

# **DATA ETHICS: CHOICES, AND VALUES**

**KATIE CREEL**

**CS 106A, FALL 2021**

# OUTLINE



Bias, Fairness and Distribution



Thick Evaluative Descriptions



Problem Formulation



Representation in Survey Data

# BIASBARS ASSIGNMENT ... SPECIAL PREVIEW 😊

QUALITY

5.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

DIFFICULTY

1.0

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



# **WHAT CAN WE LEARN FROM THIS DATA?**

**In order to know how to use a dataset  
appropriately, we need to examine it for patterns of  
bias.**

# What is in the Dataset?



A photograph of three spelunkers in a cave, silhouetted against a bright opening that reveals a lush, sunlit forest. The spelunkers are wearing headlamps and gear. The scene is dramatic, with the cave interior in deep shadow and the forest outside in bright light.

**You Have  
the Power  
to Find  
Out!**

**Values &  
Habits**

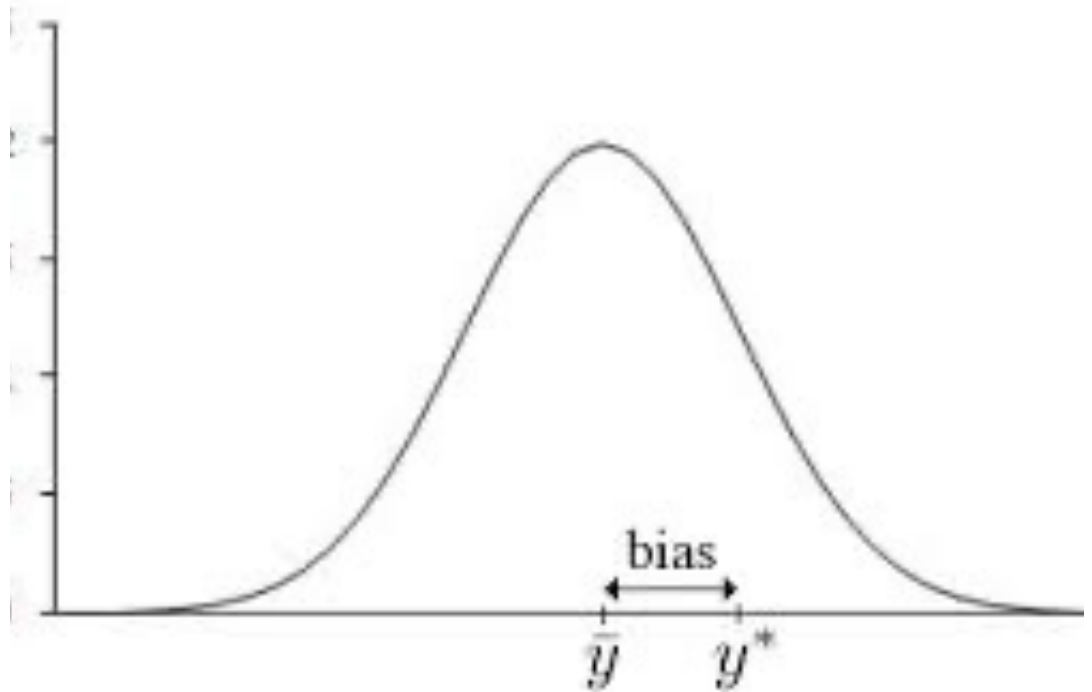
**Ethics  
Skills**

**CS  
Skills**

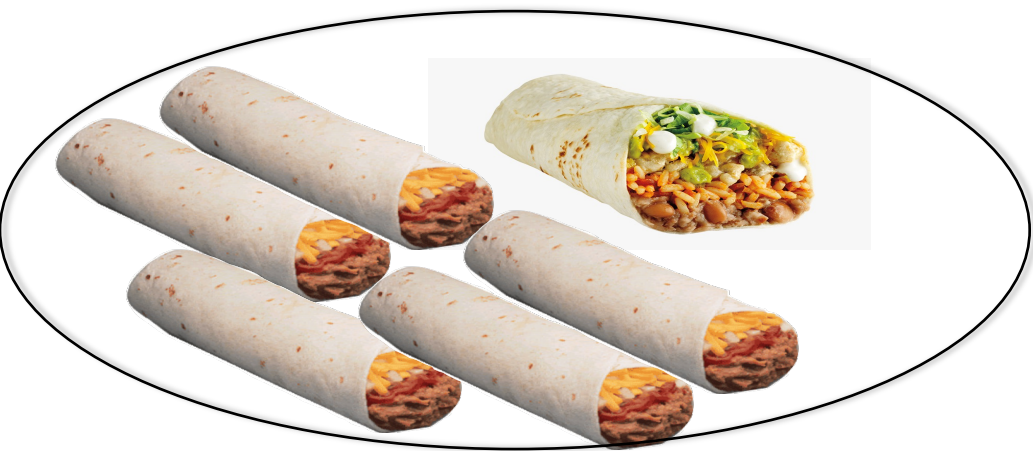
# WHAT IS BIAS?

Statistical bias is the difference between measured results and “true” value.

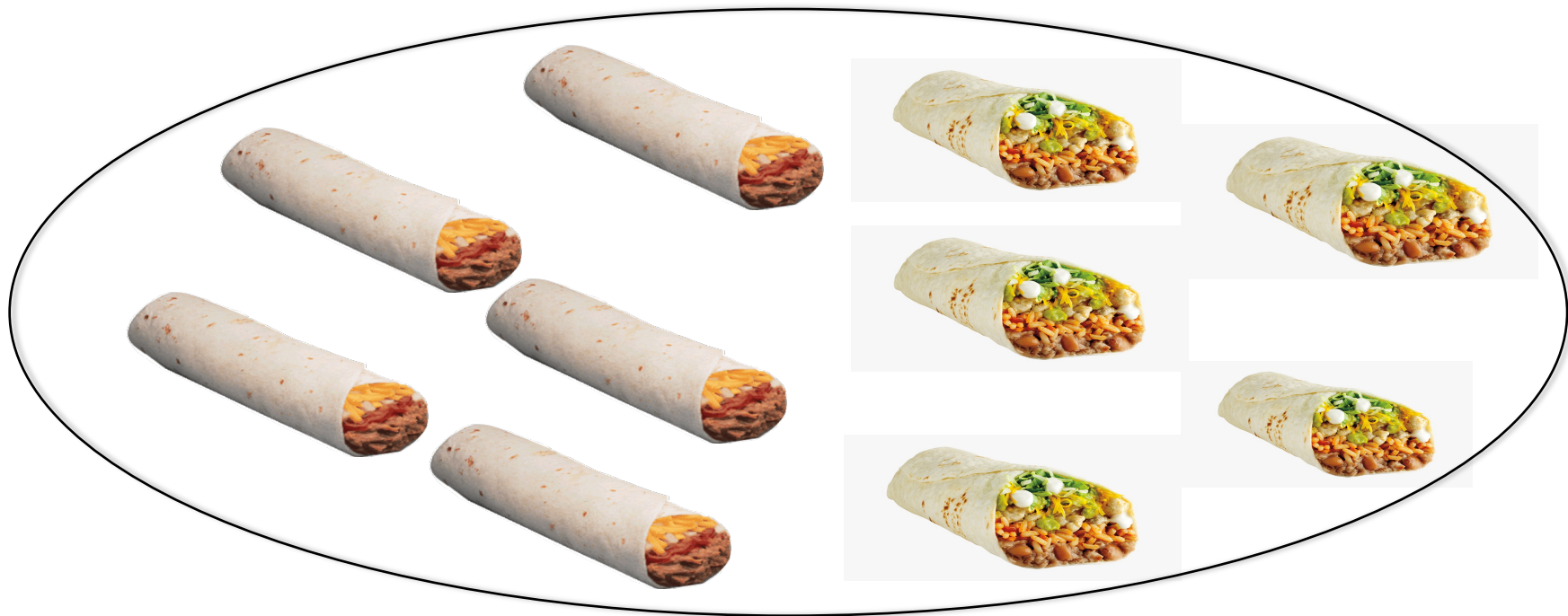
This is the “neutral” or statistical meaning of the word bias. You will see it often in discussions of patterns in data.



