



COMP20008 Elements of Data Processing

Semester 2 2018

Lecture 23: Social and Ethical Implications of Big Data Analytics



THE UNIVERSITY OF
MELBOURNE

Announcements

- Best presenter for assignment phase 3:
 - Simon Han
 - Congratulations. Excellent Work!
- Assignment phase 3 marks were released yesterday!
 - Phase3-A (5): Average 3.63 Median 4.00
 - Phase 3-B (10): Average 7.23 Median 7.33
- Final update of exam guide
 - Available on Friday
- SES
 - Reminder
 - we value your feedback about the subject



THE UNIVERSITY OF
MELBOURNE

Plan for today

- Big data analytics
 - Issues
 - Stakeholders
 - Processes and ethical implications
 - 10 simple rules for responsible big data research
- Acknowledgements
 - Slides adapted from materials developed by Ida Asadi Someh



THE UNIVERSITY OF
MELBOURNE

Warm up

- Google Duo
 - <https://www.youtube.com/watch?v=D5VN56jQMWM>
- What are the concerns people have?

	Top 3 silicon valley (2014)	Top 3 auto (car) makers (1990)
Revenue	\$247 billion	\$250 billion
Number of employees	137000	1.2 million employee
Market capitalisation	\$1.09 trillion	\$36 billion

(Zuboff 2015)

- The ability to **collect, store, and process** increasingly large and complex data sets from a variety of sources, into competitive advantage (LaValle and Lesser 2013)
- Big data management capabilities
 - Volume, Variety and Velocity (3Vs) + Veracity (4Vs)
- Algorithms to process big data
 - Advanced statistical and computational techniques to process large, unstructured and fast data

Is this a sufficient definition?

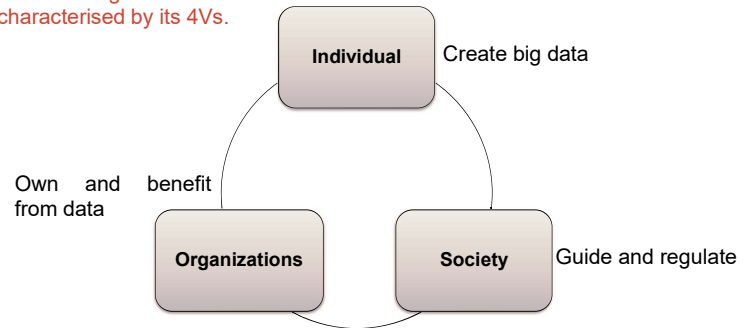
- Target exposing a teen girl's pregnancy
 - *Father: "My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"*
 - <https://www.businessinsider.com.au/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2>
 - https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp
- Facebook's 2012 secret mood experiment
 - how people react to an emotional contagion process
 - filtered users' news feeds – the flow of comments, videos, pictures and web links posted by other people
 - 689,000 users affected
 - <https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds>
- Cambridge Analytica (we have discussed previously ..)

- Positive consequences
 - Tracking criminals, higher product margins, new business models, improved healthcare and ...
- Negative consequences:
 - Misuse of personal information, breaching privacy, profiling of individuals, discrimination and ...
- Where is the boundary? What is the balance?
 - There is no agreement on what is **ethical** and what is not!

(Markus and Topi 2015; Newell and Marabelli 2015)

- Technological view (3Vs) does not help to understand unethical use
 - Technology is neutral in nature
 - It does not consider the underlying processes that are enabled
 - It does not consider the stakeholders that are involved and influenced
 - 3Vs do not consider either **people or process**
 - A new social phenomenon, a new market economy
- We need further **(non technical)** perspectives on big data analytics
 - It is not OK to exclude social from the definition

The exchange of data is characterised by its 4Vs.

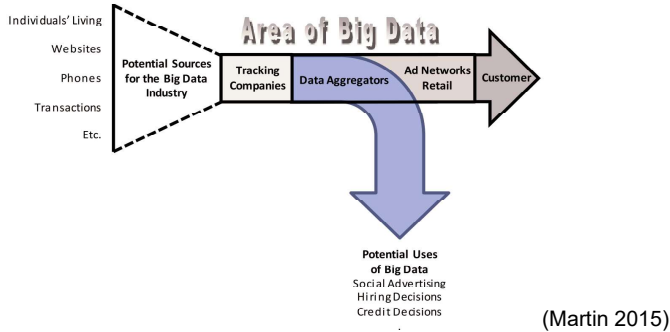


Unequal exchanges between stakeholders

- Interactions among stakeholders
- Data is **contributed, collected, extracted, exchanged, sold, shared**, and processed for the purpose of predicting and modifying human behaviour in the production of economic or social value.
- BDA involves several processes, discussed next

- Data *extraction*, not data *collection*
 - Google Streetview ("single greatest breach in the history of privacy")
 - <https://www.theguardian.com/technology/2010/may/15/google-admits-storing-private-data>
- Our everydayness quantified
- Incursions into legally and socially undefended territory
- Google has the largest **unpaid number of employees**
 - "You're not the customer, you're the product .."

