

# DATA ETHICS

Data Analysis for Journalism and Political Communication  
(Spring 2024)

Prof. Bell

# BOUIE, “QUANTIFYING THE PAIN OF SLAVERY”

Trans-Atlantic Slave Trade - Database								
Year range ▾ Ship, nation, owner ▾ Itinerary ▾ Enslaved people ▾ Dates ▾ Captain and crew ▾ Outcome ▾ Source ▾								
Results Summary statistics Tables Data visualization Timeline Maps Timelapse								
Showing 1 to 15 of 36,079 entries								
Configure columns ▾ Show 15 rows ▾ Download ▾								
Year arrived with slaves (IMP)	Voyage ID	Vessel name	Place where vessel's voyage began (IMP)	Principal place where captives were purchased (IMP)	Principal place where captives were landed (IMP)	Linked voyages	Captives arrived at 1st port	Captain's name
1817	1	Pastora de Lima	Rio de Janeiro	Mozambique	Bahia, place unspecified		290	Dias, Manoel José
1817	2	Tibério	Bahia, place unspecified	Mozambique	Bahia, place unspecified		223	Mata, José Maria da
1817	3	Paquete Real	Bahia, place unspecified	Cabinda	Bahia, place unspecified		350	Ferreira, José dos Santos
1817	4	Bom Caminho	Bahia, place unspecified	Quilimane	Bahia, place unspecified		342	Dias, Domingos Francisco
1817	5	Benigretta	Bahia, place unspecified	Cabinda	Bahia, place unspecified		516	
1817	6	Voador	Bahia, place unspecified	Mozambique	Bahia, place unspecified		515	
1817	7	Formiga	Bahia, place unspecified	Malembo	Bahia, place unspecified		204	Viana, Isidoro Antônio

Go to [SlaveVoyages.org](https://SlaveVoyages.org) and spend a few minutes exploring the data. Think about the implications of collecting and sharing this data. What are some of the benefits of sharing this data? What are some of the risks? Do the benefits outweigh the harms?

# TUSKEGEE SYPHILIS STUDY



# THE BELMONT REPORT (1978)

- Summarizes ethical principles and guidelines for conducting research with human subjects

# THE BELMONT REPORT (1978)

- Summarizes ethical principles and guidelines for conducting research with human subjects
- Codified into law as the “Common Rule”, which covers all federally-funded research

# THE BELMONT REPORT (1978)

## Respect for Persons

Individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled to protection.

# THE BELMONT REPORT (1978)

## Respect for Persons

Individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled to protection.

## Beneficence

(1) Do not harm and (2) maximize possible benefits and minimize possible harms.

# THE BELMONT REPORT (1978)

## Respect for Persons

Individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled to protection.

## Beneficence

(1) Do not harm and (2) maximize possible benefits and minimize possible harms.

## Justice

Groups who bear the burden of research should also be the beneficiaries of that research.



# OTHER CONSIDERATIONS

- Informed Consent
  - ▶ Subjects must affirmatively agree to participate in research

# OTHER CONSIDERATIONS

- Informed Consent
  - ▶ Subjects must affirmatively agree to participate in research
  - ▶ Participants must have all information necessary to make an informed decision

# OTHER CONSIDERATIONS

- Informed Consent

- ▶ Subjects must affirmatively agree to participate in research
- ▶ Participants must have all information necessary to make an informed decision
- ▶ What about studies that require deception? Studies that compensate participants?

# OTHER CONSIDERATIONS

- Informed Consent
  - ▶ Subjects must affirmatively agree to participate in research
  - ▶ Participants must have all information necessary to make an informed decision
  - ▶ What about studies that require deception? Studies that compensate participants?
- Weighing benefits and risks of research
  - ▶ Is there a way to gain the information sought in the study that minimizes risk?

# OTHER CONSIDERATIONS

- Informed Consent
  - ▶ Subjects must affirmatively agree to participate in research
  - ▶ Participants must have all information necessary to make an informed decision
  - ▶ What about studies that require deception? Studies that compensate participants?
- Weighing benefits and risks of research
  - ▶ Is there a way to gain the information sought in the study that minimizes risk?
  - ▶ If risks cannot be reduced, how great is the potential benefit to society and to the subject?

