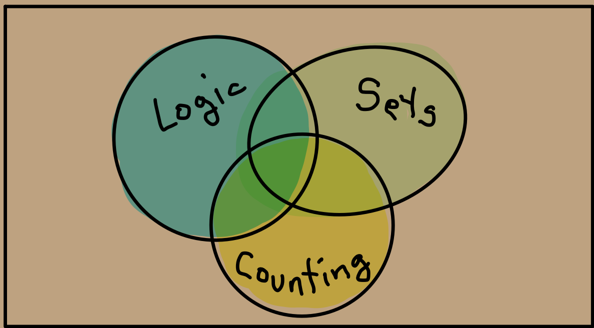
Set Theory



John Rachlin
Northeastern University

Lecture Notes

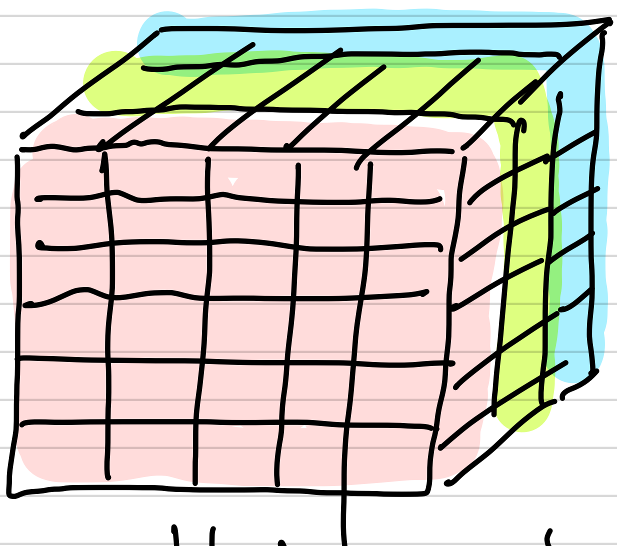
In computer science, we are interested in collections of data. These collections can take many forms:

Lists: [1, 2, 3]

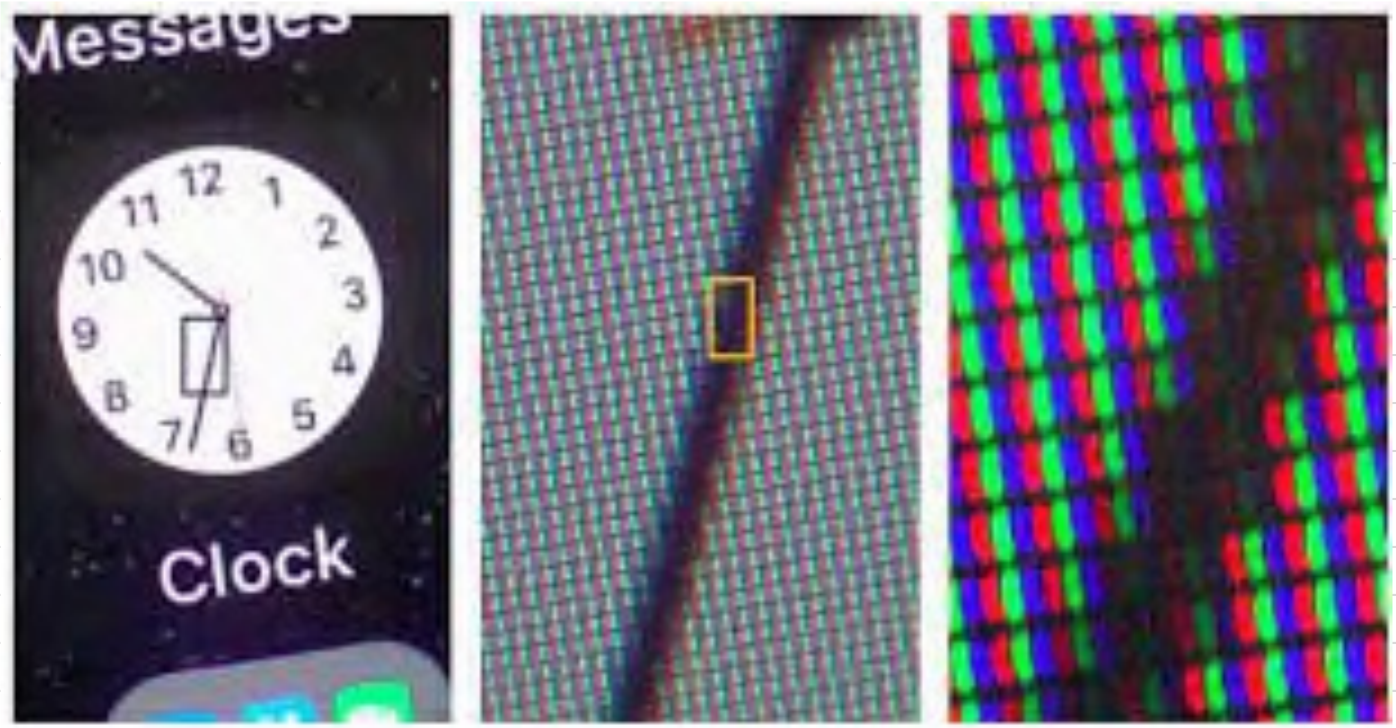
["hello", 42, ['a', 'b', 'c']]

a list w/in a list!

