



# Machine Learning - Fairness

DataBite Summer 2022

Dr. Christan Grant

[oudatalab.com](http://oudatalab.com)



*The University of Oklahoma*

# Data Bite Summer 2022 Schedule

## Day 1

Welcome +  
Intro to ML

## Day 2

Bias and Fairness  
+ Introduction to  
Probability

## Day 3

Classifiers and  
Clustering

## Day 4

Model Olympics

“With great power comes great responsibility.”

Peter Parker Principle (RIP Uncle Ben)

# AI at Google: our principles



Sundar Pichai

CEO

Published Jun 07, 2018

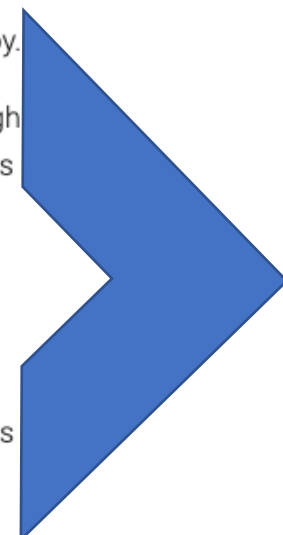
At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound. At Google, we use AI to make products more useful—from email that's spam-free and [easier to compose](#), to a digital assistant you can [speak to naturally](#), to photos that [pop the fun stuff out](#) for you to enjoy.

Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to [predict the risk of wildfires](#). Farmers are using it to monitor the [health of their herds](#). Doctors are starting to use AI to help [diagnose cancer](#) and [prevent blindness](#). These clear benefits are why Google invests heavily in AI research and development, and makes AI technologies widely available to others via our tools and open-source code.

We recognize that such powerful technology raises equally powerful questions about its use. How AI is developed and used will have a significant impact on society for many years to come. As a leader in AI, we feel a deep responsibility to get this right. So today, we're announcing seven principles to guide our work going forward. These are not theoretical concepts; they are concrete standards that will actively govern our research and product development and will impact our business decisions.

We acknowledge that this area is dynamic and evolving, and we will approach our work with humility, a commitment to internal and external engagement, and a willingness to adapt our approach as we learn over time.

<https://www.blog.google/technology/ai/ai-principles/>



# Objectives for AI applications



- 1. Be socially beneficial.**
- 2. Avoid creating or reinforcing unfair bias.**
- 3. Be built and tested for safety.**
- 4. Be accountable to people.**
- 5. Incorporate privacy design principles.**
- 6. Uphold high standards of scientific excellence.**
- 7. Be made available for uses that accord with these principles.**

# AI Applications we won't pursue

1. **Technologies that cause or are likely to cause harm.** Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.
2. **Weapons** or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.
3. Technologies that gather or use information for **surveillance** violating *internationally* accepted norms.
4. Technologies whose purpose contravenes widely accepted principles of *international law* and human rights.

# Bias Types

# Reporting Bias

occurs when the frequency of events, properties, and/or outcomes captured in a data set does not accurately reflect their real-world frequency. This bias can arise because people tend to focus on documenting circumstances that are unusual or especially memorable



# Automation Bias

is a tendency to favor results generated by automated systems over those generated by non-automated systems, irrespective of the error rates of each.

# Selection bias

occurs if a data set's examples are chosen in a way that is not reflective of their real-world distribution.

# Selection bias

occurs if a data set's examples are chosen in a way that is not reflective of their real-world distribution.

- **Coverage bias** -- occurs when data is not selected in a representative fashion.

# Selection bias

occurs if a data set's examples are chosen in a way that is not reflective of their real-world distribution.

- **Sampling bias** -- occurs when proper randomization is not used during data collection.

# Selection bias

occurs if a data set's examples are chosen in a way that is not reflective of their real-world distribution.

- **Non-response bias** -- occurs when data are unrepresentative due to participation gaps in the data collection process.

# Group Attribution Bias

is a tendency to generalize what is true of individuals to an entire group to which they belong.

# Group Attribution Bias

is a tendency to generalize what is true of individuals to an entire group to which they belong.

- **In-group bias** -- A preference for members of a group to which you also belong, or for characteristics that you also share.

# Group Attribution Bias

is a tendency to generalize what is true of individuals to an entire group to which they belong.

- **Out-group homogeneity bias** -- A tendency to stereotype individual members of a group to which you do *not* belong, or to see their characteristics as more uniform.



# Confirmation Bias

is where model builders unconsciously process data in ways that affirm preexisting beliefs and hypotheses.

