DSC 40A

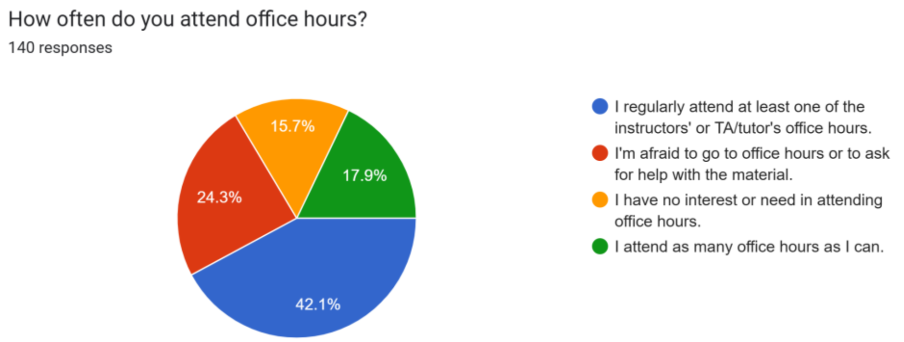
Theoretical Foundations of Data Science I

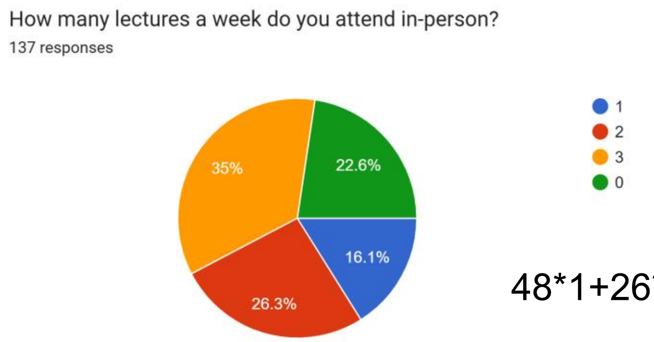
Announcements

- Homework 6 due Monday
- Homework 7 released 11/27 and due 12/6.

(No slip day for HW 7)

Course Survey





48*1+26*2/3+22*1/3=72

How helpful are OH in helping you understand/practice course content?

- Extremely helpful 49
- Helpful 39

Favorite aspect of the course:

- *Office hours, they are really good for learning while having fun.*
- *I think office hours are fun and a good vibe*
- *The ability to go to office hours for help*
- *office hours is where the class grows interesting. I love attending office hours to do work with the tutors and to essentially see how everyone thinks about the problems at hand.*

Feedback for staff

- *I find the staff to be very helpful with their explanations*
- *They're all really cool people and I enjoy spending time in office hours with them.*
- *I think the DSC 40A staff have been really helpful so far in OH*

Agenda

- Law of total probability.
- Bayes theorem.

Question

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!
If the direct link doesn't work, click the "Lecture Questions" link in the top right corner of dsc40a.com.

Getting to Campus

- You conduct a survey:
  - How did you get to campus today? Walk, bike, or drive?
  - Were you late?

Walk & Late

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

all sum to 100%

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

$6\% + 3\% + 36\% = 45\%$

What is the probability that a randomly selected person is late?

A. 24%
B. 30%
C. 45%
D. 50%

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

- Since everyone either walks, bikes, or drives,

P(Late) = P(Late AND Walk) + P(Late AND Bike) + P(Late AND Drive)

$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$

- This is called the **Law of Total Probability**.

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

= 30%

Suppose someone tells you that they walked. What is the probability that they were late?

