Existing Dataset (used to train/test)

Included features (feat)

		Feat 1	Feat 2	Feat 3	Feat i…	Known outcome	Uncollected key features	Not deemed important by power brokers	Hard to or non-quantifiable aspects	Unable to access
		e.g., sex	e.g., marital status	e.g., years in school	e.g., Additional features of interest	e.g., **clinic:** miscarriage risk **college:** dropout likelihood	e.g., Unknown explanatory features	e.g., minoritized community driven interests	e.g., emotional wellbeing, sense of agency, etc.	e.g., proprietary data or legal concerns
Included cases	Person 1	1	1	12	…	0.95				
	Person 2	0	null	null	…	0.1				
	Person 3	0	2	8	…	0				
	Person i…	X_1	NULL	X_2	…	$\hat{y}$				

Outside of dataset — **Absent features**

Outside of dataset — Absent cases

| Uncounted Person 1 |
| Uncounted Person 2 |
| Uncounted Person 3 |
| Uncounted Person i… |

Figure annotations:
- 1.
- 2.
- 3.
- 4.

Side questions:
1. Who is in the data?
2. How are they in the data?
3. What are the politics of the collected data?
4. What are the politics of how people show up in data?

Springer, Emily. (2022). Visualizing In- and Exclusion in AI Datasets: A Sociotechnical Framework.

TechnoSocio
ADVISORY

Integrating the social into AI system models

Springer. Integrating the social into OECD's new AI System model. 2024.