# OTHER CONSIDERATIONS

- Informed Consent

- ▶ Subjects must affirmatively agree to participate in research
- ▶ Participants must have all information necessary to make an informed decision
- ▶ What about studies that require deception? Studies that compensate participants?

- Weighing benefits and risks of research

- ▶ Is there a way to gain the information sought in the study that minimizes risk?
- ▶ If risks cannot be reduced, how great is the potential benefit to society and to the subject?
- ▶ Examples: Stanford Prison Experiment, Milgram Experiment

# OTHER CONSIDERATIONS

- Informed Consent

- ▶ Subjects must affirmatively agree to participate in research
- ▶ Participants must have all information necessary to make an informed decision
- ▶ What about studies that require deception? Studies that compensate participants?

- Weighing benefits and risks of research

- ▶ Is there a way to gain the information sought in the study that minimizes risk?
- ▶ If risks cannot be reduced, how great is the potential benefit to society and to the subject?
- ▶ Examples: Stanford Prison Experiment, Milgram Experiment  
vs. randomized drug trials, “challenge” trials in medicine

# OTHER CONSIDERATIONS

- Informed Consent

- ▶ Subjects must affirmatively agree to participate in research
- ▶ Participants must have all information necessary to make an informed decision
- ▶ What about studies that require deception? Studies that compensate participants?

- Weighing benefits and risks of research

- ▶ Is there a way to gain the information sought in the study that minimizes risk?
- ▶ If risks cannot be reduced, how great is the potential benefit to society and to the subject?
- ▶ Examples: Stanford Prison Experiment, Milgram Experiment  
vs. randomized drug trials, “challenge” trials in medicine

- Selection of subjects

- ▶ Do not select favored subjects for highly beneficial research and unfavored subjects for risky research



# CASE STUDIES

- 1 Home DNA Testing
- 2 Crisis Text Line
- 3 Diversity in Faces (DiF) dataset

# CASE STUDIES

- 1 What are the relevant ethical principles and practices in this case?
- 2 In what ways/why are there concerns about a violation of ethical principles in this case?
- 3 What are some ways that data could have been used more ethically in this case?

# CASE STUDIES

- 1 Home DNA Testing
- 2 Crisis Text Line
- 3 Diversity in Faces (DiF) dataset

# DE-IDENTIFICATION

- Researchers often promise ***anonymity*** or ***confidentiality*** to participants in order to reduce the risks of participating in research

# DE-IDENTIFICATION

- Researchers often promise ***anonymity*** or ***confidentiality*** to participants in order to reduce the risks of participating in research
- There are two types of identifiers that must be removed before sharing this data:

# DE-IDENTIFICATION

- Researchers often promise ***anonymity*** or ***confidentiality*** to participants in order to reduce the risks of participating in research
- There are two types of identifiers that must be removed before sharing this data:
  - ① **Direct identifiers:** information that would be sufficient on its own to disclose an identity, such as names, addresses, and phone numbers.

# DE-IDENTIFICATION

- Researchers often promise ***anonymity*** or ***confidentiality*** to participants in order to reduce the risks of participating in research
- There are two types of identifiers that must be removed before sharing this data:
  - 1 **Direct identifiers:** information that would be sufficient on its own to disclose an identity, such as names, addresses, and phone numbers.
  - 2 **Indirect identifiers:** information that *in combination* would be sufficient to disclose an identity

# DE-IDENTIFICATION

- Researchers often promise **anonymity** or **confidentiality** to participants in order to reduce the risks of participating in research
- There are two types of identifiers that must be removed before sharing this data:
  - ① **Direct identifiers:** information that would be sufficient on its own to disclose an identity, such as names, addresses, and phone numbers.
  - ② **Indirect identifiers:** information that *in combination* would be sufficient to disclose an identity
- Examples: credit card transactions, Netflix movie ratings, tennis match-fixing



# DE-IDENTIFICATION

- Researchers often promise **anonymity** or **confidentiality** to participants in order to reduce the risks of participating in research
- There are two types of identifiers that must be removed before sharing this data:
  - ① **Direct identifiers:** information that would be sufficient on its own to disclose an identity, such as names, addresses, and phone numbers.
  - ② **Indirect identifiers:** information that *in combination* would be sufficient to disclose an identity
- Examples: credit card transactions, Netflix movie ratings, tennis match-fixing
- De-identification is the process of removing direct and indirect identifiers

