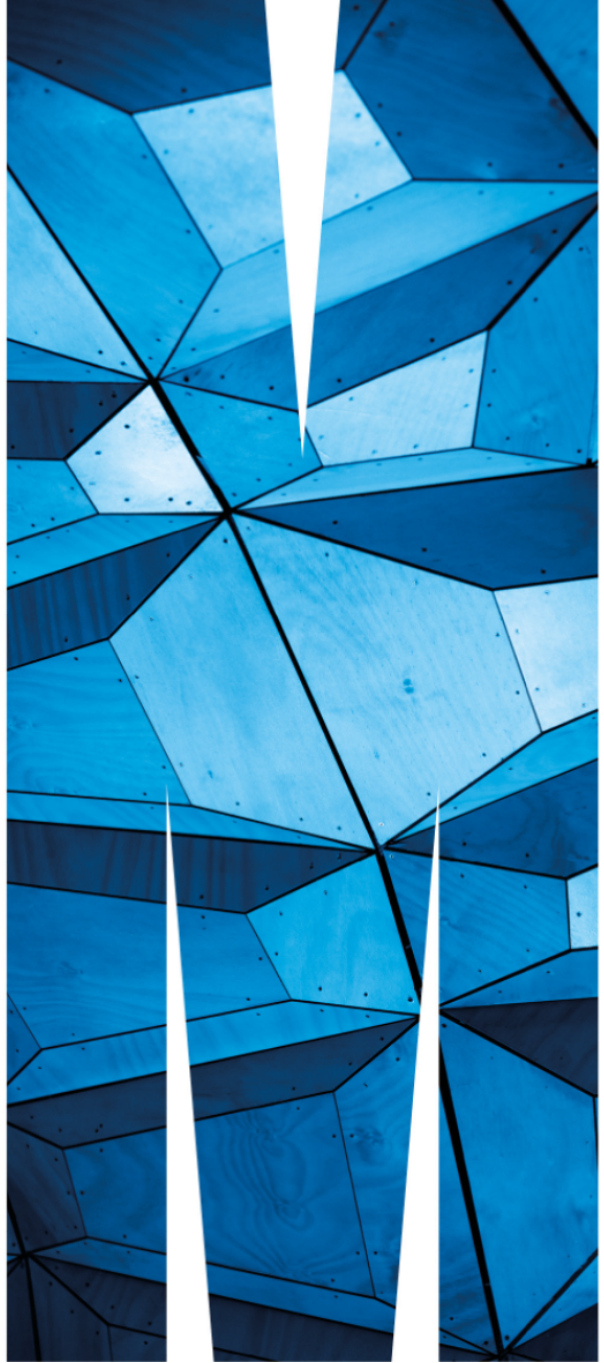
MONASH University

ETC5512: Wild Caught Data

Ethical use of data

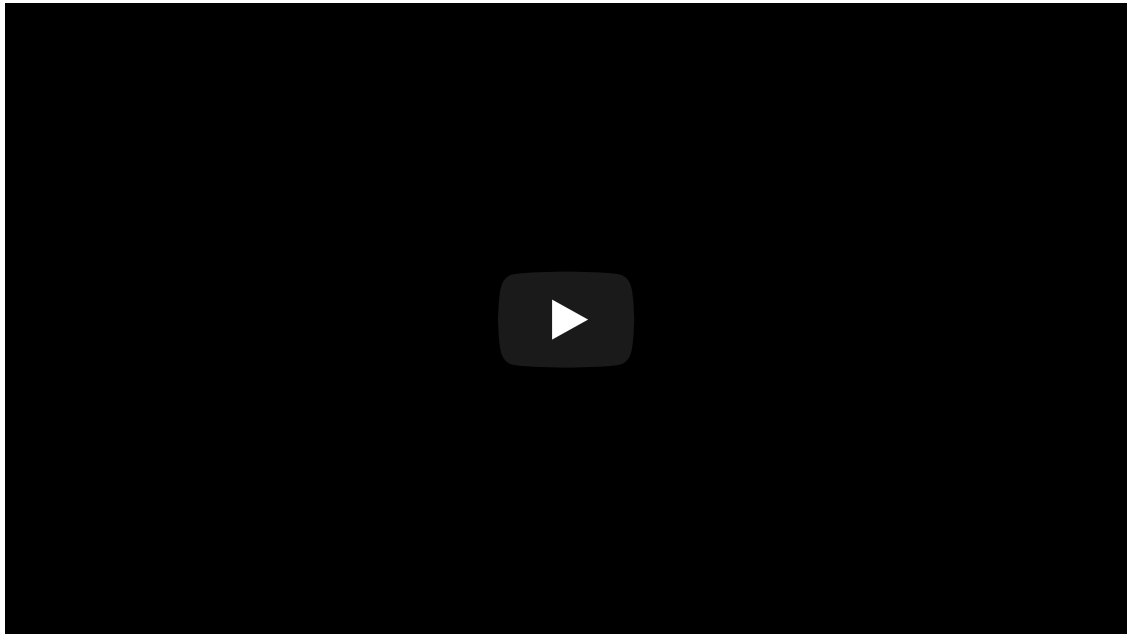


Lecturer: *Lauren Kennedy*

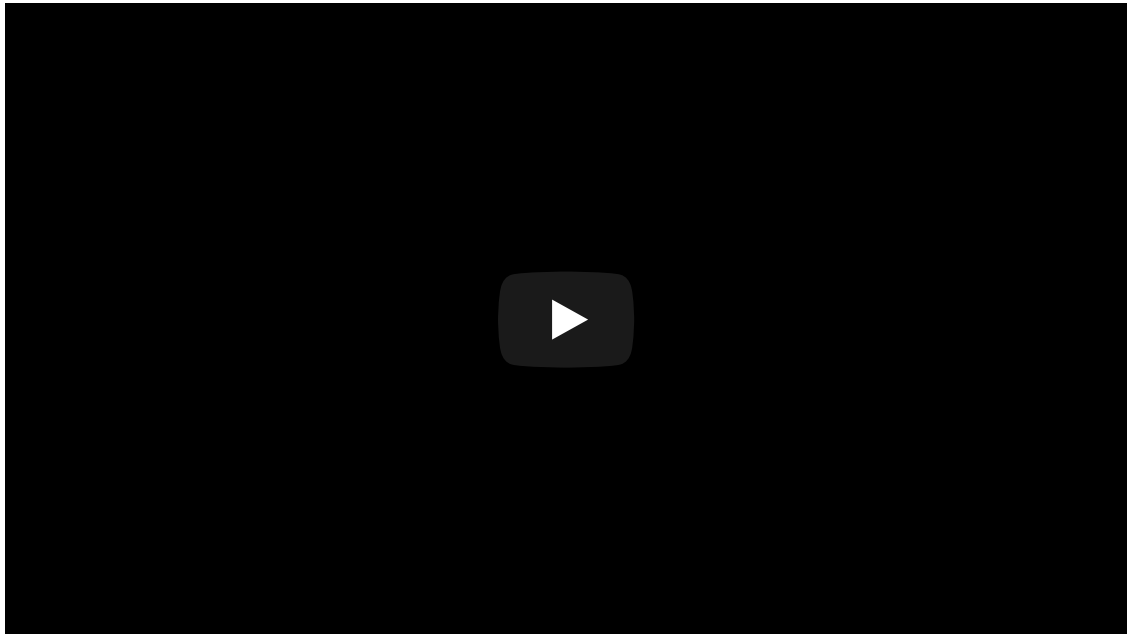
Department of Econometrics and Business Statistics

ETC5512.Clayton-x@monash.edu

Week 3



What would you do? How did you decide what to do?



Why consider ethical data practices?

- Data is increasingly accessible and available
  - Can use for understanding our world
  - Can use to make decisions (data driven decision making)
  - Can use to inform policy decisions
- Research ethics is often discussed in experimental settings
  - The fundamental ethical concerns do not change
- Data has an increasing impact on our lives
  - Data for good initiatives
  - Unethical uses of data
  - Open data increases the reach of data but risks privacy

Ethical Perspectives

- Virtue Based Ethics (Plato, Aristotle, Confucius, Mencius)
  - Act in a way to be a better person (culturally based)
  - The same aim can lead to conflicting actions
- Deontological Ethics(Immanuel Kant)
  - Universal ethical laws
  - Rules should apply to everyone
  - Preserve human autonomy and dignity-
- Consequential ethics and utilitarianism (John Stuart Mill)
  - Consider the outcomes rather than the intentions
  - Argue some unethical behavior okay if justified by the return

O'Keefe, K., & Brien, D. O. (2018). Ethical data and information management: concepts, tools and methods. Kogan Page Publishers, chapter 2

