Big Data and Privacy

Rebecca Balebako

Advisor: Dr. Lorrie Cranor
Agenda

• Quiz
• Important dates
• Exam feedback
• Benefits of big data
• Concerns about big data
• What is PCAST
• PCAST big data recommendations
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td>1965</td>
<td>male</td>
<td>02141</td>
<td>short of breath</td>
</tr>
<tr>
<td>black</td>
<td>1965</td>
<td>male</td>
<td>02141</td>
<td>chest pain</td>
</tr>
<tr>
<td>black</td>
<td>1965</td>
<td>female</td>
<td>02138</td>
<td>painful eye</td>
</tr>
<tr>
<td>black</td>
<td>1965</td>
<td>female</td>
<td>02138</td>
<td>wheezing</td>
</tr>
<tr>
<td>black</td>
<td>1964</td>
<td>female</td>
<td>02138</td>
<td>obesity</td>
</tr>
<tr>
<td>black</td>
<td>1964</td>
<td>female</td>
<td>02138</td>
<td>chest pain</td>
</tr>
<tr>
<td>white</td>
<td>1960-69</td>
<td>male</td>
<td>02138</td>
<td>short of breath</td>
</tr>
<tr>
<td>white</td>
<td>1960-69</td>
<td>human</td>
<td>02139</td>
<td>hypertension</td>
</tr>
<tr>
<td>white</td>
<td>1960-69</td>
<td>human</td>
<td>02139</td>
<td>obesity</td>
</tr>
<tr>
<td>white</td>
<td>1960-69</td>
<td>human</td>
<td>02139</td>
<td>fever</td>
</tr>
<tr>
<td>white</td>
<td>1960-69</td>
<td>male</td>
<td>02138</td>
<td>vomiting</td>
</tr>
<tr>
<td>white</td>
<td>1960-69</td>
<td>male</td>
<td>02138</td>
<td>back pain</td>
</tr>
</tbody>
</table>

Important Dates

• November 18\textsuperscript{th} – Homework 7 due
  – Time consuming: \textit{start today}
• November 20 – Draft project paper due
• Dec 4\textsuperscript{th} – Poster Fair
• Dec 11\textsuperscript{th} – Final exam 1-4pm
• Dec 12\textsuperscript{th} – Final project due
Comments on Exam
Benefits of big data

• Scientific American “How Big Data Can Transform Society for the Better’ Oct 13

• Understanding the spread of Malaria in Kenya through mobile phone usage patterns (Wesolowski, Science 2012)

• Better public transportation through GPS tracking

• Better public health through search queries
Concerns about big data

- Incremental Effect
- Automated Decision-Making
- Predictive Analysis
- Lack of Access and Exclusion
- Analytics
- Chilling Effect

Omer Tene and Jules Polonetsky,
Solutions to the concerns?
REPORT TO THE PRESIDENT
BIG DATA AND PRIVACY:
A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President
President’s Council of Advisors on
Science and Technology

May 2014
PCAST = President’s advisory committee on science and tech.

- PCAST, staffed by OSTP
- Since FDR in 1933, many US Presidents have had a science and technology advisory board
- PCAST named and chartered by George H. W. Bush in 1990
- Topics: Climate, MOOCs, agriculture, nuclear security, big data

Seed Magazine, Recasting PCAST by Robert Koenig
President's Advisory Committee Straddles Worlds Of Politics, Science
Politics, Science (The Scientist, Vol:10, #6, p. 3 & 5, March 18, 1996)
20 current PCAST members

- Three Nobel laureates,
- Four MacArthur Prize fellows,
- Two university presidents
- 4 members of the national scientific, engineering, or medical academies
- Google’s chief executive, Eric Schmidt,
- Microsoft’s research chief, Craig Mundie.
Decisions about and by PCAST

• Who is a member
• Number of members
• Access to president
• Funding available
• Private or public meetings?
  – The next public PCAST meeting will take place on November 14, 2014.
• Big studies or small reports
PCAST report on big data

REPORT TO THE PRESIDENT
BIG DATA AND PRIVACY:
A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President
President’s Council of Advisors on Science and Technology

May 2014
What’s new about big data

• The quantity and variety of data that are available to be processed.

• The scale of analysis, inferences, and conclusions

• Data fusion: “when data from different sources are brought into contact and new facts emerge”
~6000 mega Data centers

http://www.Google.com/about/datacenters
Data: “Born digital” versus “born analog”

• Born digital - data packets, such as intentionally typed text, clicks, GPS location
  – Overcollection of data is a red flag

• Born analog – pictures, video, health info, voice calls
  – Signal to noise = More information than is strictly needed
  – Overcollection of data may be a good design choice
Security Options for Big Data

• Technologies that allow encrypted databases to be queried
  – E.g. Zero-knowledge systems
  – difficult to extend to complex/unstructured systems

• Differential privacy
  – Trade-off between anonymity and accuracy

• Anonymization
  – “PCAST does not see it as being a useful basis for policy”

• Deletion
  – “The only viable assumption …is that data, once created, are permanent”
Other FIPPs that don’t seem viable

• Notice and Consent
• Collection Limitation
  – Except perhaps attention to collection practices
• A right to forget/ expiration
Specific options that might work

• Privacy profiles
  – E.g. Developed by ACLU or Consumer Reports
  – standard, machine-readable interfaces

• Focus on use
  – With metatags
PCAST Policy Recommendations

1. Focus more on use of data than collection and analysis

2. Policy should be on intended outcomes, not technology solutions
PCAST Policy Recommendations cont.

3. Strengthen U.S. research in privacy-related technologies

4. Encourage increased education and training opportunities concerning privacy protection
PCAST Policy Recommendations, cont.

5. US should take the lead through standards and procurement practices