Data Science Ethics

Deep Dive

Ezinne Nwankwo

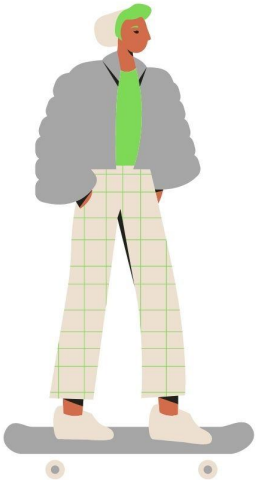
Where does bias come from?

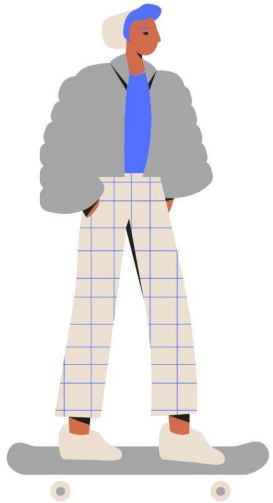
Data Collection vs Data Analysis

Where does bias come from?

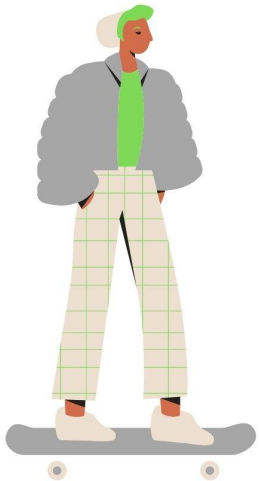
Data Collection vs Data Analysis

Redlining Analogy



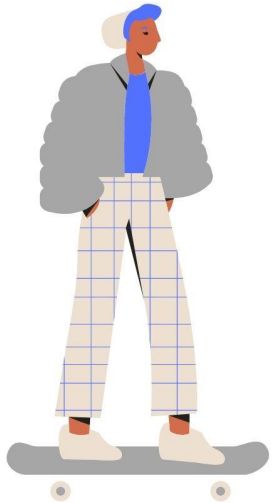


Redlining Analogy

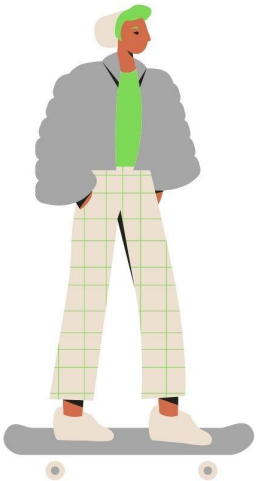


Individual 1

- Credit score: 650
- Income: $70,000/yr
- Hair color: green
- Shirt color: green

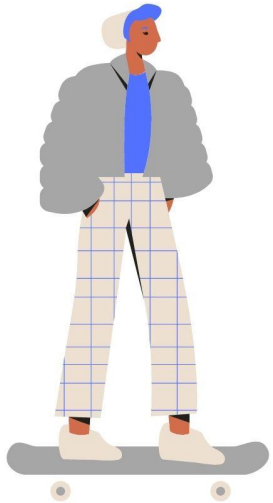


Redlining Analogy

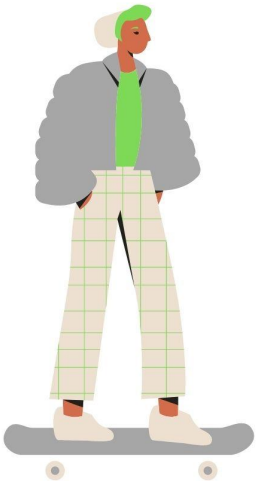


Individual 2

- Credit score: 650
- Income: $70,000/yr
- Hair color: blue
- Shirt color: blue

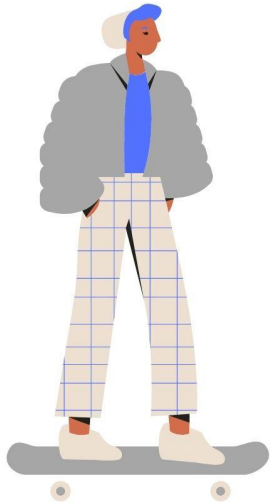


Redlining Analogy



Individual 2

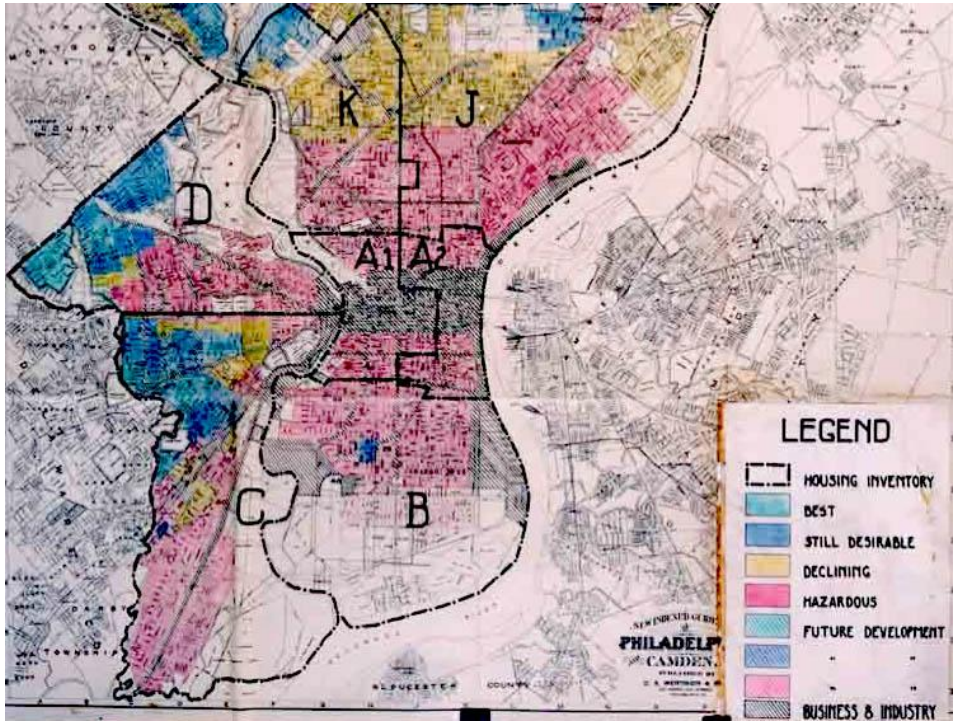
- Credit score: 650
- Income: $70,000/yr
- **Race**: blue
- **Zip code**: blue