# Confirmation Bias

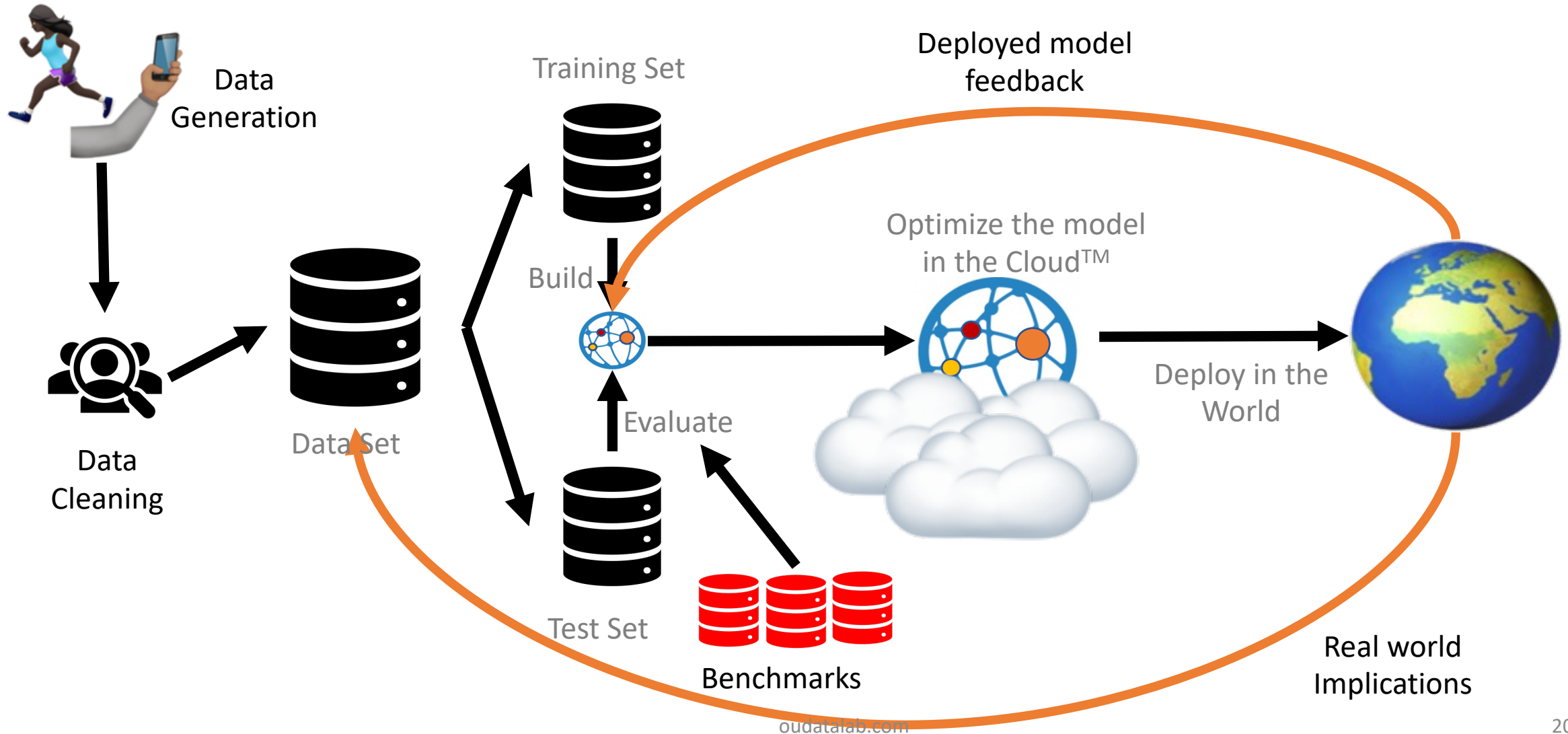
is where model builders unconsciously process data in ways that affirm preexisting beliefs and hypotheses.

- **Experimenter's bias** – a model builder may actually keep training a model until it produces a result that aligns with their original hypothesis.