Arrays:



An image might be stored as a 3 dimensional array height x width x rgb



Here is an iphone display under the microscope. Notice how each pixel is rendered as an intensity value for red, green, and blue.

Databases are centralized collections of data. The layout of data and query languages (SQL) use to ask questions of the data are built around the concepts of set theory.*

* Im referring to so-called "Relational DBs"

In mathematics we want to reason about numbers and their properties.

"n^2 is an even # if and only if n is even"

A bit more formally:

$$\forall n \in \mathbb{Z} : n^2 \text{ is even} \longleftrightarrow n \text{ is even}$$

or with predicates:

$$\forall n \in \mathbb{Z} : \text{even}(n^2) \longleftrightarrow \text{even}(n)$$

Proof

| Direct Proof of $p \to q$
| Assume p
| Show q follows

Assume x is an <u>arbitrary</u> even #.

Then $x = 2k$ for some $k \in \mathbb{Z}$

So $x^2 = (2k)^2 = 4k^2 = 2 \cdot 2k^2$

$= 2m$

$(m = 2k^2 \in \mathbb{Z})$

$\therefore$ x^2 is even.

But $p \leftrightarrow q \equiv p \to q \wedge \neg p \to \neg q$.

Assume x be odd

$x = 2k + 1$ $k \in \mathbb{Z}$

$x^2 = (2k+1)^2$

$= 4k^2 + 4k + 1$

$= 2(2k^2 + 2k) + 1$

$= 2m + 1$ $m = 2k^2 + 2k \in \mathbb{Z}$

$\therefore$ x^2 is odd.

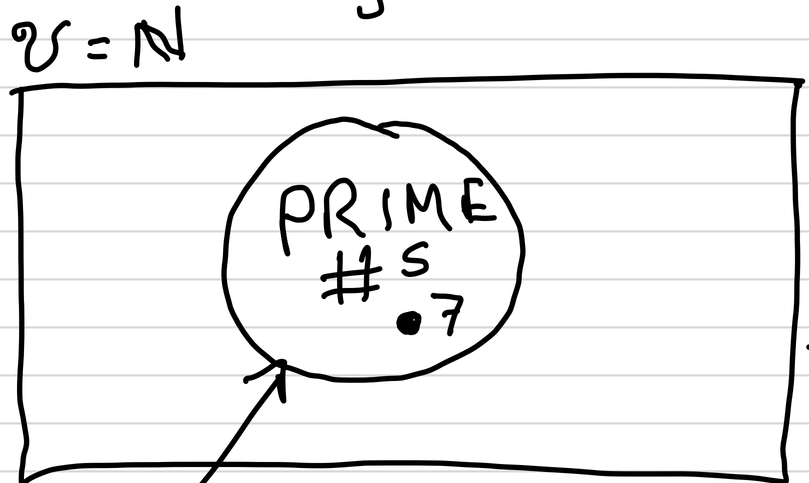
So math involves reasoning about properties (predicates)

collections (sets)

- odd, even
- integers, rational, real
- prime

Graphically I might represent membership or non-membership with a diagram.

VENN DIAGRAM



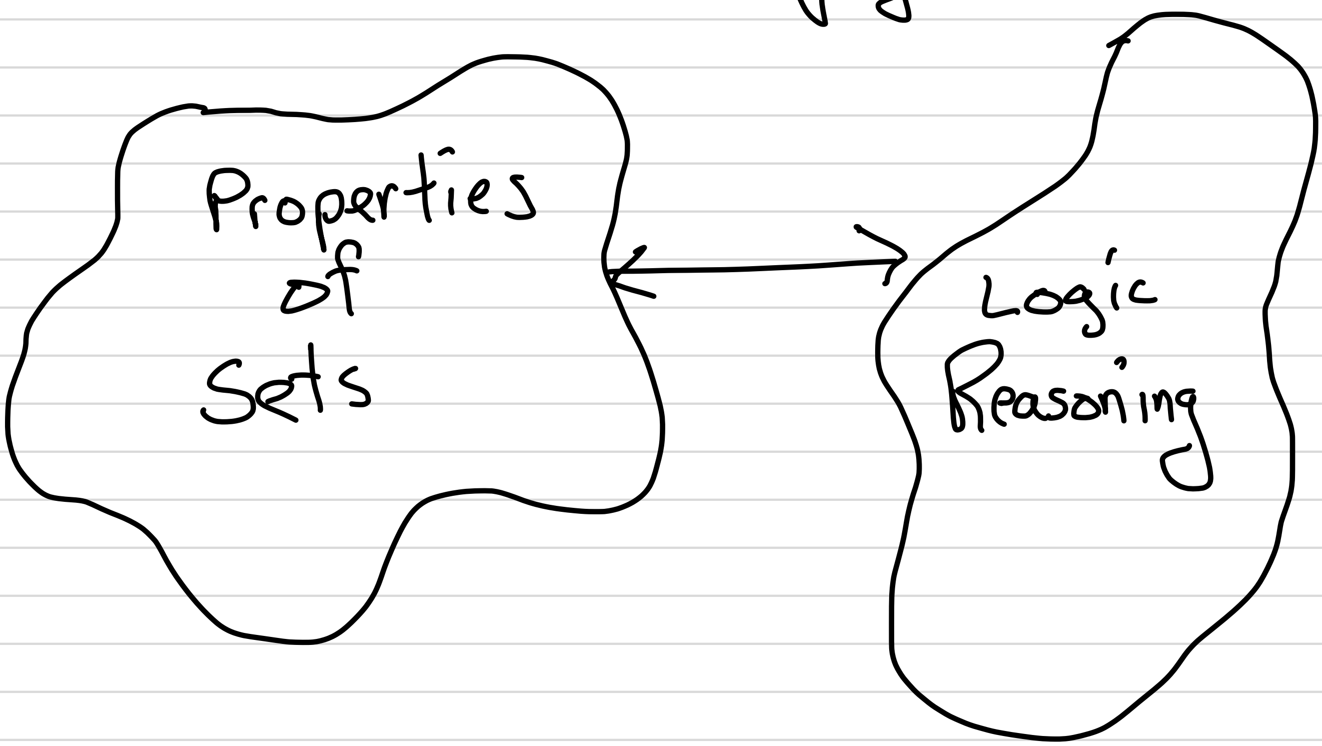
my "domain" e.g $\mathbb{N}$ 0, 1, 2, ... The "Universe"

