft by 2100.
  – With ice melting in other regions, the total rise of the sea could reach 5-6 ft by 2100. Rise of sea could exceed 1+ ft per decade by the middle of 2200.

• Rising sea level could drown the world’s coastlines, including many great cities (NYC, Miama, New Orleans, London, Venice, Shanghai, Hong Kong, Sydney, etc.)

• Some NASA data shows that Antarctic ice sheet has been growing
  – Satellite data shows net gain of 112 billion tons of ice a year from 1992 to 2001. Snow accumulation adding more ice to East Antarctica. Interior region of West Antarctica is being lost as glaciers thin out.
  – Increases slowed between 2003 to 2008 to 82 billion tons a year. Losses could offset gains in years to come.
References


• “Iceberg the size of Delaware is “hanging by a thread” in Antarctica – and it could lead to a 10cm rise in global sea levels”, Daily Mail, http://www.dailymail.co.uk/sciencetech/article-4094942/Iceberg-size-Delaware-hanging-thread-Antarctica-lead-10cm-rise-global-sea-levels.html

Deconstructed Presentation
Grading Detail – Presentations

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

1. Summary and main points
   • What is the article about?
   • What are the main points of the article?

2. What are the data issues?
   • How is data used to support the article’s point of view?
   • What data infrastructure, policy, practice, etc. is needed to be there for the data to play its role?

3. What is the broader context?
   • What questions arise from reading this article?
   • What is the point of view presented in the article?

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

Presentation Grading Metrics:

Talk (5 pts):
• Is the presentation compelling?
• Does the presentation tell an interesting story?
• Did the speaker use the timeframe effectively?

Visuals (4 pts):
• Are the slides well-organized and informative?
• Do the slides help tell the story?
• Are the slides visually interesting?

Content (4 pts):
• Does the speaker understand the topic?
• Has the speaker leveraged appropriate additional materials as needed to support their presentation?

Q&A (2 pts):
• Is the speaker well prepared for questions? Can they respond to them articulately?
The crack in this Antarctic ice shelf just grew by 11 miles. A break could be imminent. Washington Post, 1/6/17

Growing rift can shrink Larsen C ice shelf in Antarctica

• Growing rift on the floating Larsen C ice shelf moving towards break with loss of ~2000 square miles of ice (almost as large as Delaware)
  – Larsen C ice shelf is > 1000 feet thick
  – 4th largest ice shelf in Antarctica

• Only 12 miles of remaining ice now connect the shelf to the “emerging mega-iceberg”.
  – Rift has grown 50 miles since 2011 and has widened to 1000+ feet

• Scientists fear the loss of ice shelves around the frozen continent will allow glaciers inland to slide faster towards the sea as temperatures rise, raising world sea levels.
  – “Calving” of ice shelf could lead to a 10 cm rise in global sea levels.
What happens when the ice shelf breaks ("calves")?

- Larsen C Ice Shelf will lose more than 10% of its area
- Ice front will be at its most "retreated" position ever recorded.
- Landscape of Antarctic Peninsula will change.
  - New configuration will be less stable than it was prior to the rift
  - Larsen C could disintegrate (similar to Larsen B after calving)
What would make it calve?

• British Antarctic Survey: Ice is being thawed both by warmer air above and warmer waters below.

• Scientists believe that jumps of the rift tip occurring in shorter time intervals the longer the rift gets.
  – Due perhaps to longer “lever” for the forces acting to advance the rift – up and down tides, strong winds towards the sea, etc.
  – Scientists believe it will go in the next few months

Image from WP article
How do we know all this?

- Project MIDAS team (funded by U.K.) monitoring rift on Larsen C ice shelf.
- Data: Break not visible on optical data (cloud-free Landsat images) but radar analysis has shown a growth of the rift of about 18 km/11 mi).
- Growth expands 13 mile growth in the rift earlier this year.

Map of Antarctica showing the amount of melting of ice shelves from below. Blue shades represent melt rates of greater than 5 meters per year. Arrow points to Larsen C Ice Shelf

Image from Daily Mail

Fran Berman, Data and Society, CSCI 4370/6370
What is the data backstory?

DATA ISSUES

- Models used to predict the rise of sea level validated by reproduction of previous high sea levels of the past.
  - Provides confidence that model will reasonably accurately predict future sea levels

- Data collected from satellites used in predictive models. Satellites equipped with altimeters that measure elevation by shooting a beam of radio waves or light waves to the ice surface. Altimeter records the time it take the waves to bounce off the surface and back to the satellite.

- Researchers can then map how ice elevation changes over time.

- Data must be curated, archived, annotated to provide useful information for models

Image from WP article
Fran Berman, Data and Society, CSCI 4370/6370
What are the repercussions of the rift on the Antarctic?

• Floating ice shelf fed by glaciers that sit above sea level on the Antarctic Peninsula

• When shelf loses mass, glaciers can flow more quickly and sea levels can rise
  – Losses of mass from ice shelf alone do not have the same effect as the shelf is already floating on water

• West Antarctic ice sheet sits in a deep bowl that extends far below sea level
  – If it loses its protective fringes of floating ice, the result is likely to be the formation of vast, sheer cliffs of ice facing the sea
  – Cliff height could make them unstable and ice sheet could experience rapid shrinkage as unstable cliffs collapse into the water

Image from NYT reference
Does this have anything to do with Climate Change?

- For the last half century, climate scientists have thought that the West Antarctic ice sheet is potentially vulnerable to disintegration from a relatively small amount of global warming, and capable of raising the sea level by 12+ ft should it break up.
  - Researchers believed that this scenario would take 100’s or 1000’s of years to occur.
  - High emissions of heat-trapping gases could launch a disintegration of the ice sheet within decades, heaving enough water into the ocean to raise the sea level as much as 3 ft by 2100.
  - With ice melting in other regions, the total rise of the sea could reach 5-6 ft by 2100. Rise of sea could exceed 1+ ft per decade by the middle of 2200.

- Rising sea level could drown the world’s coastlines, including many great cities (NYC, Miami, New Orleans, London, Venice, Shanghai, Hong Kong, Sydney, etc.)

- Some NASA data shows that Antarctic ice sheet has been growing
  - Satellite data shows net gain of 112 billion tons of ice a year from 1992 to 2001. Snow accumulation adding more ice to East Antarctica. Interior region of West Antarctica is being lost as glaciers thin out.
  - Increases slowed between 2003 to 2008 to 82 billion tons a year. Losses could offset gains in years to come.
References


• “Iceberg the size of Delaware is “hanging by a thread” in Antarctica – and it could lead to a 10cm rise in global sea levels”, Daily Mail, http://www.dailymail.co.uk/sciencetech/article-4094942/Iceberg-size-Delaware-hanging-thread-Antarctica-lead-10cm-rise-global-sea-levels.html

Your Turn ...
Presentation Articles for January 27

• January 27:


Presentation Articles for February 1

February 1:


Presentation Articles for February 3

February 3:

  http://www.lexology.com/library/detail.aspx?g=fa622c15-f2c4-4397-9d64-6cd341fcaf3f

- “Four Steps to Precision Public Health”, Nature, [Kusum B]
  http://www.nature.com/news/four-steps-to-precision-public-health-1.21089

- “Can IBM’s Watson do it all?”, Fast Company, [Erica B]

- “mHealth’s Year in Review: From Texting to Wearables to Telehealth’s Tricks (and Treats)”, mHealth Intelligence, [Tim T]
No class Wednesday. Class next Friday.

- *Next time (January 27):* Big data applications / Data and the election; Data and Target; Discussion

- *Read for January 27 Discussion:*