Ethical Perspectives

- Common good and justice based ethics
  - Action should contribute to some greater good
- Shareholder/stockholder theory (Milton Friedman)
  - Increase value to shareholders by maximizing profits
- Stakeholder theory
  - Organization responsible to stakeholders
  - Employees, stockholders, society, environment

O'Keefe, K., & Brien, D. O. (2018). Ethical data and information management: concepts, tools and methods. Kogan Page Publishers, chapter 2

Applying different ethical practices

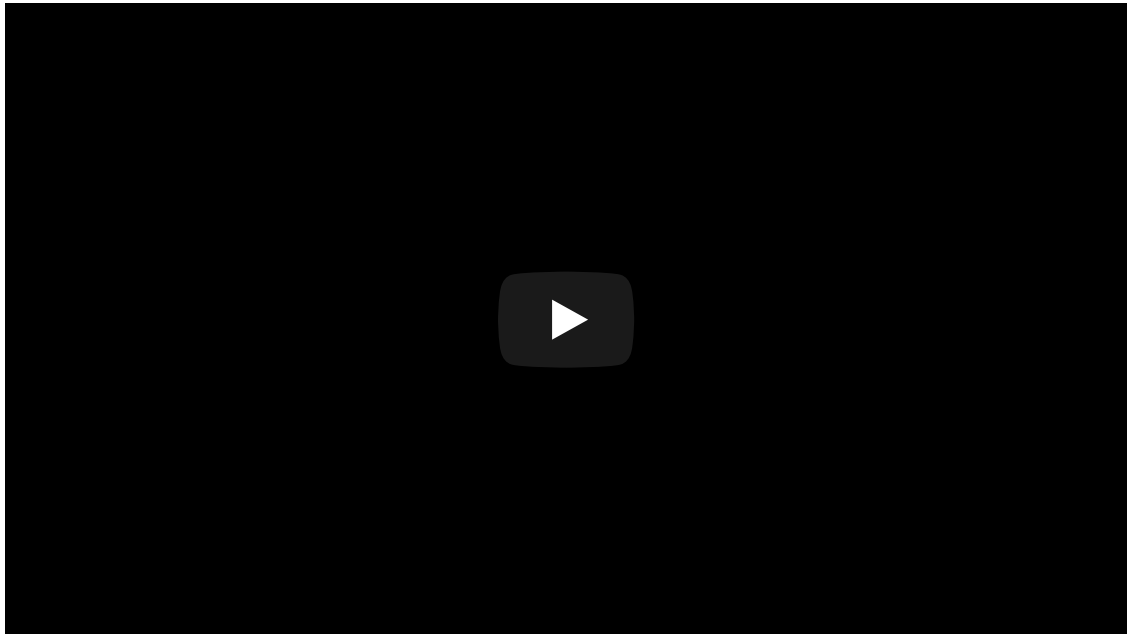
A large company surveys all of their current employees, measuring demographics and personality factors. They hope to identify key personality factors that correspond with a successful time with the company. Their hope is to use this data to identify which prospective employees they should hire.

Stakeholder theory

Shareholder theory

Deontological Ethics

How do we decide what is ethical?



Belmont report (1979)

- Identifies what constitutes as research
- Basic Ethical Principles
- Application to demonstrate the use of these principles

https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html#xbound

What constitutes as research?

> ***Research***
>
> *"an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge "*

> ***Practice***
>
> *"interventions that are designed solely to enhance the well-being of an individual patient or client and that have a reasonable expectation of success"*

https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html#xbound

Basic Ethical Principles

Autonomy is about being able to deliberate and make personal goals, and then act upon them

1. Respect for Persons
    - Acknowledge and respect autonomy
    - Protect those with diminished autonomy
2. Beneficence
    - Do not cause harm
    - Minimize possible harms and maximise possible benefits
3. Justice
    - Who ought to receive the benefits of research and bear its burdens?

https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html#xbound

Applications

1. Informed consent
    - Participants have a right to understand what is going to happen
    - Application of respect for persons
2. Assessment of Risks and Benefits
    - Are the risks justified?
    - Is the study well designed?
    - Are there alternative lower risk designs possible?
    - Are potential benefits and risks communicated to participants?
    - Application of beneficence

https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html#xbound

Applications

1. Selection of Subjects
    - Fairly offer research participation to all eligible
    - Some subjects preferred to as best able to bear risk (e.g., adults)
    - "injustice arises from social, racial, sexual and cultural biases institutionalized in society"
    - Careful not to overburden vulnerable subjects

Australia- National Statement on ethical conduct in research involving humans

Human research is conducted with or about people, or their data or tissue.