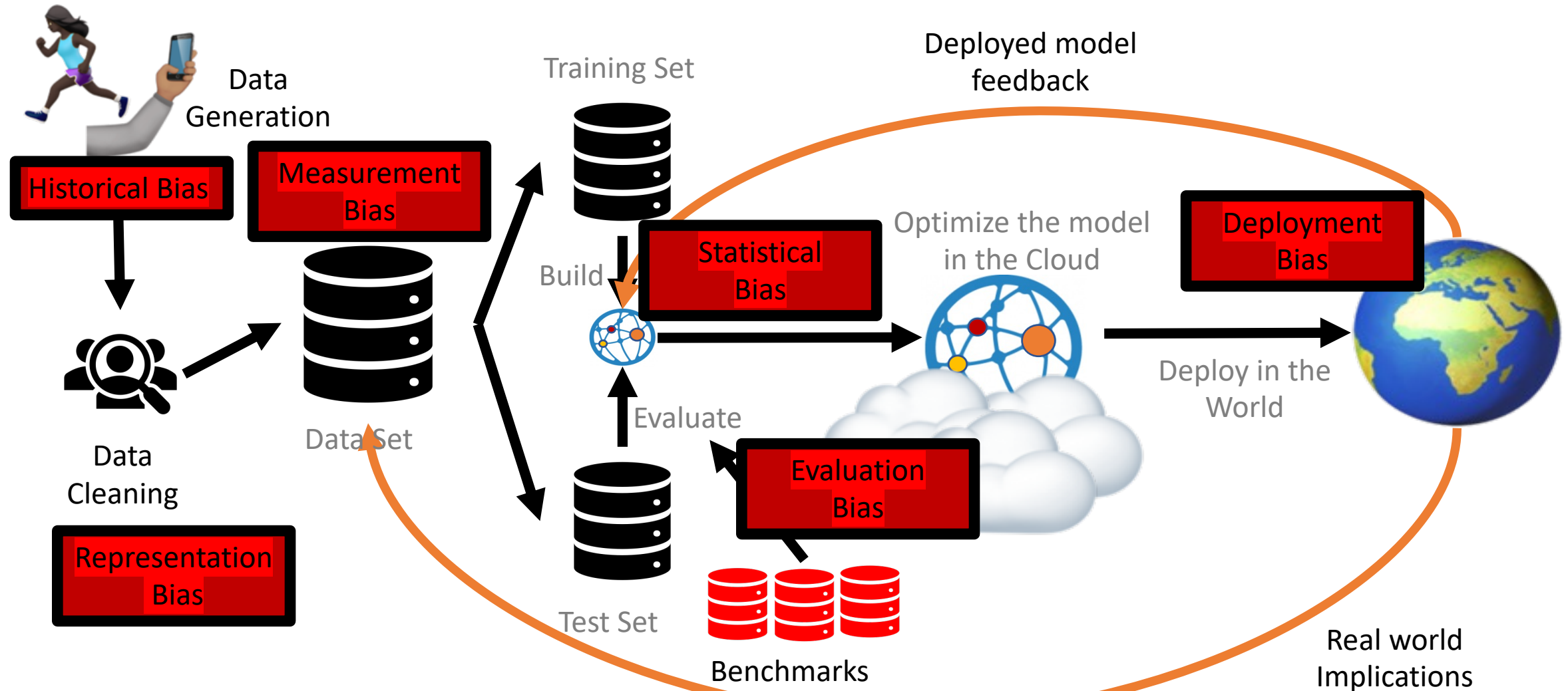
# Bias Types

- Reporting
- Automation
- Selection (coverage, non-response, sampling)
- Group attribution (in-group, out-group)
- Implicit (confirmation, experimenters)