This circle encompasses the natural #s that are prime: 2, 3, 5, 7, 11, . . .

Prime(x) neither T or F

Prime(7): True Prime(12): False.

So now we begin to see a connection between logical predicates and sets: Predicates define membership in a set. This suggest an underlying deep connection:

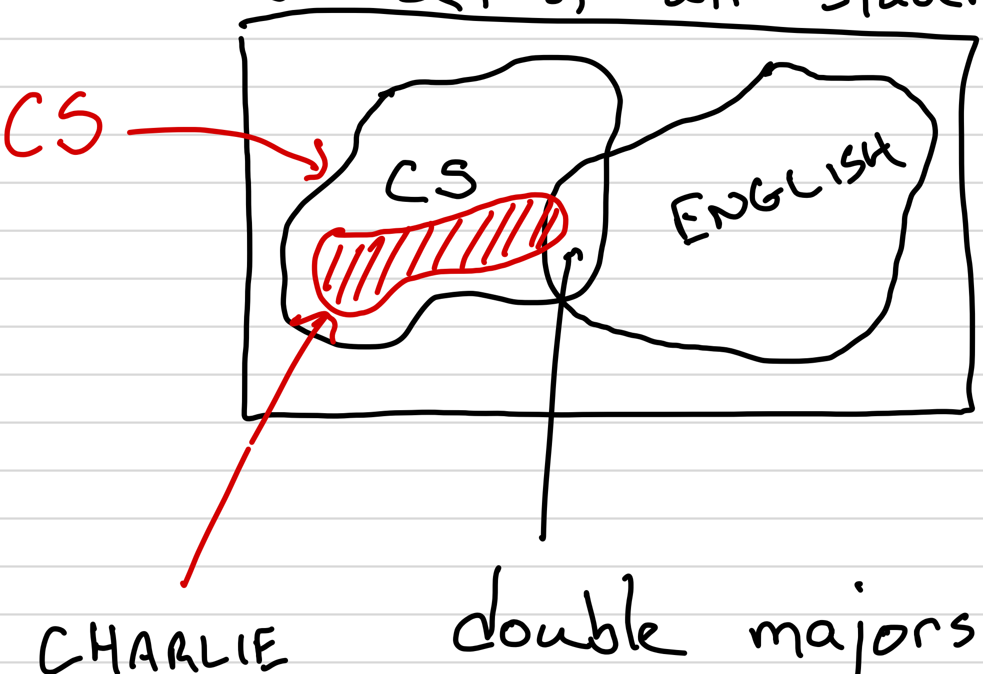


Consider:

A) Charlie knows only CS students.

In pictures.

U = set of all students (domain)



double majors are both CS students and English students so its ok (valid) if charlie knows some double majors.