A. 6%
B. 20% ←
C. 25%
D. 45%

$$\text{Prob}(\text{late} \mid \text{walk}) = \frac{\text{Prob}(\text{late} \,\&\, \text{walk})}{P(\text{walk})} = \frac{6\%}{30\%} = 20\%$$

Need Prob(walk)

$$\text{Prob}(\text{walk}) = \text{Prob}(\text{walk} \,\&\, \text{late}) + \text{Prob}(\text{walk} \,\&\, \text{not late})$$

$$= 6\% + 24\% = 30\%$$

$$\text{Prob}(\text{walk} \,\&\, \text{late}) = \text{Prob}(\text{late} \mid \text{walk}) \cdot P(\text{walk})$$

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

- Since everyone either walks, bikes, or drives,

P(Late) = P(Late AND Walk) + P(Late AND Bike) + P(Late AND Drive)

Using mult. rule:

P(Late) = P(Late|Walk)*P(Walk) + P(Late|Bike)*P(Bike)
+P(Late|Drive)*P(Drive)

Partitions

- A set of events $E_1, E_2, \ldots, E_k$ is a **partition** of S if
  - $P(E_i \cap E_j) = 0$ for all i,j ← All pairs of events are mutually exclusive
  - $P(E_1) + P(E_2) + \ldots + P(E_k) = 1$ $= P(S)$