# The (closer) to full picture of the pipeline.

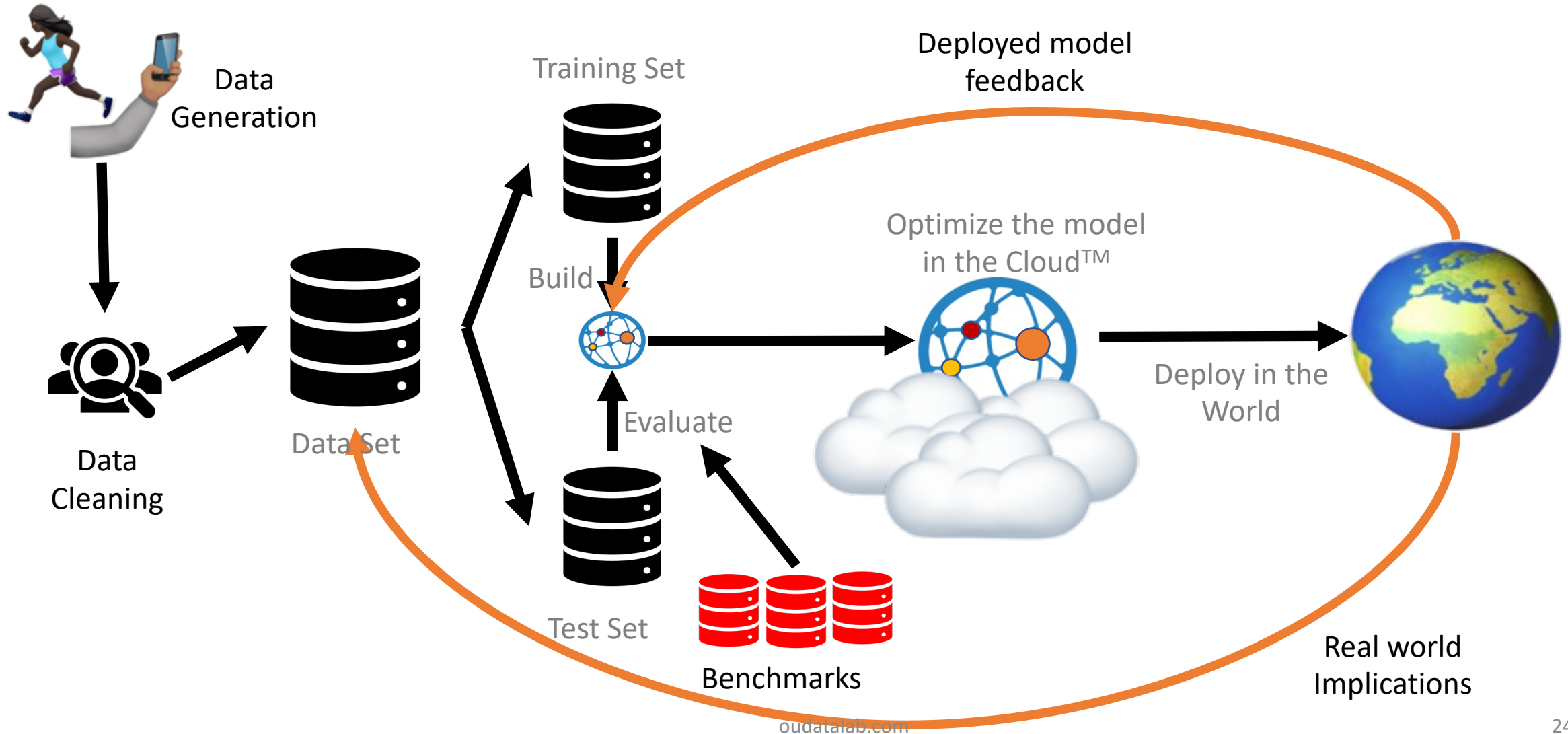


# There is Bias hiding in every step!



# Bias in the AI/ML Pipeline

# The (closer) to full picture of the pipeline.



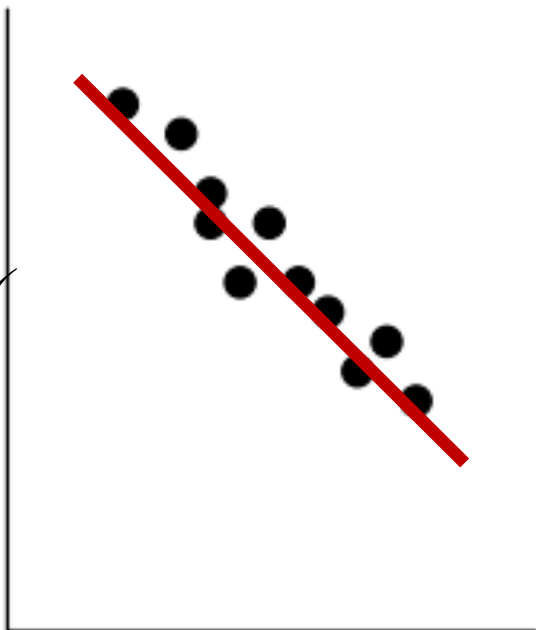
# Biases in Classification

Simpson's Paradox by example (Statistical Bias)