The diagram makes the solution more obvious:

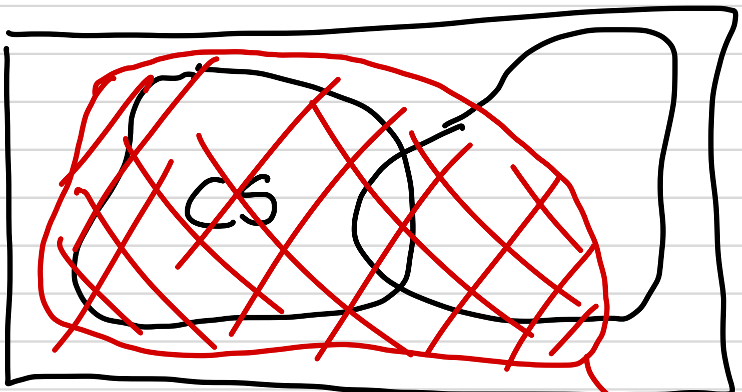
$$\forall x \; Charlie(x) \rightarrow CS(x)$$

Or by the contrapositive:

$$\forall x \; \neg CS(x) \rightarrow \neg Charlie(x)$$

What about:

B) Charlie knows <u>all</u> CS students.



i.e., if you're a CS student, then charlie knows you!

$$\forall x \; CS(x) \rightarrow Charlie(x)$$

CHARLIE

Note:

A predicate such as Charlie(x) is <u>not</u> a set. But the values of x that make the predicate true are a set

We could say:

$$CHARLIE = \{ x \in STUDENT \mid Charlie(x) \text{ is true} \}$$

Definition: Set

A <u>set</u> is an unordered collection of distinct objects.

$S_1 = \{0000, 0001, \dots 9999\}$ 4 digit pins

$S_2 = \{\text{red}, \text{blue}, \text{green}, \text{yellow}\}$

$= \{\text{blue}, \text{green}, \text{yellow}, \text{red}\}$

<u>Common sets</u>

$\mathbb{N} = \{0, 1, 2, 3, \dots\}$ natural #s

$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ integers

$\mathbb{Z}^+ = \{1, 2, 3, \dots\}$ positive #s

$\mathbb{Q} = \{p/q \mid p \in \mathbb{Z}, q \in \mathbb{Z}, q \neq 0\}$ rationals

where element of

$\mathbb{R}$ = real #s

$\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$ integers mod n

$\emptyset = \{\}$ The empty set

- Set builder Notation.

$$S = \{ v \mid \text{condition on } v \}$$

e.g.

$$S = \{ x \mid x \in \mathbb{Z},\ |x| < 5 \}$$
$$= \{ -4, -3, 2, -1, 0, 1, 2, 3, 4 \}$$

(Its easier to just define the rules rather than enumerating all the elements).

- Cardinality : The size of the set

$|S|$ = "size" of the set
= # of elements.

may be finite or infinite.

- membership.

A set is a collection of elements.

In the above example,

$3 \in S$ (3 is an element of S)

$7 \notin S$ (7 is not an element of S)

- Venn Diagram.

U = universe of possible elements:

$\{1, 2, 3, \ldots, 10\}$.

OR $\{x \in \mathbb{Z} \mid x = 2k \text{ where } k \in \mathbb{Z}\}$

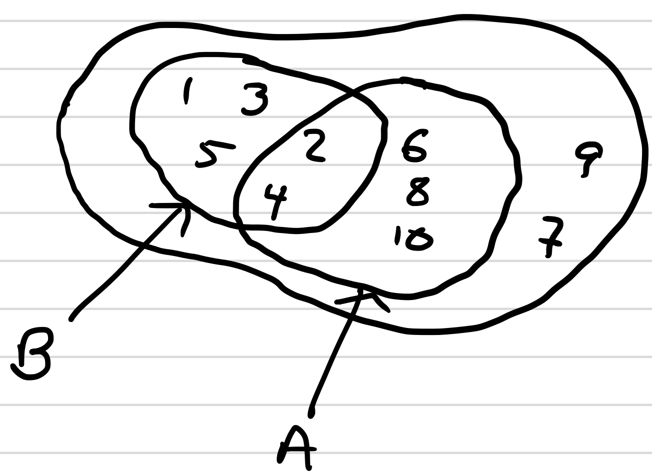
A = "evens" $= \{x \mid x \in U,\ x = \text{even}\}$

$= \{2, 4, 6, 8, 10\}$

B = "≤ 5" $= \{x \mid x \in U,\ x \le 5\}$

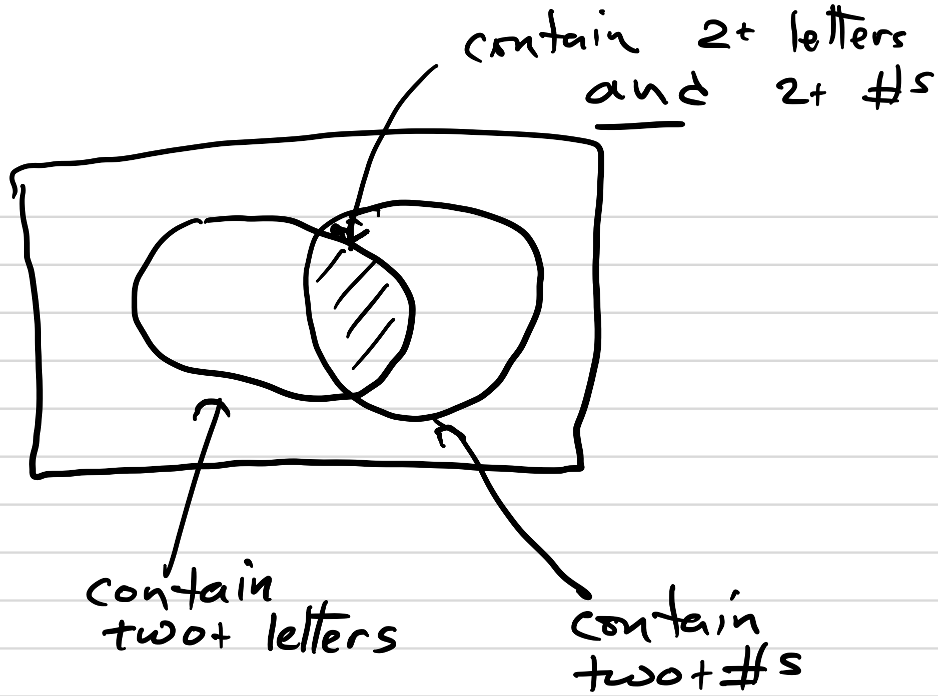
$= \{1, 2, 3, 4, 5\}$.