Redlining Analogy

- Substituting hair color and shirt color with race and zip code are examples of what some people call redundant encoding.

- Historically, the practice of arbitrarily denying or limiting financial services to specific neighborhoods, generally because its residents are people of color or poor is called redlining.



Moral of the story

1. Understand where your data comes and what the variables relate (or correlate) to.
2. Question the data being collected and what it points to.

Where does bias come from?

Data Collection vs Data Analysis

Effect on sub-populations

		Graduate school admissions	
		admitted	denied
Gender	F	1512	2809
	M	3715	4727

Positive Outcomes

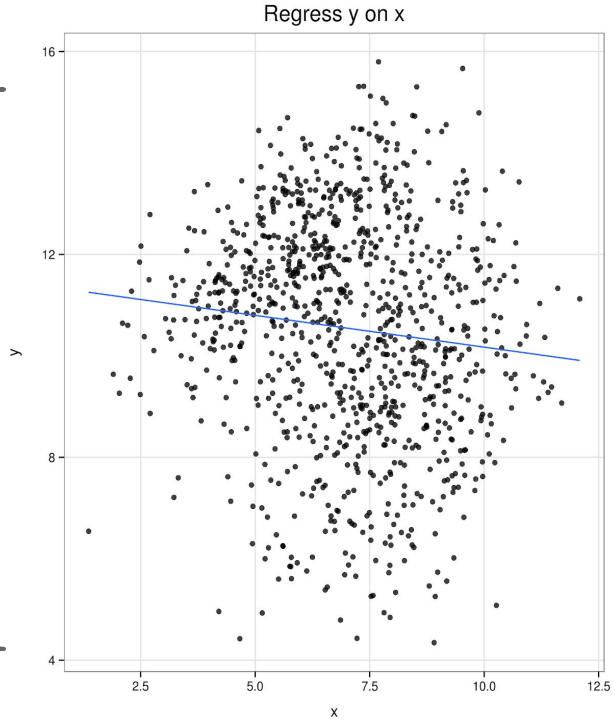
35 % of women were admitted

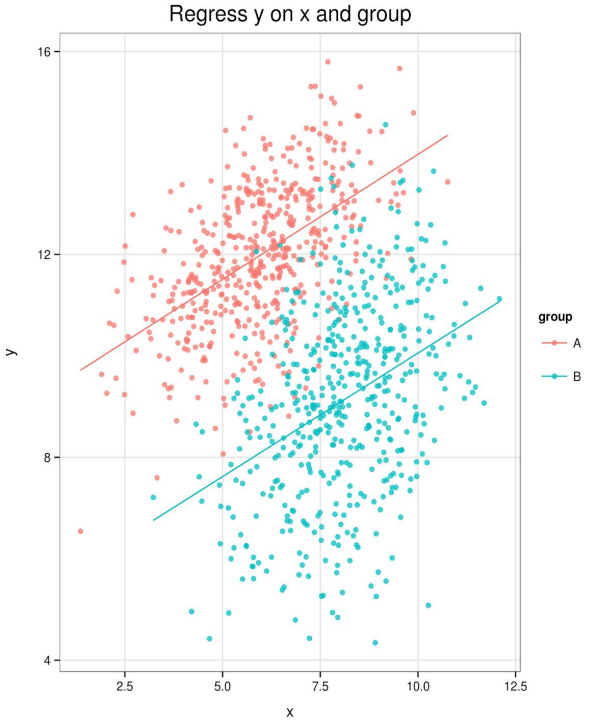
44 % of men were admitted

UC Berkeley 1973: Upon looking at admittance rates of different genders for different departments, the rates actually ended up being higher for women than men

Simpson's Paradox

Definition: When conclusions drawn from using aggregate data over an entire population is different (maybe even opposite) to those of underlying sub groups





Thank For Listening!

Sources

[1] Peter J Bickel, Eugene A Hammel, and J William O'Connell. 1975. Sex bias in graduate admissions: Data from Berkeley. Science 187, 4175 (1975), 398-404.

[2] Nazanin Alipourfard, Peter G Fennell, and Kristina Lerman. 2018. Can you Trust the Trend?: Discovering Simpson's Paradoxes in Social Data. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. ACM, 19-27

[3] FairML Textbook. https://fairmlbook.org/pdf/fairmlbook.pdf