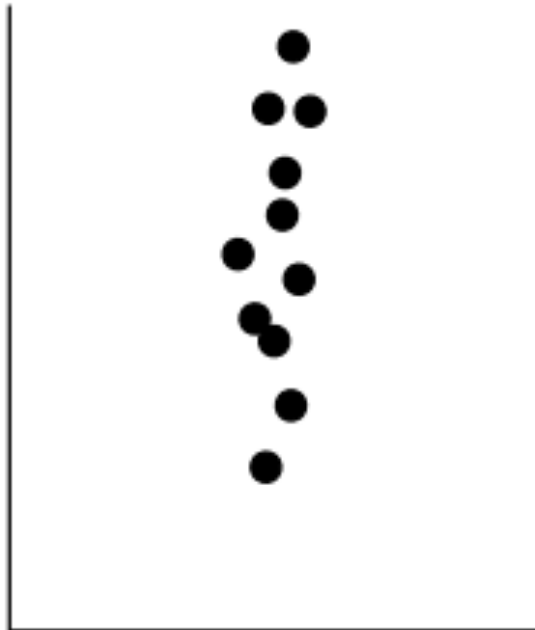


*Y*

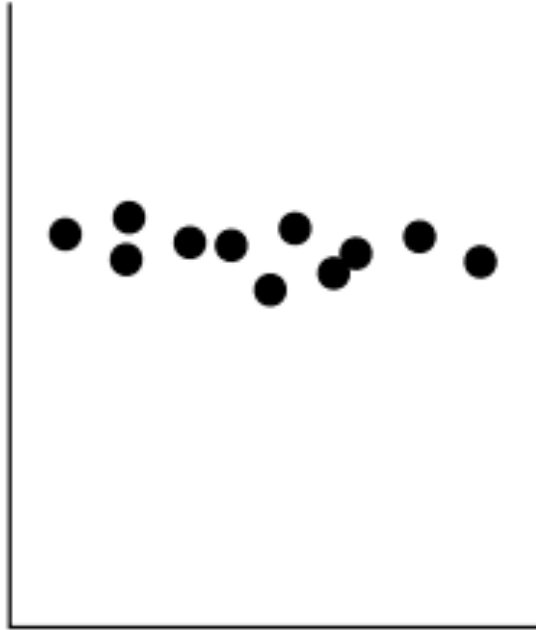
Negative  
Correlation



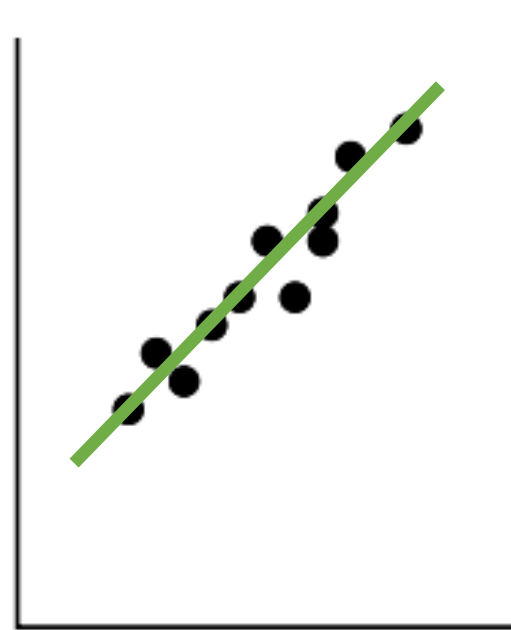
No  
Correlation



No  
Correlation



Positive  
Correlation



*x*