U:



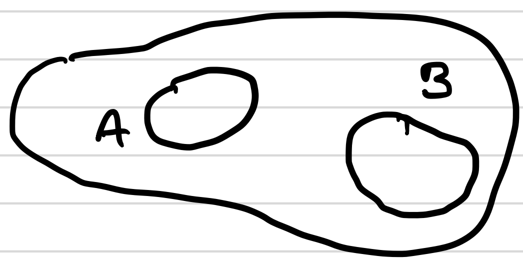
We want to be able to reason about the sizes of these chunks.

Passwords:



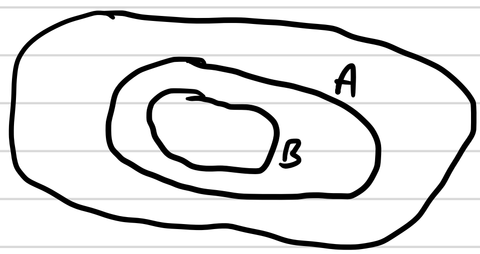
7 Definitions

a) A and B are **disjoint** if they share no elements in common. (An empty overlap)



b) B is a **subset** of A if every element of B is an element of A

$B \subseteq A$



$\therefore A \subseteq A$

c) **Proper Set**: $B \subset A$: B is a subset of A and A contains other elements.

⑧ Set Operations

a) Union : $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

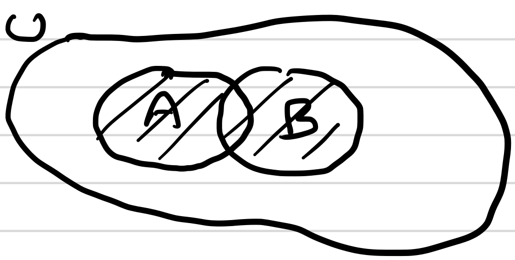
$U = \{1 \ldots\ldots 10\}$

$A = \{2, 4, 6, 8, 10\}$

$B = \{1, 2, 3, 4, 5\}$

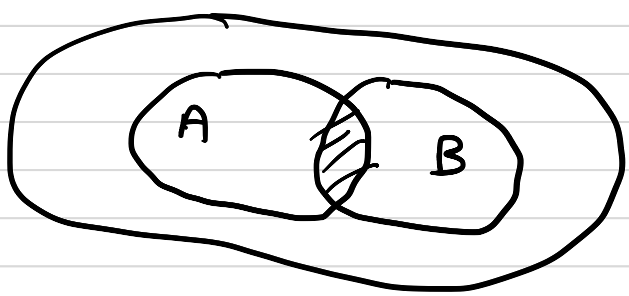
$A \cup B = \{1, 2, 3, 4, 5, 6, 8, 10\}$

⇑ no repetitions



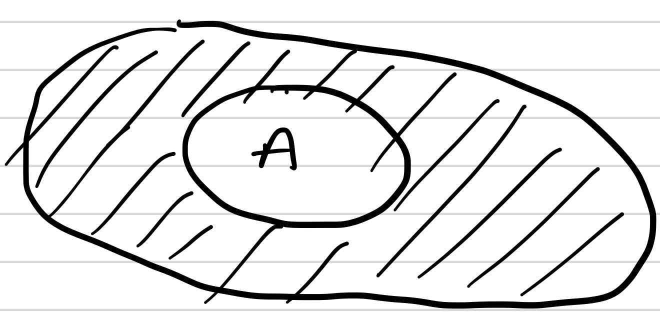
b) Intersection : $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