- Sell personal data until it turns into waste
- Big data as a new industry (secondary markets)



- <http://intelligence.towerdata.com/>
- <http://www.iriworldwide.com/en-us/>
- <https://www.intelius.com/>

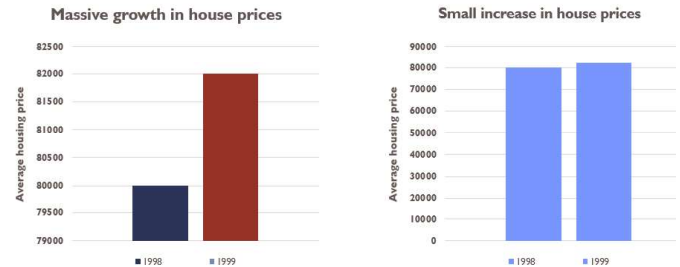
- Big Data Quality (Veracity)
 - Data accuracy for aggregated data
 - What's the quality criteria for a social media post?
 - Completeness of our digital identity
 - Mosaic effect
 - "When is an ambulance not an ambulance?"
 - Game: <http://celebrityguesswho.com/#2>
 - Meaning dependent on the context
 - Is how I act on social media a true representation of who I am?



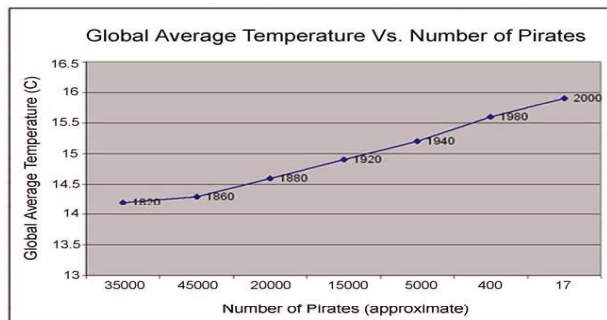
- Data Analysis
 - Predictions based on the past
 - How can I redefine myself?
 - In what context is it legitimate to make a prediction about someone?
 - Predictions often based on correlations (not causations)
 - What about outliers? (what if I don't fit into a predefined category??)
- Data Visualization
 - Decision making and presentation biases



- <https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts>



- Do you want to stop global warming? **BECOME A PIRATE!**



- Pervasive monitoring now possible using sensors, Internet of Things technology
 - Everyone is observed, organisations make money of observing others, collect data, sell data, make offers, induce dependence
 - What happens to social trust?
 - Surveillance is the precise opposite of the trust-based relationships
 - Free market economy versus **Surveillance Economy**
 - Who I am, what I am, what I like to do, where I like to go, who I know, ...
 - <http://theconversation.com/someones-looking-at-you-welcome-to-the-surveillance-economy-16357>
- (Zuboff 2015 and Martin 2015)

- Where did experiments traditionally take place?
 - We have been seeing big companies running a large number of experiments on its **users investigating different aspects**
- Rewards and punishments

(Zuboff 2015)

- Interactions among stakeholders
- Data is **contributed, collected, extracted, exchanged, sold, shared**, and processed for the purpose of predicting and modifying human behaviour in the production of economic or social value.

- Factors to consider for
 - Individuals
 - Organisations
 - Society
- For reference – we won't go through them all

Data Control	The extent to which an individual is empowered to audit the access to, storage, exploitation and modification of data about the individual.
Awareness	The extent to which an individual is mindful in consenting as to what data is collected about them and how it is used.
Trust	The extent to which an individual can have confidence that the parties who have access to their data respect the individual's rights.
Privacy	The extent to which an individual is able to restrict the disclosure of their personal information. <i>"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place" ?</i>
Choice	The extent to which an individual is able to make choices without being unfairly discriminated against or constrained by the use of big data and analytics.
Anxiety	The extent of psychological discomfort engendered by the collection and use of personal information for big data analytics purposes.