# Trends in data

# In real world data it is not so simple

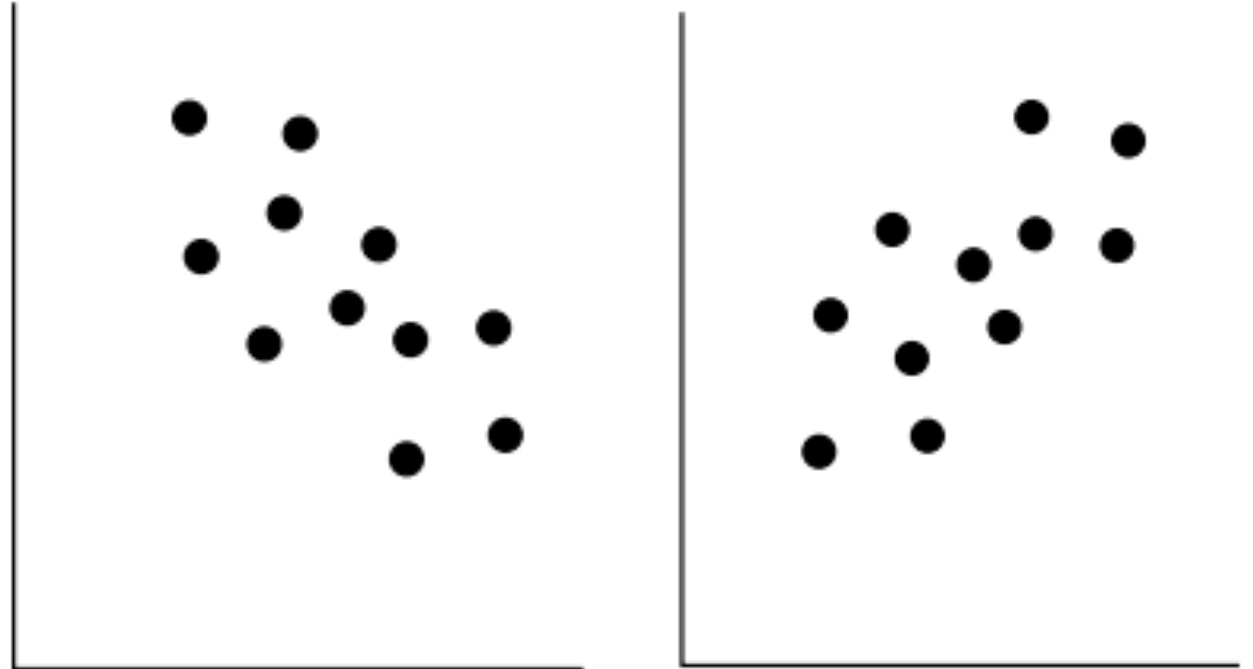
- Correlations can be weak
- We use Pearson's Correlation coefficient to determine the trends.

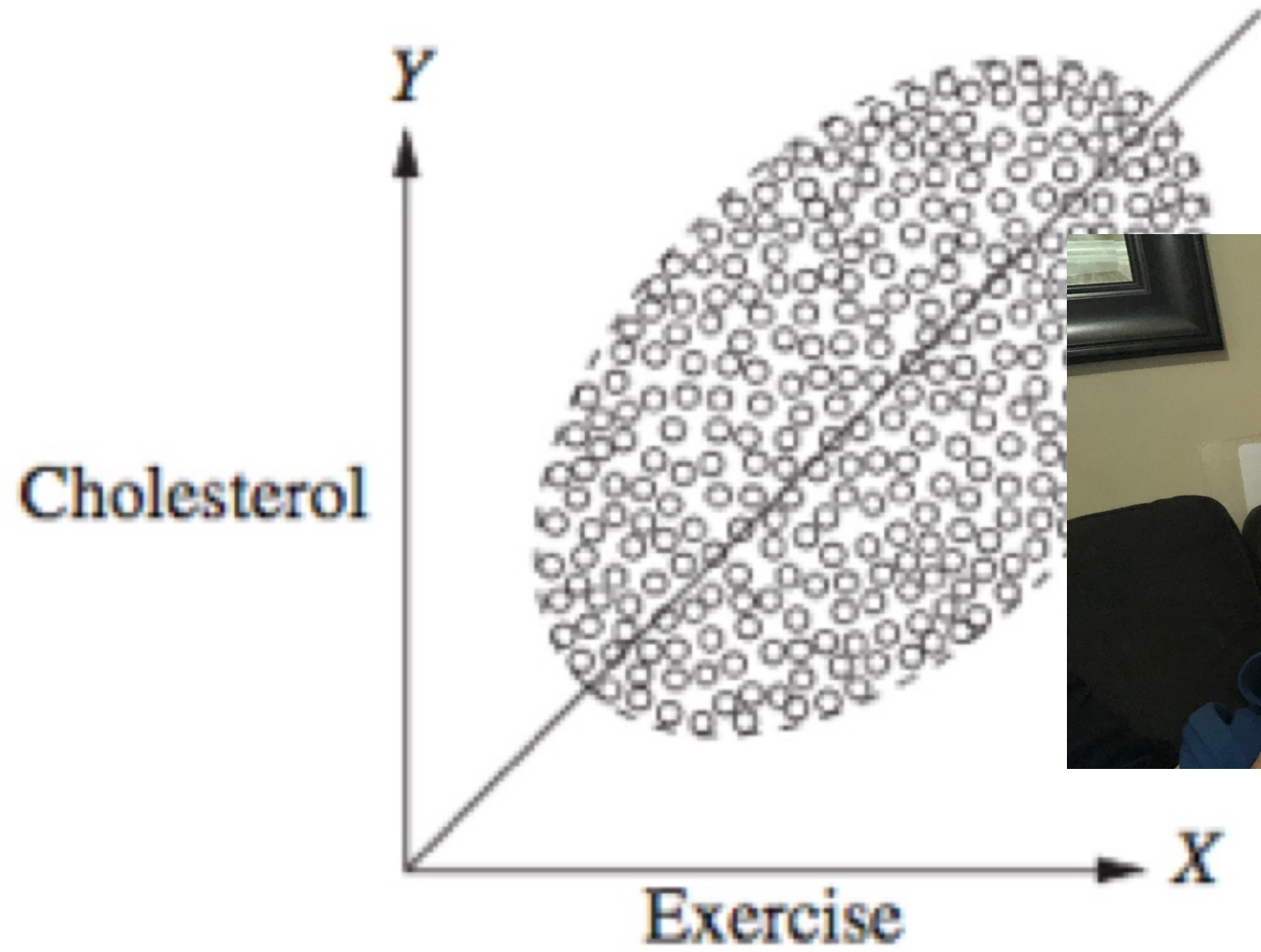
$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