# EXAMPLE: SAMPLING BIAS



# EXAMPLE: SAMPLING BIAS



**What Kinds of Bias  
Raise  
Ethical Concerns?**



# DISCRIMINATORY BIAS IN DATA



**Biased measurement or classification**

**+ use of that bias to compound existing injustice or to fail to treat all as having equal moral worth (Hellman 2020)**

**=> *Discriminatory or Unfair Bias***

**HOW DOES  
DEFINING BIAS  
HELP US  
UNDERSTAND OUR  
PROFESSOR  
RATINGS DATA?**

# I HAVE DATA ABOUT PEOPLE! NOW WHAT? CHECKING BIAS

## Decide how to use the data given the bias

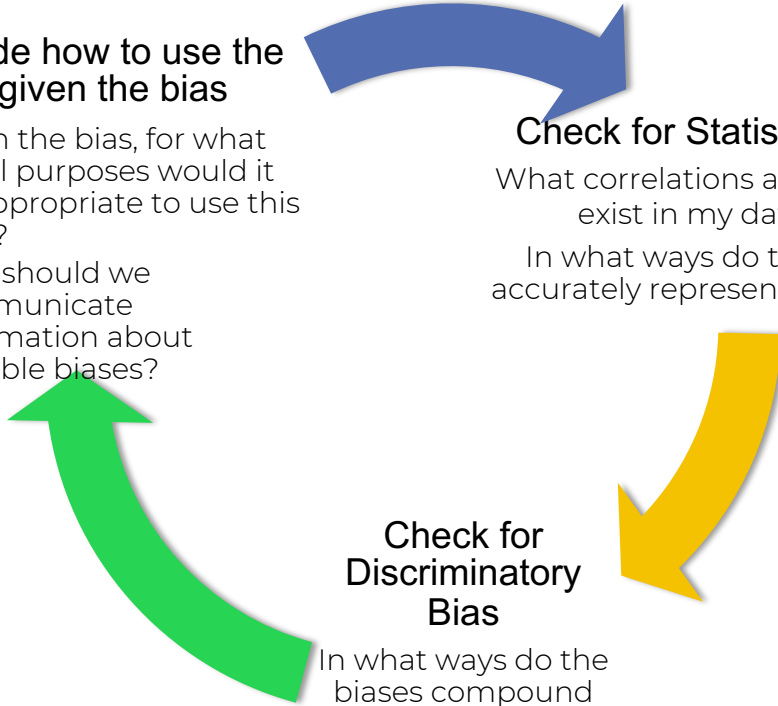
- Given the bias, for what social purposes would it be appropriate to use this data?
- How should we communicate information about possible biases?

## Check for Statistical Bias

What correlations and patterns exist in my dataset?  
In what ways do they fail to accurately represent the world?

## Check for Discriminatory Bias

In what ways do the biases compound existing injustice or fail to treat all people as being of equal moral worth?



**WHAT  
FRAMEWORKS CAN  
WE USE TO  
EVALUATE FAIR  
DISTRIBUTIONS?**

# Equality of Opportunity:

everyone has same opportunity to develop skills needed for the job, apply for the job, and get promoted.



**Parity: Everyone is equally likely to be a good teacher, so we should expect to end up with number of good teachers (and high rankings) proportionate to population.**



# OTHER DEFINITIONS OF FAIRNESS (IN CS 109!)

## Simulating loan decisions for different groups

Drag the black threshold bars left or right to change the cut-offs for loans.  
Click on different preset loan strategies.