$= \{2, 4\}$



c) Complement : $\overline{A} = \{x \mid x \in U \text{ and } x \notin A\}$

$= \{1, 3, 5, 7, 9\}$

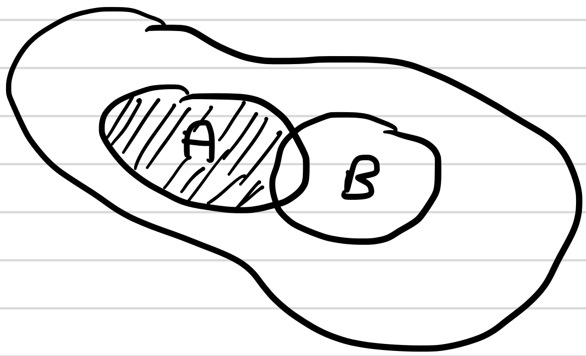


$\overline{B} = \{6, 7, 8, 9, 10\}$

d) Set difference

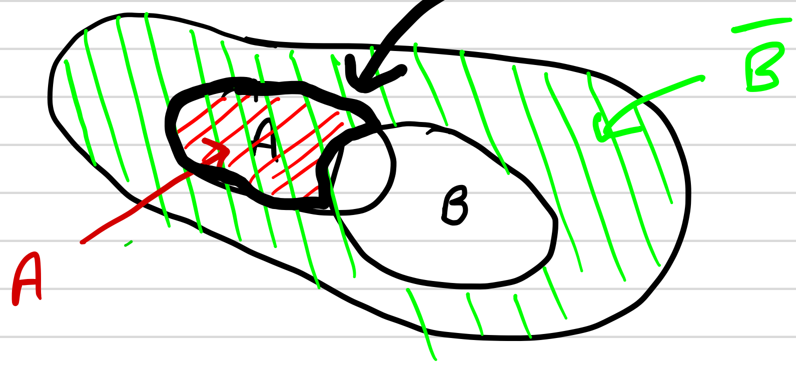
$$A - B = \{x \mid x \in A \text{ and } x \notin B\}$$

$$= A \setminus B$$



$$= \{6, 8, 10\}$$

$$= A \cap \overline{B}$$



e) Symmetric difference

: element occurs in exactly 1 set

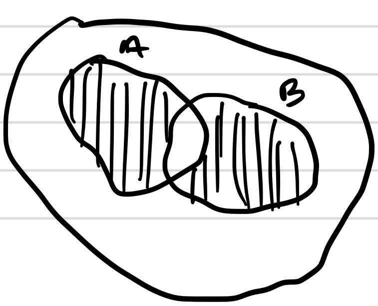
$$A \triangle B = \{x \mid x \in A \text{ and } x \notin B \text{ or } x \in B \text{ and } x \notin A\}$$

$$= (A - B) \cup (B - A)$$

$$= (A \cap \overline{B}) \cup (B \cap \overline{A})$$

$$= (A \cup B) - (A \cap B)$$

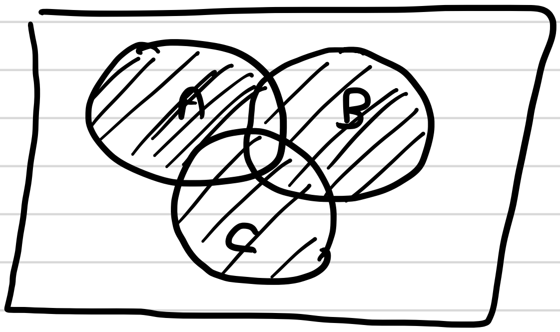
$$= (A \cup B) \cap (\overline{A \cap B})$$