- Governs how to ethically do research involving humans in Australia
- Discusses Risk and Benefit, Consent
- Discusses ethical considerations for specific participants (e.g., children, pregnant women, Aboriginal and Torres Strait Islander Peoples)
- Discusses when and how ethical review can be conducted
- Discusses Risk and the level of review
  - Neglible risk - no foreseeable risk other than inconvenience
  - Low risk - only risk is discomfort
  - Potential risks must be balanced with potential benefit

https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#block-views-block-file-attachments-content-block-1

Who decides what is ethical?

- Human Research Ethics Committee (HREC) in Australia. Can be small differences in other countries so be mindful! (e.g., IRB in US)
- In Australia, all research with more than low risk needs to be reviewed by a HREC
- Low risk studies might be reviewed under different levels
- Studies with negligible risk might be exempt under certain conditions

Who is on a HREC? Sections 5.1.29 = 5.1.30

- Minimum 8
  - even numbers of men and women
  - 1/3 outside of the institution
- Chairperson
- Two lay people (do not do medical, scientifc, legal or academic work)
- At least one person with knowledge/experience in treatment/professional care (e.g., a nurse)
- One person who is in a pastoral care role (Aboriginal elder, minister)
- One lawyer
- Two people with current research experience

Why do you think the HREC requires such a diverse group members?

General Data Protection Laws

- Passed in 2018 to provide strong protection on the collections, use and management of data
- Rights of the individual
  - Information and access - you can access your data and see how it is processed
  - Right to request erasure of data
- Companies can now be held accountable
- In Australia, need to comply if you offer goods and services in EU, collect data from EU individuals or are established in the EU

Balancing Risk and Benefit

National Statement on ehtical conduct in research involving humans "risk is a potential for harm, discomfort or inconvenience (discussed below). It involves:

- the likelihood that a harm (or discomfort or inconvenience) will occur; and
- the severity of the harm, including its consequences."

Risks should be:

- Identified/Assessed
- Minimized
- Justified - does the potential benefit outweigh the potential risk?
- Monitored and managed if the research is approved

Your turn

A study investigates whether there has been an increased occurrence in loneliness during the COVID-19 lockdowns in Melbourne, Australia. The research method uses a survey administered during lockdown and non-lockdown periods to measure participant loneliness. What is one potential risk? How could this risk be minimized/managed/monitored? Are there benefits that justify this risk?

- **RISK:** Respondents reports extreme psychological stress
- **BENEFIT:** Understanding loneliness patterns can assist in providing preventitive and other mental health services
- **MITIGATION/MANAGEMENT:** Respondents are given a list of freely available mental health services to use if needed at the end of the study.

Consent

Belmont Report

- Information
- Comprehension
- Voluntary (without cohersion)

National Statement on ethical conduct in research involving humans "...consent should be a voluntary choice, and should be based on sufficient information and adequate understanding of both the proposed research and the implications of participation in it."

Your turn

You are working on a study involving high schoolers aged between 15 and 17. The study aims to investigate the impact of a learning intervention on scores in maths subjects. The students are told they will be randomly assigned to either an intervention or a control condition of maths exercises, and the researchers request access to their future scores as part of the study. Students will be compensated $100 for participation. What should be considered to obtain consent?

- Do teenagers or their parents consent?
- Is $100 a reasonable compensation or could it coerse students into participating
- Are alternatives to participating presented to students?
- Do participants sacrifice learning time to participate?

Justice, data sources and representation

- From the Belmont report we see that justice relates to the risks/benefits of the study at the same probability for all participants
- They also note the need to be careful not to exhaust vulnerable populations by overstudying.
- Big datasets often have challenges of representation - not everyone in the population has an equal chance of representation
- This can be one cause of biased predictions and unfair algorithms - the algorithms are less accurate for particular subsets of the population
- However, much of big data doesn't cost the participant to give - it's created through scraping etc, and the lack of representation reflects sociological and cultural power imbalances

Your turn

A large company surveys all of their current employees, measuring demographics and personality factors. They hope to identify key personality factors that correspond with a successful time with the company. Their hope is to use this data to identify which prospective employees they should hire.