### Loan Strategy

Maximize profit with:

**MAX PROFIT**

No constraints

**GROUP UNAWARE**

Blue and orange thresholds are the same

**DEMOGRAPHIC PARITY**

Same fractions blue / orange loans

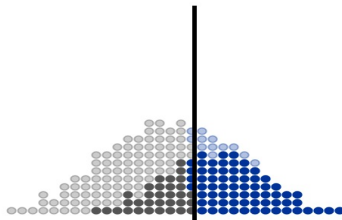
**EQUAL OPPORTUNITY**

Same fractions blue / orange loans to people who can pay them off

### Blue Population

0 10 20 30 40 50 60 70 80 90 100

loan threshold: 59

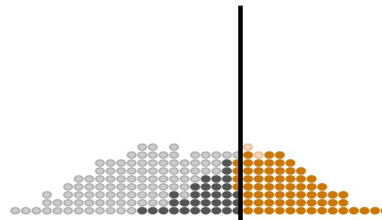


denied loan / would default (grey square)  
denied loan / would pay back (dark grey square)  
granted loan / defaults (light blue square)  
granted loan / pays back (dark blue square)

### Orange Population

0 10 20 30 40 50 60 70 80 90 100

loan threshold: 53



denied loan / would default (grey square)  
denied loan / would pay back (dark grey square)  
granted loan / defaults (light orange square)  
granted loan / pays back (dark orange square)

# **FAIRNESS BEYOND THE NUMBERS**

QUALITY

5.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

DIFFICULTY

1.0

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



QUALITY

5.0

DIFFICULTY

1.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



QUALITY

5.0

DIFFICULTY

1.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that ~~has never done~~ anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



**AWESOME**

QUALITY

**5.0**

DIFFICULTY

**1.0**



**CS1**



**AWESOME**

May 21st, 2015

Attendance: **No** **Mandatory** Grade: **A** Textbook: **No** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

**CLEAR**

**GREAT TEACHER**



# DESCRIPTIVE VS NORMATIVE LANGUAGE

## Descriptive language

- “is”
- “was”
- what people did
- what happened

## Normative language

- “right”
- “wrong”
- “good”
- “bad”
- “should”
- “should not”

# Descriptive or Normative?

QUALITY

5.0



CS101



AWESOME

21st, 2015

Attendance: **Not Mandatory** Grade: **A** Texts:

**CLEAR**

DIFFICULTY

1.0

This class was awesome. A beginner like me that had never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

**GREAT TEACHER**





**Descriptive**



**Normative**

# THICK NORMATIVE TERMS

- “Thick” normative terms are both descriptive AND normative
- We compare people in many ways, not just numerically
- Thick normative terms express morally or aesthetically “loaded” judgments

## EXAMPLES:

Cowardly :: Cautious

Polite :: (?)

Rude :: (?)

Chill :: (?)

Kind :: (?)

etc

# DESCRIPTIVE OR NORMATIVE?



**Do programs like the ones you are writing contain descriptive claims?**



**Do they contain normative claims or values?**



**How about thick normative terms?**

# **HOW ARE VALUES EMBEDDED IN DESIGN?**

# **1. PROBLEM FORMULATION STATEMENTS ARE NORMATIVE & EMBED VALUES**

**Formulating a problem means describing the desired solution as good or worthy of being done.**

- What is the problem to be solved?**
- For whom is this a problem? Who would benefit from its solution?**
- Who can agree that this is a problem worth solving?**

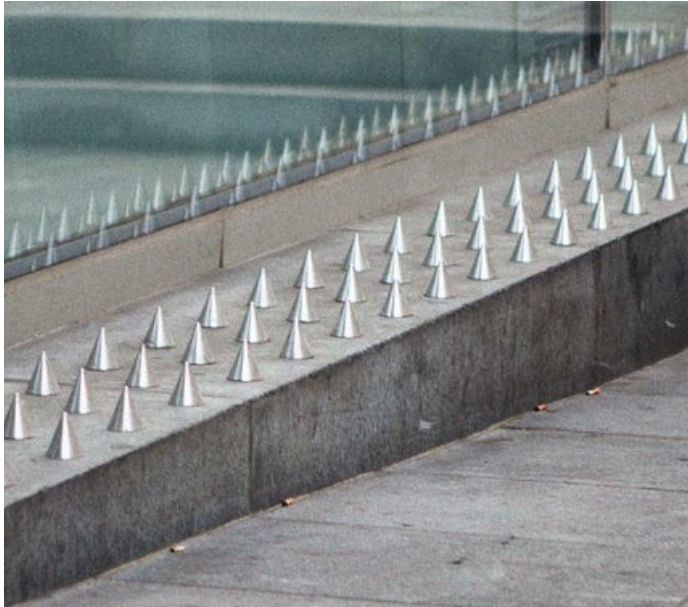
# WHAT IS THE PROBLEM TO BE SOLVED?