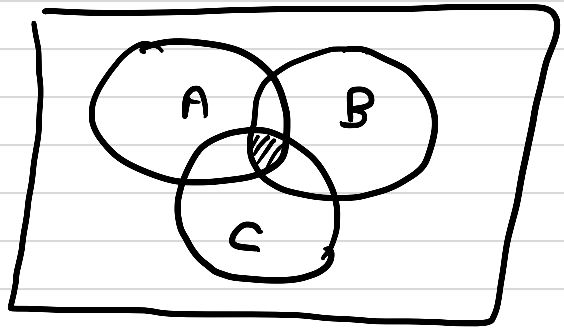
$$\Downarrow$$

$$\{1, 3, 5, 6, 8, 10\}$$

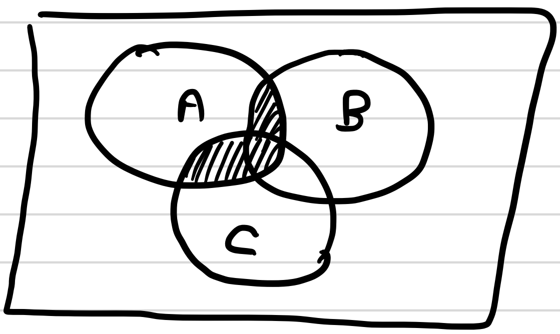
Set operations on multiple sets



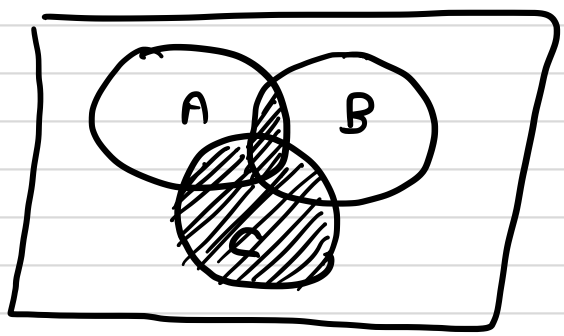
$A \cup B \cup C$



$A \cap B \cap C$



$A \cap (B \cup C)$



$(A \cap B) \cup C$

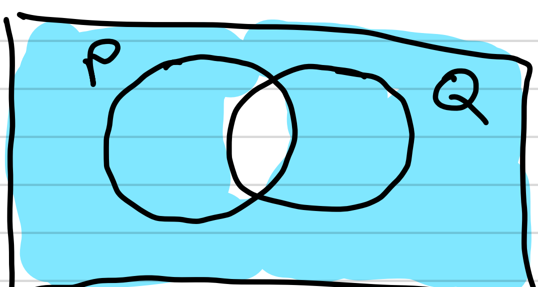
Set Equivalence $\equiv$ Logical Equivalence

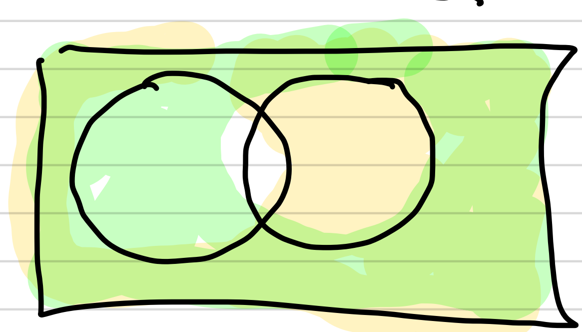
$\cup$	$\equiv$	$\vee$
$\cap$	$\equiv$	$\wedge$
$\overline{A}$	$\equiv$	$\neg a$
$A \subseteq B$	$\equiv$	$a \rightarrow b$
$A = B$	$\equiv$	$a \leftrightarrow b$ or $a \equiv b$
$\mathcal{U}$	$\equiv$	T
$\emptyset$	$\equiv$	F

And all the usual laws of logical equivalence apply!

De Morgans

$$\neg(P \wedge Q) \equiv \neg P \vee \neg Q$$

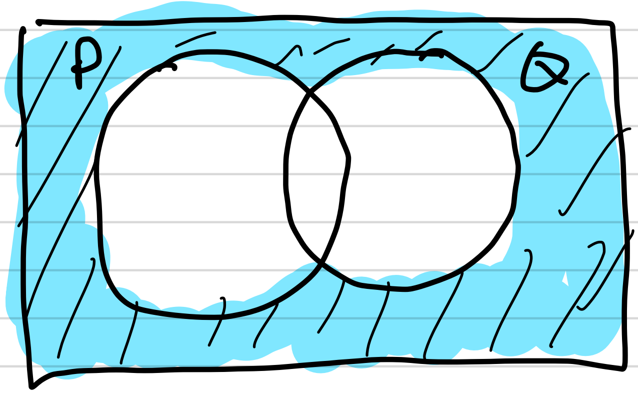




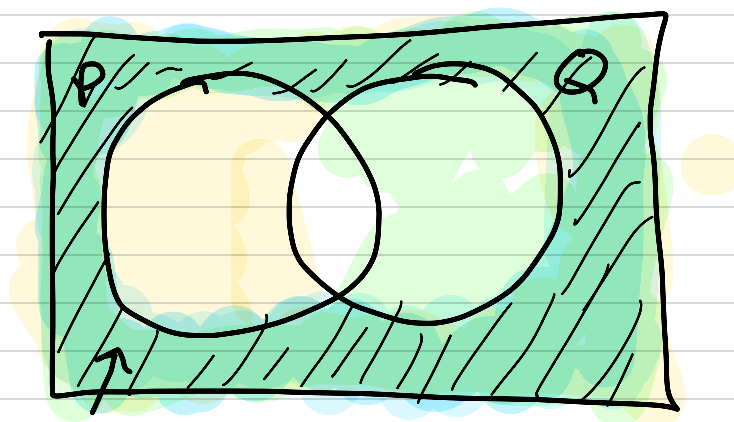
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

De Morgans - Continued

$$\neg(P \vee Q) \equiv \neg P \wedge \neg Q$$



$\equiv$



$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

Also:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$A \cap \mathcal{U} = A \qquad A \cup \mathcal{U} = \mathcal{U}$

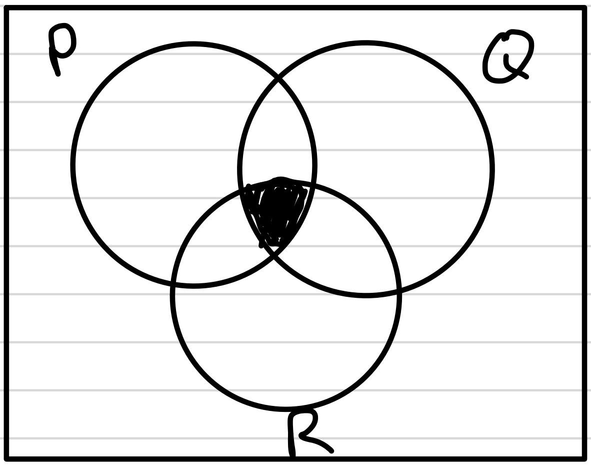
$A \cap \emptyset = \emptyset \qquad A \cup \emptyset = A$