$$P(S) = 1 = \sum_i P(E_i)$$

Every outcome in S belongs to only one of the E_i's

Example

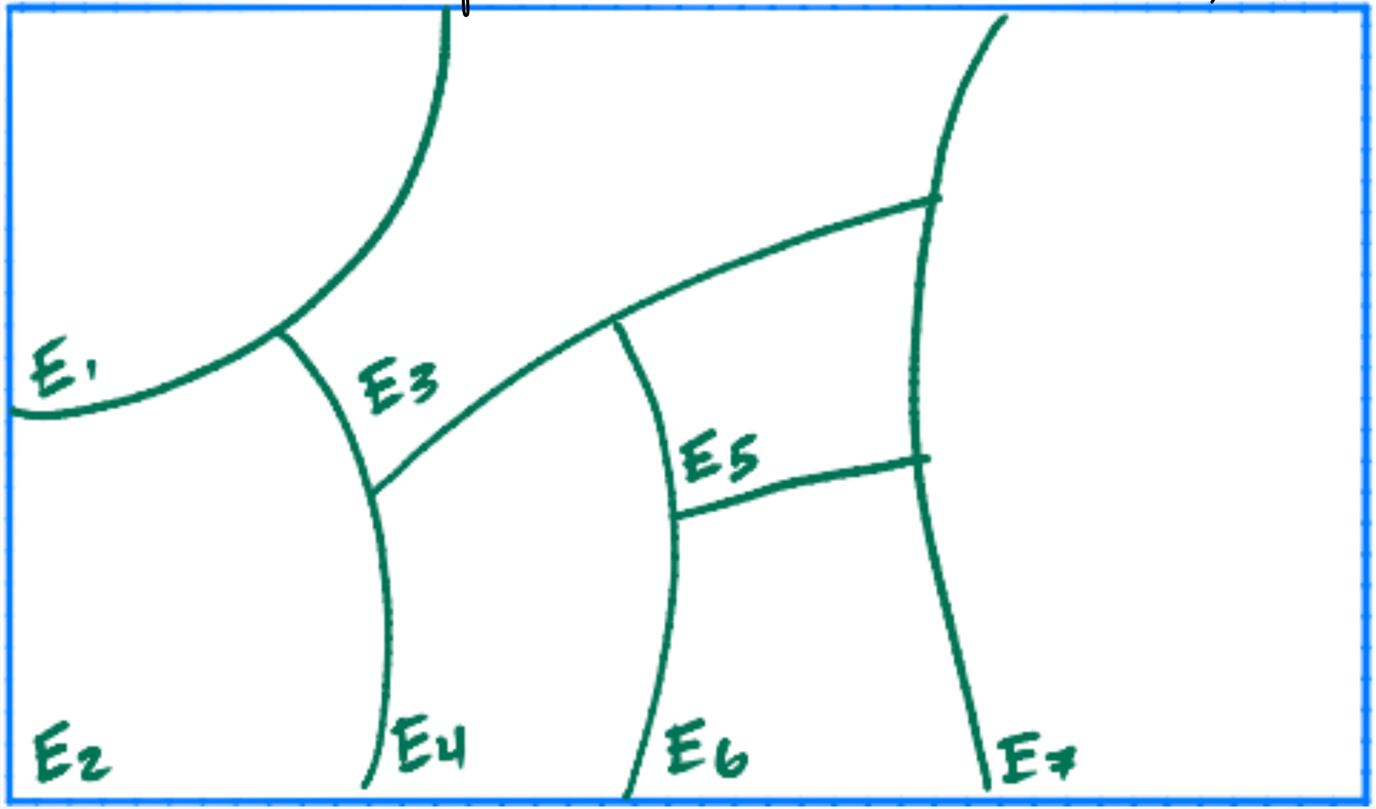
Walk	30%
Bike	10%
Drive	60%

Late	Not late
45%	55%

Partitions

No overlap between pairs of events $\Rightarrow E_i \cap E_j = \phi$ for all i, j

S



Law of Total Probability

- If A is an event and $E_1, E_2, \ldots, E_k$ is a **partition** of S, then

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \ldots + P(A \cap E_k)$$
$$= \sum_{i=1}^{k} P(A \cap E_i)$$

