Worksheet 15 - data science ethics

Wednesday, November 3, 2026

DS 002R - Jo Hardin

Name: _____

Names of people you worked with: _____

For next week's quiz: what is the hardest part of functions and `map()` for the simulations?

Task: 12 principles given by the Data Values and Principles manifesto. For each of the 4 examples, note which principles are addressed or violated (a few principles per example).

1. Use data to improve life for our users, customers, organizations, and communities.
2. Create reproducible and extensible work.
3. Build teams with diverse ideas, backgrounds, and strengths.
4. Prioritize the continuous collection and availability of discussions and metadata.
5. Clearly identify the questions and objectives that drive each project and use to guide both planning and refinement.
6. Be open to changing our methods and conclusions in response to new knowledge.
7. Recognize and mitigate bias in ourselves and in the data we use.
8. Present our work in ways that empower others to make better-informed decisions.
9. Consider carefully the ethical implications of choices we make when using data, and the impacts of our work on individuals and society.
10. Respect and invite fair criticism while promoting the identification and open discussion of errors, risks, and unintended consequences of our work.
11. Protect the privacy and security of individuals represented in our data.
12. Help others to understand the most useful and appropriate applications of data to solve real-world problems.

a. Predicting sexuality (neural network)
b. Predicting race (**wru** R package)
c. Data scraping (OKCupid)
d. Errors in spreadsheet (Excel)

Solution:

Predicting sexuality

- principle 1: Does the prediction of sexual orientation based on facial recognition improve life for communities?
- principle 9: As noted in the abstract, the researchers did consider the ethical implications of their work,
- principle 11: but did they protect the privacy and security of the individuals presented in their data?
- principle 7: The exclusion of non-white faces from the study casts doubt on whether the authors recognized their own biases.

Predicting race

- using this software to discriminate against historically marginalized people would violate some combination of principles 3, 7, and 9.
- is it ethical to use this software to try and help underrepresented groups if those same principles are not violated?
- The authors of the **wru** package admirably met principle 2, but they may not have fully adhered to principle 9.

Data scraping

stakeholders: OKCupid users (principles 1, 9, 11)

- research involving humans requires that the human not be exposed to any risk for which consent has not been explicitly given. The OkCupid members did not provide such consent.
- the data contain information that makes it possible to identify individual humans, there is a realistic risk of the release of potentially embarrassing information, or worse, information that jeopardizes the physical safety of certain users.

stakeholders: OKCupid itself (principle 4)

- were the terms of agreement (set by OKCupid) violated?

Errors in spreadsheet

- principle 10 was met: Reinhart and Rogoff shared their work when it was challenged
- principle 2 is violated: Microsoft Excel, the tool used by Reinhart and Rogoff, is an unfortunate choice because it mixes the data with the analysis.