- Are there any concerns with the data source?
- Using existing employee data reinforces any previous discriminatory hiring patterns

Counterfactual: What would have happened if we hired people who were different to those previously hired?

Focussing on statistical and data practices

- So far we've focused on ethical and data practice in research generally.
- The basic principles apply to ALL research, but many of the examples focus on experimental research
- How can we translate this principles to a more statistical/data focus?

Ethical guidelines for statisticians

1. Professional Integrity and Accountability
2. Integrity of data and method
3. Responsibilities to Science/Public/Funder/Client
4. Responsibilities to Research Subjects
5. Responsibilities to Research Team Colleagues
6. Responsibilities to Other Statisticians or Statistics Practitioners
7. Responsibilities Regarding Allegations of Misconduct

Alternative view

1. Prioritize open data and methods
2. Be clear about the information and assumptions that go into statistical methods
3. Respect for data
4. Publication of criticisms
5. Respect the limitations of statistics

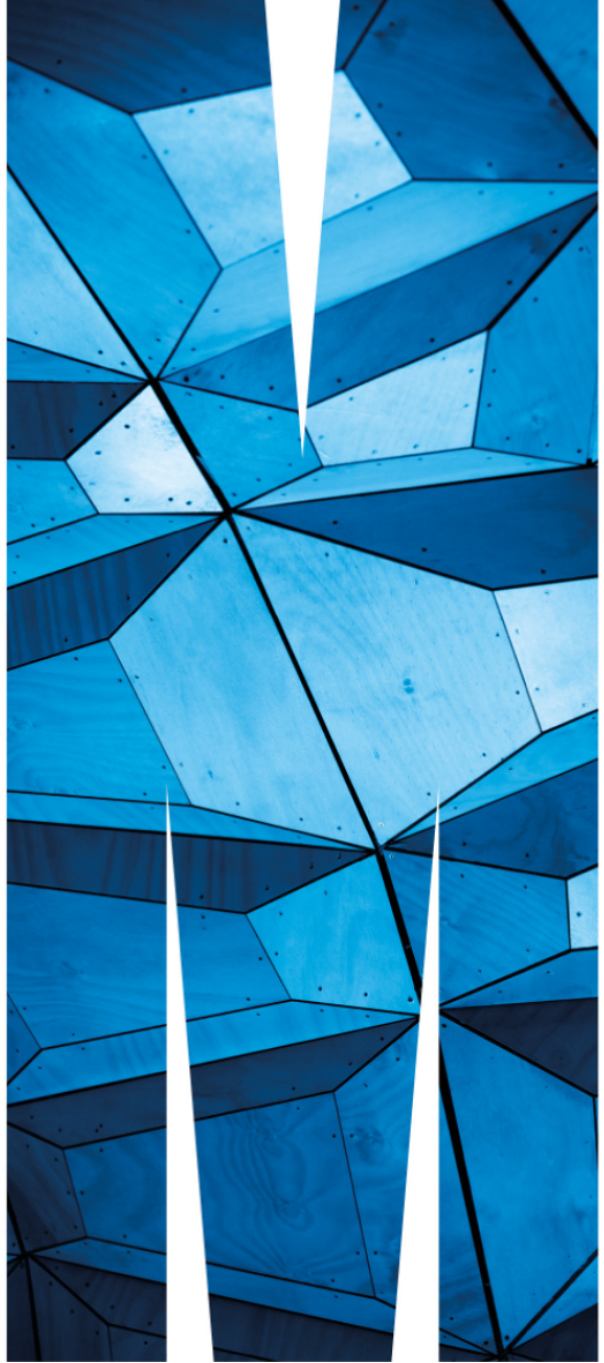
Yet another perspective

- Much of what we spent the first have of the lecture discussing surrounds data collection
- Identify other areas including:
  - Storage
  - Analysis
  - Modeling
  - Deployment
- These areas are less likely to be mentioned, but are equally important.

How to be an ethical data scientist

1. Firstly ensure that you are following the ethical frameworks of the company you are in (e.g., I work for Monash so I ensure that my research conforms to Australian standards through Monash procedures)
2. Ensure you understand the different areas where unethical thinking/practices can be introduced
3. Checklists like https://deon.drivendata.org/#default-checklist can be useful!
4. Keep learning and exploring different perspectives to you own

MONASH University



Lecturer: *Lauren Kennedy*

Department of Econometrics and Business Statistics

ETC5512.Clayton-x@monash.edu

Week 3