Data Quality	The extent to which organisations ensure the accuracy, currency, completeness and validity of big data.
Data Sourcing	The extent to which organisations collect, buy, and aggregate data from multiple sources in a manner that respects the rights of individuals.
Data Sharing	The extent to which organisations share, sell or otherwise disclose data in a manner that respects the rights of individuals.
Decision Making	The extent to which big data analytics and resulting organisational decisions respect the rights of individuals.
Ethical Culture	The extent to which organisations have values, norms and shared beliefs that promote ethical big data analytics practices through education, training and other means.
Ethical Data Governance	The extent to which organisations articulate ethical standards, decision rights and responsibilities for sourcing, analysing and sharing big data.
Behaviour	The extent to which organisational actors behave consistently with their organisation's ethical culture and standards.
Reputation	The extent to which relevant stakeholders (e.g. customers) believe an organisation will manage and utilise data about them ethically.
Competitive Pressure	The extent to which organisations are subject to pressure to compete using big data analytics unethically.

Power Imbalance	The extent to which power in society is imbalanced by the use of big data analytics by a dominant group, organisation or government.
Coercion	The extent to which participation and functioning in society is dependent on contributing one's own data to a collection for analysis.
Social Awareness	The extent to which members of a society are aware of the role of big data analytics in directing and regulating behaviour in the society.
Surveillance	The extent to which the lives of individuals in a society are observed, monitored, measured and profiled.
Principles and Guidelines	The extent to which effective principles and guidelines exist to protect the rights of individuals impacted by big data analytics.
Authority	The extent to which an entity (e.g., government, professional association) acts to enforce, through sanctions or other means, the rights embodied in the established principles and guidelines for big data analytics.
Social Mindset	The extent to which society collectively feels anxious and oppressed (or oppositely assured and empowered) about the use of big data analytics.

- EU General Data Protection Regulation (enforced since May 2018)
- Aims to regulate and **protect data privacy** for all EU citizens.
 - Penalty 4% of annual global turnover of the organizations.
- The consent
 - should be clear, concise, not too long and intelligibly written—should attach **the reasons of data collection** and analyses.
 - individuals have the right to **withdraw the consent** with the same easiness that they have previously agreed with.
- Accessing individual's data
 - Individuals have the right to ask for a **copy of their personal data** together with information regarding the processing and purpose of data collection and analyses from a controller
 - Individuals have the **right of data portability**, which means that they can transfer their data from one controller to another.

<http://www.eugdpr.org/eugdpr.org.html>

- From Zook et al (Plos Comp Bio 2017)
 - <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399>

- Acknowledge that data are people and can do harm
 - All data are people until proven otherwise
 - Social media
 - Heart rates from Youtube videos
 - Ocean measurements that change property risk profiles

- Recognize that privacy is more than a binary value
 - Privacy is contextual and situational
 - Single Instagram photo versus entire history of social media posts
 - Privacy preferences differ across individuals and societies

- Guard against the reidentification of your data
 - Metadata associated with photos
 - Reverse image search – connect dating and professional profiles
 - Difficult to recognize the vulnerable points a-priori!
 - Battery usage on a phone – can reveal a person's location
 - Unintended consequence of 3rd party access to phone sensors
 - When datasets thought to be anonymized are combined with other variables, it may result in unexpected reidentification

- Practice ethical data sharing
 - Seeking consent from participants to share data

- Consider the strengths and limitations of your data; big does not automatically mean better
 - Document the provenance and evolution of your data. Do not overstate clarity; acknowledge **messiness and multiple meanings**.
 - is a Facebook post or an Instagram photo best interpreted as an approval/disapproval of a phenomenon, a simple observation, or an effort to improve status within a friend network?

- Debate the tough, ethical choices/issues
 - importance of debating the issues within groups of peers
 - Examples mentioned earlier
 - Facebook emotional contagion
 - Exposing teen girl's pregnancy
 - More recently, Google Duplex
 - <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>

- Develop a code of conduct for your organization, research community, or industry
 - Are we abiding by the terms of service or users' expectations?
 - Does the general public consider our research "creepy"?

- Design your data and systems for auditability
 - Plan for and welcome **audits of your big data practices**.
 - Systems of auditability clarify how different datasets (and the subsequent analysis) differ from each other, aiding understanding and creating better research.
 - "For example, many types of social media and other trace data are unstructured, and answers to even basic questions such as network links depend on the steps taken to collect and collate data."

- Engage with the broader consequences of data and analysis practices
 - Recognize that doing big data research has **societal-wide effects**

- Know when to break these rules
 - Natural disaster
 - Public health emergency
 - Hostile enemy
 - ...
- It may be important to temporarily put aside questions of individual privacy in order to serve a larger public good.

- We need to empower individuals
 - Educate individuals, raise social awareness
 - Provide data access and control (e.g. Google activity)
 - <http://www.abc.net.au/4corners/stories/2017/04/10/4649443.htm>
- Define and develop a culture of acceptable data use
 - Organizations should internalize the costs
 - Genuine consent from individuals
 - Be transparent and clearly communicate intent of data collection and analytics/ai
 - Adopt rules for responsible big data research
 - Provide data control to individuals