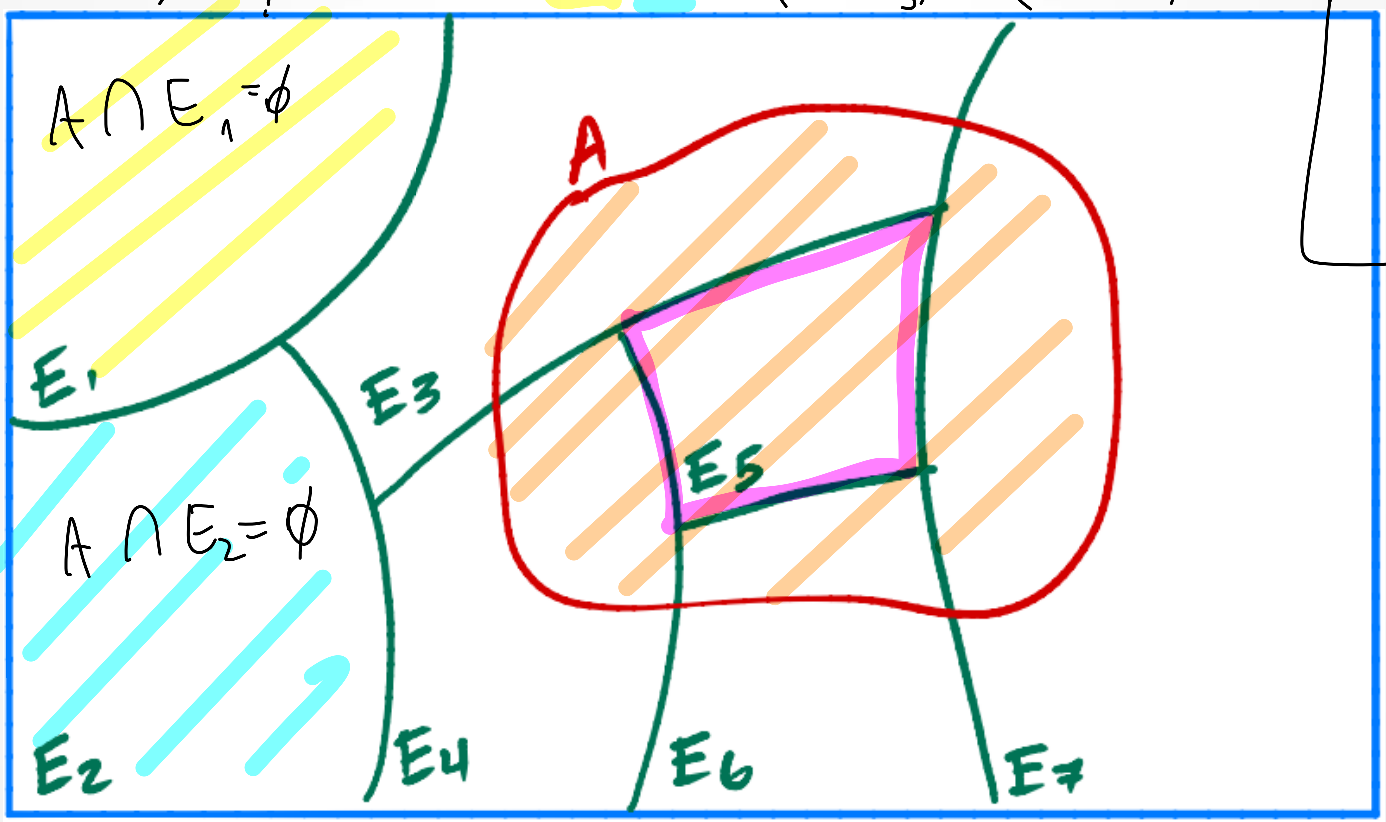
Law of Total Probability

$A \subset S$

partition

$E_1, \ldots, E_k$

$$P(A) = \sum_i P(A \cap E_i) = 0 + 0 + P(A \cap E_3) + P(A \cap E_4) + P(A \cap E_5) + P(A \cap E_6) + P(A \cap E_7)$$



$E_5 \subset A$

$P(A \cap E_5) = P(E_5)$

Law of Total Probability

- If A is an event and $E_1, E_2, \ldots, E_k$ is a **partition** of S, then

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \ldots + P(A \cap E_k)$$

joint probabilites

$$= \sum_{i=1}^{k} P(A \cap E_i)$$

multiplication rule

joint prob $\Rightarrow$ cond. prob.

- Written another way,

$$P(A) = P(A \mid E_1) \cdot P(E_1) + \ldots + P(A \mid E_k) \cdot P(E_k)$$

$$= \sum_{i=1}^{k} P(A \mid E_i) \cdot P(E_i)$$

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

Suppose someone is late. What is the probability that they walked? Choose the best answer.

A. Close to 5%
B. Close to 15%
C. Close to 30%
D. Close to 40%

$P(\text{late}|\text{walk}) = 20\%$

$$P(\text{walk}|\text{late}) = \frac{P(\text{walk} \,\&\, \text{late})}{P(\text{late})} = \frac{6\%}{45\%} \approx 13\%$$

$P(A|B) \neq P(B|A)$

Getting to Campus

$\& = AND = \cap$

- Suppose all you know is
  - P(Late) = 45%
  - P(Walk) = 30%
  - P(Late|Walk) = 20%
- Can you still find P(Walk|Late)?

$$P(walk|late) = \frac{P(Walk \,\&\, Late)}{P(Late)} = \frac{P(late|Walk) \cdot P(walk)}{P(late)} = \frac{0.2 \cdot 0.3}{0.45} \approx 13\%$$

Bayes' Theorem

Bayes' Theorem follows from the multiplication rule, or conditional probability.

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

Bayes' Theorem:

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

Bayes' Theorem

Bayes' Theorem follows from the multiplication rule, or conditional probability.

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

$$P(A) = \sum_i P(A|E_i)P(E_i)$$

$$= P(A|B)P(B) + P(A|\bar{B}) \cdot P(\bar{B})$$

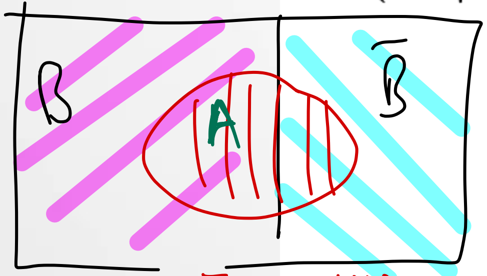
Bayes' Theorem:

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

$$= \frac{\boxed{P(A|B) * P(B)}}{\boxed{P(B) * P(A|B)} + \boxed{P(\bar{B}) * P(A|\bar{B})}}$$

not B

(complement of B)



B, $\bar{B}$ partition