0 < |r| < 0.3 weak correlation

0.3 < |r| < 0.7 moderate correlation

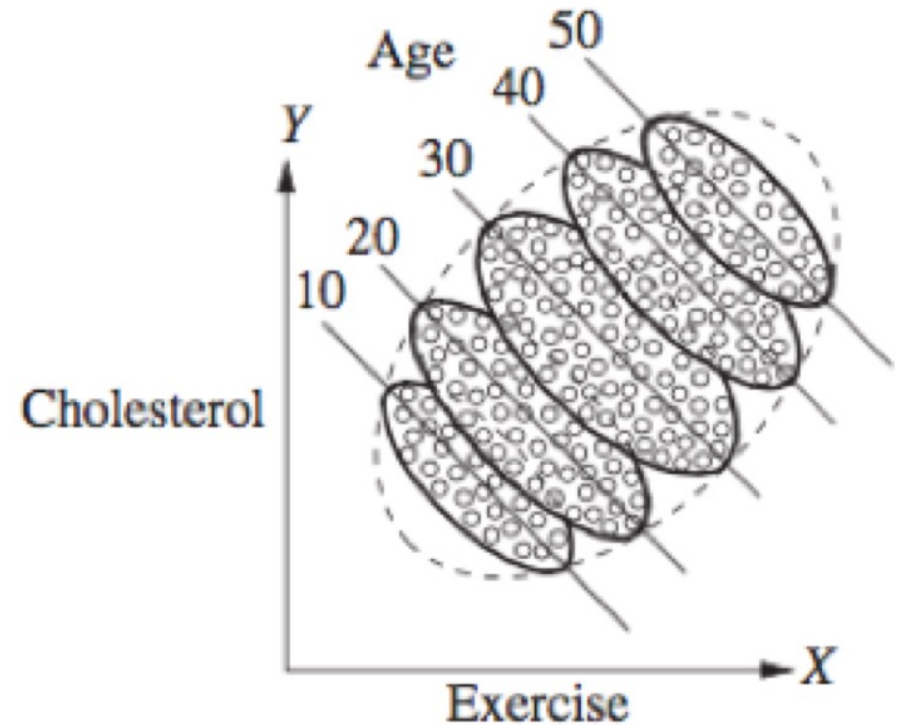
|r| > 0.7 Strong correlation





# Trends in data can be misleading.

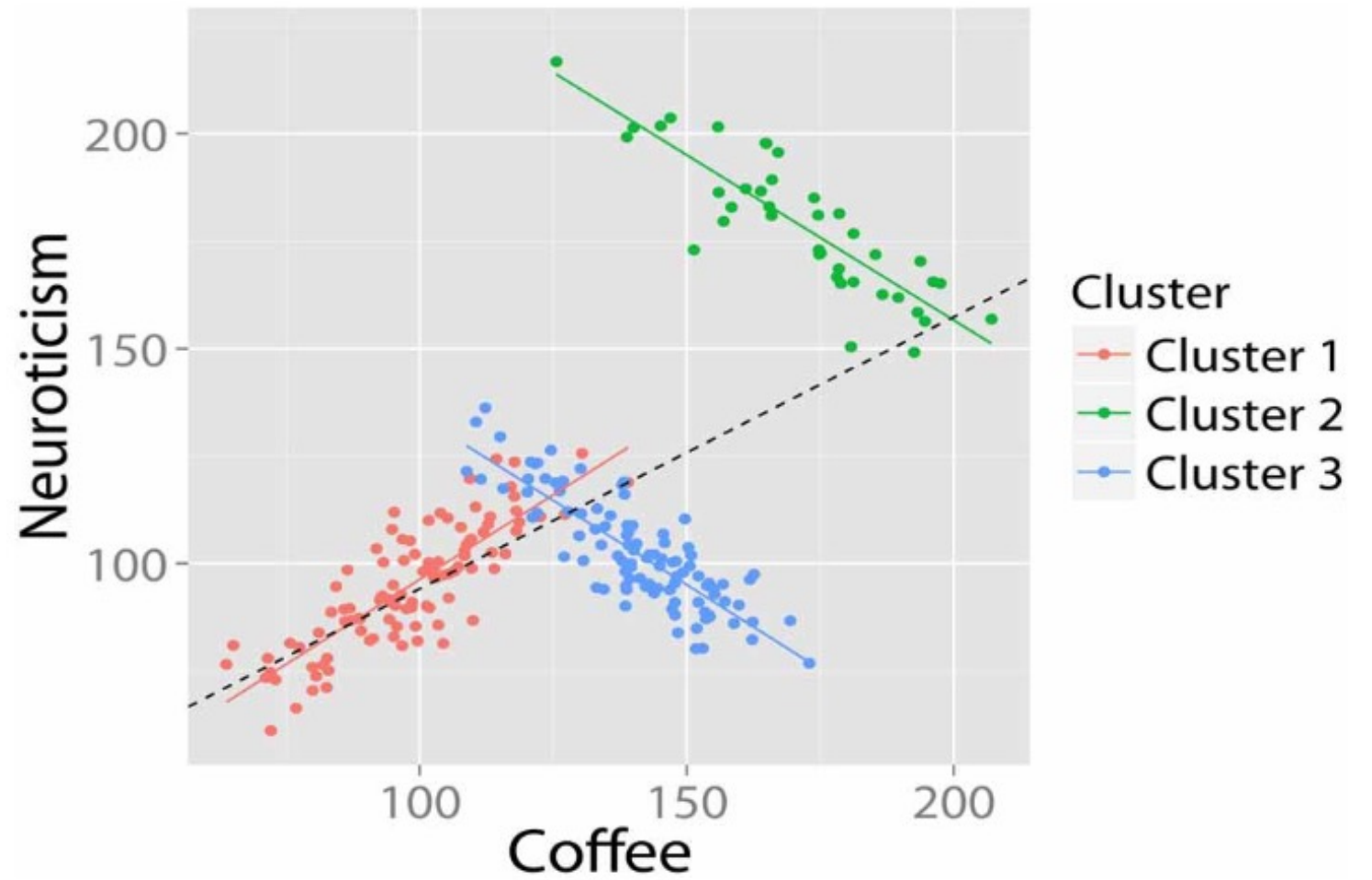
When we separate the data by another parameter, we uncover a more appropriate trend.



# Simpson's Paradox

- When a trend between two variables is reversed in *all* subgroups of the data.
- If the trend is reversed for *some* subgroups, it is a mix effect.

# Mix Effect



# Simpson's Paradox (Rate-based)



	Hits / At Bats			
David Justice				
Derek Jeter				

# Mix Effects (Rate-based)

	<b>Applicants</b>	<b>Admitted</b>
<b>Men</b>	8442	<b>44%</b>
<b>Women</b>	4321	<b>35%</b>

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
<b>A</b>	825	62%	108	<b>82%</b>
<b>B</b>	560	63%	25	<b>68%</b>
<b>C</b>	325	<b>37%</b>	593	34%
<b>D</b>	417	33%	375	<b>35%</b>
<b>E</b>	191	<b>28%</b>	393	24%
<b>F</b>	373	6%	341	<b>7%</b>






# Introduction to Probability

# Rolling a Die Creates a Random Variable

Random Variable



X	Probability(X)
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

- If we roll a 6-sided die, what is the probability of rolling a 1?
- What is the probability of rolling an even number?



# Die Rolls are Uniform Probabilities

- When we roll a 6-sided die, what is the "most likely" value?
- Imagine rolling the die 100 times, what would the "average roll be?"
- $3.5 = 1*(1/6) + 2*(1/6) + 3*(1/6) + \dots + 6*(1/6)$
- $3.5 = (1/100)*(100*1*(1/6) + 100*2*(1/6) + \dots + 100*6*(1/6))$
- The expected value of a random variable can be thought of as the *mean* or *average*.



```
import random
```

```
rolls = [random.randint(1,6) for i in  
range(0,100000)]
```

```
average_rolls = sum(rolls)/len(rolls)
```

```
print(average_rolls)
```

most

ould the

(1/6)

) + ... +

can be

# Relationships Among Random Variables

- **Independent variables:** knowing one event has happened does not change the probability that the other happens
  - Probability of rolling a 1 and flipping a head
  - When X and Y are independent,  $P(X \text{ and } Y) = P(X)P(Y)$
- **Dependent variables:** knowing one event has happened gives us new information, affecting the probability that the other happens
  - Probability that the sum of two die rolls being a 5, if the first roll was a 3

# Conditional Probability

The probability of X given Y has occurred is  $P(X|Y)$ , for example,  
 $P(\text{sum} = 5 \text{ first die} = 3) = P(\text{sum is 5 if first die is 3}) = 1/6$

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Joint Probability

Conditional Probability

Marginal Probability

# Probability: Example

- By looking at a table of all possibilities, we found that

$$P(\text{sum} = 5 | \text{first die} = 3) = \frac{1}{6}$$

- Now, using  $P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)}$ , we can calculate this without needing to find every possible value of two dice rolls:

$$\begin{aligned} P(\text{sum} = 5 | \text{first die} = 3) &= \frac{P(\text{second die} = 2 \text{ and first die} = 3)}{P(\text{first die} = 3)} \\ &= \frac{1/36}{1/6} \\ &= \frac{1}{6} \end{aligned}$$

# Conditional Probabilities are not Joint Probabilities

If we let  $X$  be the first die roll of value 3, and  $Y$  be the second of value 2, and  $Z$  be the sum, then

- Conditional probability:

$$P(Z|X) \longrightarrow P(\text{sum} = 5 | \text{first die} = 3) = \frac{1}{6}$$

- Joint probability:

$$P(X \text{ and } Y) \longrightarrow P(\text{roll 2 and 3}) = \frac{2}{36}$$

- Probability of  $Z$

$$P(Z) = P(X + Y) \longrightarrow P(\text{sum} = 5) = \frac{4}{36}$$



# Conditional, Joint, & Marginal Probabilities are Related

Let  $X$  and  $Y$  be random variables.