# DE-IDENTIFICATION

## 1 Remove direct identifiers

Antenatal Card: B2-2296 Report Date: 2006-07

Patient Name	Patient Id	Age	Birth Date	Address	Husband	Clinic Name
Patient2332		20-25		Region 3	married	Inst_OC3

Previous Pregnancies

Month/Year	Duration	Health During Pregnancy	Mode of Delivery	Type of Labor	Birth Weight	Outcome
1 / 19	Full Term	health_type 3	Spontaneous Vaginal	Spontaneous	Infant 1: 2.5	Alive
1 / 19	Full Term	health_type 4	Spontaneous Vaginal	Spontaneous	Infant 1: 3.0	Alive
12 / 1	Full Term	health_type 2	Spontaneous Vaginal	Spontaneous	Infant 1: 2.3	Alive

Lab Tests

Date Request	Lab Type	Date Results	Results	Site	Clinician
2006-0	Hb - 1st screen	2006-11-13	12.6	InstOC3	Dr_16
24rd week					

Drug Interventions

Date Given	Folic	Iron
2006-0	X	X
18th week		

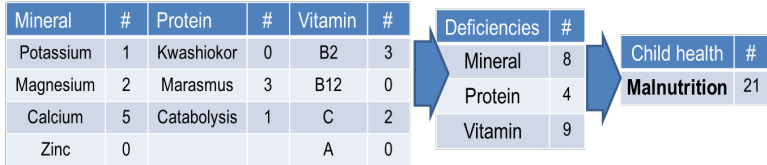
# DE-IDENTIFICATION

- 1 Remove direct identifiers
- 2 Aggregate or reduce the precision of a variable
  - ▶ Generalize the meaning of categories

SubjID	Region	RegGen
2253	21239	212
2254	21238	212
2255	21135	211
2256	06058	060

# DE-IDENTIFICATION

- 1 Remove direct identifiers
- 2 Aggregate or reduce the precision of a variable
  - ▶ Generalize the meaning of categories
  - ▶ Collapse categories



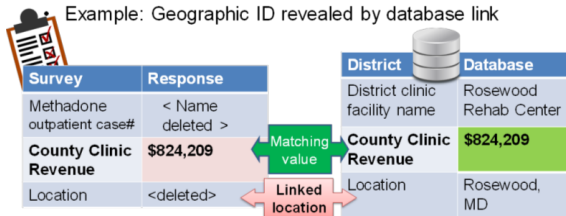
# DE-IDENTIFICATION

- 1 Remove direct identifiers
- 2 Aggregate or reduce the precision of a variable
  - ▶ Generalize the meaning of categories
  - ▶ Collapse categories
  - ▶ Restrict the upper or lower ranges

Age	Actual Wealth	Top-coded Wealth
24	24,778	24,778
31	26,750	26,750
42	26,780	26,780
64	35,469	<b>30000+</b>
27	43,695	<b>30000+</b>

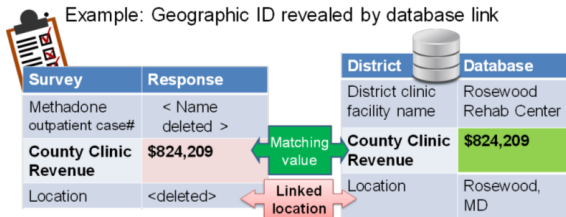
# DE-IDENTIFICATION

- 1 Remove direct identifiers
- 2 Aggregate or reduce the precision of a variable
  - ▶ Generalize the meaning of categories
  - ▶ Collapse categories
  - ▶ Restrict the upper or lower ranges
- 3 Anonymize keys that link to other datasets



# DE-IDENTIFICATION

- 1 Remove direct identifiers
- 2 Aggregate or reduce the precision of a variable
  - ▶ Generalize the meaning of categories
  - ▶ Collapse categories
  - ▶ Restrict the upper or lower ranges
- 3 Anonymize keys that link to other datasets
- 4 Maintain a master log of all replacements, aggregations, or removals and keep it in a secure location separate from the de-identified data files



# DE-IDENTIFICATION EXERCISE