**“HOMELESS PEOPLE  
ARE SLEEPING HERE  
AND WE (WHO IS WE?)  
WANT THEM TO STOP”**



# WHAT IS THE PROBLEM TO BE SOLVED?

**PEOPLE ARE SLEEPING  
HERE AND WE (WHO IS  
WE?) WANT THEM TO  
STOP”**



**“SOME PEOPLE IN OUR  
COMMUNITY DON'T HAVE  
A PLACE TO SLEEP AND  
WE (WHO IS WE?) THINK  
THEY SHOULD”**



# WHAT IS THE PROBLEM TO BE SOLVED?

**PEOPLE ARE SLEEPING  
HERE AND WE (WHO IS  
WE?) WANT THEM TO  
STOP”**



**“SOME PEOPLE IN OUR  
COMMUNITY DON'T HAVE  
A PLACE TO SLEEP AND  
WE (WHO IS WE?) THINK  
THEY SHOULD”**



Note: Who is included in each problem formulation? Who can agree it's a problem?

# 1. CHOOSING WHICH PROBLEM TO SOLVE EXPRESSES VALUES



- What are the problem(s) to be solved?
- For whom are these problems?
- Who would benefit from their solution(s)?
- For each problem, who can agree that the problem is worth solving?

# 2. CHOICE OF DATA EMBEDS VALUES

CS 106A Mid-Quarter Evaluation

\* Required

Instructor Feedback

These first few questions come from Mehran Sahami, Chris Piech, and Juliette Woodrow. Your anonymous feedback will be reviewed by them, so please be specific and give suggestions for improvement.

What has Mehran been doing so far that has worked well for you? What should he continue doing? \*

Your answer

What has Mehran been doing so far that has not worked well for you? Is there anything that he should stop or start doing?