1. If  $X$  and  $Y$  are independent then 
$$P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)} = \frac{P(X)P(Y)}{P(Y)} = P(X)$$

*Example: let  $X$  be the outcome of rolling a 6-sided die and let  $Y$  be the outcome of flipping a coin. Suppose we know that  $Y$  is "heads." What is the probability that we roll a 3? This tells us that  $P(\text{roll } 3|\text{heads}) = P(\text{roll } 3)$ . This matches intuition — flipping a coin does not change the outcome of rolling a die.*

# Conditional, Joint, & Marginal Probabilities are Related

Let  $X$  and  $Y$  be random variables.

$$2. P(X) = \sum_Y P(X|Y)P(Y)$$

*Example: let  $X$  be the sum of rolling two dice and let  $Y$  be the outcome of the first die roll. If we want to know the probability that  $X$  is 3, then this tells us*

*$P(\text{sum is 3}) = \sum_Y P(\text{sum is 3}|\text{first roll was } Y)P(\text{first roll was } Y)$ . From here, we know that the sum can never be 3 unless the first roll is 1 or 2. Thus,*

$$\begin{aligned} P(\text{sum is 3}) &= P(\text{sum is 3}|\text{first roll was 1})P(\text{first roll was 1}) + P(\text{sum is 3}|\text{first roll was 2})P(\text{first roll was 2}) \\ &= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{2}{36} \end{aligned}$$

# Conditional, Joint, & Marginal Probabilities are Related

Let  $X$  and  $Y$  be random variables.

$$3. P(X) = \sum_Y P(X \text{ and } Y)$$

*Example: let  $X$  be the outcome of a first die roll and let  $Y$  be the outcome of a second die roll. If we want to find the probability that  $X$  is 3, this tells us that*

$$\begin{aligned} P(X = 3) &= \sum_{y=1}^6 P(X = 3 \text{ and } Y = y) \\ &= \sum_{y=1}^6 \frac{1}{36} = \frac{6}{36} = \frac{1}{6} \end{aligned}$$

# Conditional, Joint, & Marginal Probabilities are Related

Let  $X$  and  $Y$  be random variables.

1. If  $X$  and  $Y$  are independent then  $P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)} = \frac{P(X)P(Y)}{P(Y)} = P(X)$

2.  $P(X) = \sum_Y P(X|Y)P(Y)$

Conditional Probability

3.  $P(X) = \sum_Y P(X \text{ and } Y)$

Joint Probability

Marginal Probability

# Conditional Probabilities and Bayes' Theorem

Sometimes we want to find  $P(X|Y)$  when we already know  $P(Y | X)$

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

# Bayes' Theorem: Example

Sometimes we want to find  $P(X|Y)$  when we already know  $P(Y|X)$ .

For instance,  $P(\text{first die} = 3 | \text{sum} = 5) = \frac{1}{4}$ .

We can verify this using Bayes' Theorem.

$$\begin{aligned} P(\text{first die} = 3 | \text{sum} = 5) &= \frac{P(\text{sum} = 5 | \text{first die} = 3)P(\text{first die} = 3)}{P(\text{sum} = 5)} \\ &= \frac{\frac{1}{6} \cdot \frac{1}{6}}{\frac{4}{36}} \\ &= \frac{1}{4} \end{aligned}$$

## Sample **Exercise**: Peanut Chocolate Detector

**Assumptions:** Suppose we have a new device that distinguishes whether or not a type of chocolate contains peanuts. If a chocolate contains peanuts, 99% of the time it correctly reports a positive result. Likewise, if a chocolate does not contain peanuts, 99% of the time it correctly reports a negative result. Imagine 1% of all chocolates contain peanuts.

**Question:** If the device reports that a chocolate contains peanuts, what is the probability that the chocolate *actually does* contain peanuts?

## Sample Exercise: Peanut Chocolate Detector

$p$  = random variable indicating whether peanuts are in a chocolate bar

$d$  = random variable indicating whether we detected peanuts in a chocolate bar.

We were given  $P(p) = 0.01$ ,  $P(d|p) = 0.99$ , and  $P(\text{not } d|\text{not } p) = 0.99$ .

We can then calculate  $P(d|\text{not } p) = 0.01$  and  $P(\text{not } p) = 0.01$ .

Then,  $P(d) = P(d|p)P(p) + P(\text{not } d|\text{not } p)P(\text{not } p)$ .

By Bayes' Theorem,  $P(p|d) = \frac{P(d|p)P(p)}{P(d)} = \frac{0.99 \cdot 0.01}{0.0198} = 0.5$ .





Thanks!



@oudatalab