Your answer

What has Chris been doing so far that has worked well for you? What should he

**Surveys are cheap  
to run**

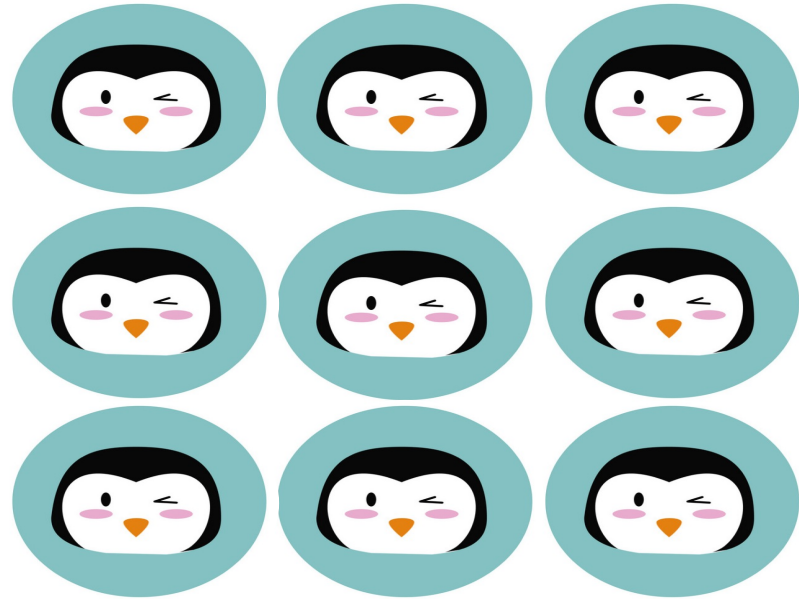
**They measure  
opinions**

**What are other  
ways to measure  
quality of  
professors?**

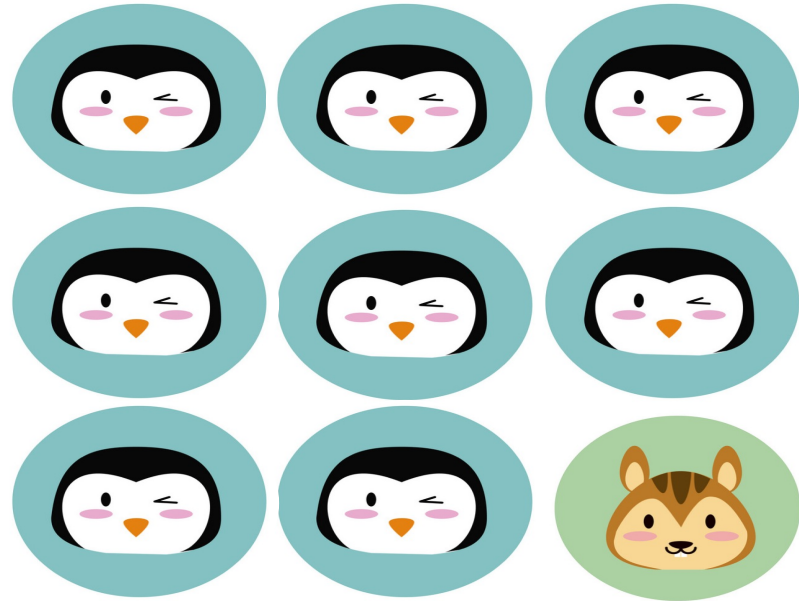
# **SURVEY DATA AND REPRESENTATIONAL HARMS**

- **Who is represented in this data?**
- **Who can express themselves in it?**
- **Which people can represent or express themselves and the cultures that are meaningful to them in this system?**

# REPRESENTATION IN SURVEY DATA

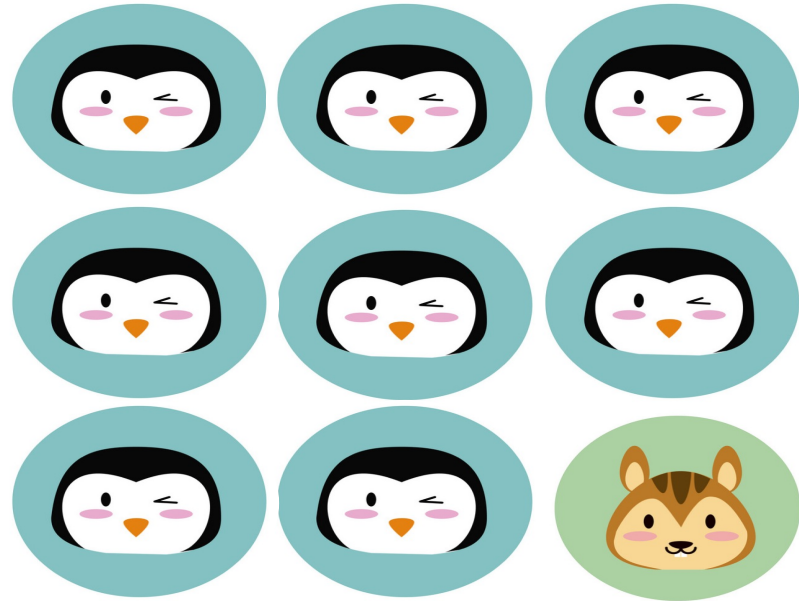


# REPRESENTATION IN SURVEY DATA



If I had a different experience, my numerical ratings may not affect the survey data – but they matter!

# REPRESENTATION IN SURVEY DATA



I want to see myself represented in search results, but other people are using different terms to describe me!



**THANK YOU!**

**You can always get in touch at [kcreel@stanford.edu](mailto:kcreel@stanford.edu)  
Or plan a visit to my office hours at  
[calendly.com/kathleencreel](https://calendly.com/kathleencreel)**