$A \cup B = B \cup A \qquad A \cap B = B \cap A$

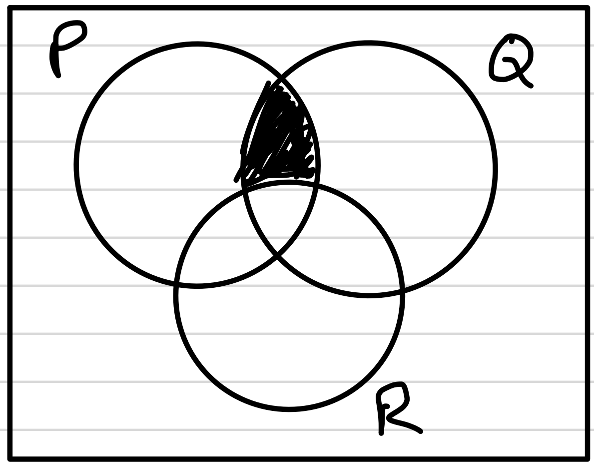
$A \cap A = A \qquad A \cup A = A$

Does "Rachlin's" Law work??

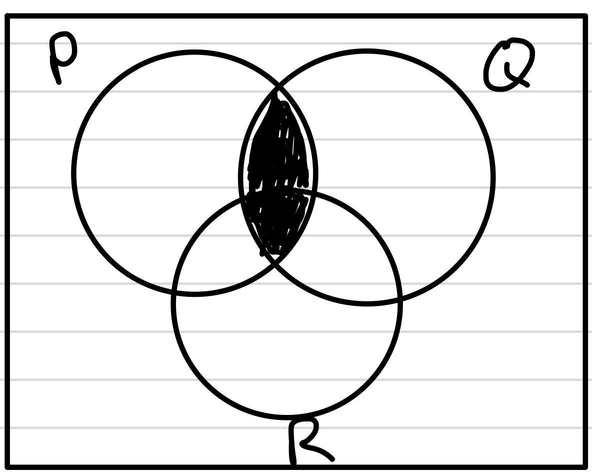
$$(p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r) = p \wedge q$$



$\cup$



$$(P \cap Q \cap R) \cup (P \cap Q \cap \overline{R}) =$$



$$P \cap Q$$

Yes!

Power Sets *

f) $\mathcal{P}(A)$ = set of all subsets of A

e.g., $A = \{a, b, c\}$

$$\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}$$

Can we generate this systematically?

$$|\mathcal{P}(A)| = 2^{|A|}$$

We enumerated all the
0 - element subsets
1 - element subsets
2 - element subsets
3 - element subsets.

We associate each subset with a bit string:

$\{\emptyset$	$\{a\}$	$\{b\}$	...	$\{a,c\}$	....	$\{a,b,c\}\}$
000	100	010		101		111

So clearly $|\mathcal{P}(A)| = 2^{|A|}$

What is $\mathcal{P}(\emptyset)$?

$$\mathcal{P}(\emptyset) = \{\emptyset\}$$

$|\emptyset| = 0$ $\quad |\mathcal{P}(A)| = 1 = 2^{|\emptyset|} = 2^0 = 1$ ✓

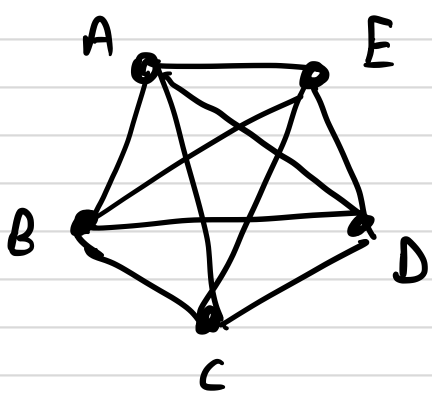
Remember: A Powerset always contains sets as elements.

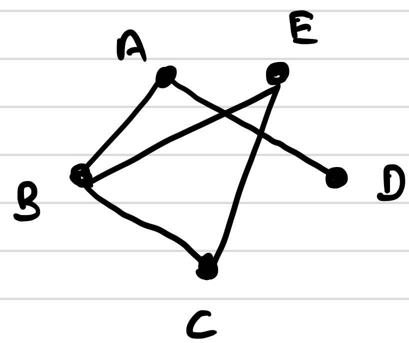
OPTIONAL

How many graphs are there on the vertices $V = \{A, B, C, D, E\}$

$|V| = 5$

$$|E| = \frac{|V|(|V|-1)}{2} = 10$$





	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
Complete graph :	1	1	1	1	1	1	1	1	1	1
Other Graph	1	0	1	0	1	0	1	0	1	0

For each pair (u, v) we decide whether to include an edge, so we have

$$\underbrace{2 \cdot 2 \cdot 2 \cdots\cdots \cdot 2}_{\text{choices}} = 2^{10} = 1024$$

g) Cartesian Product.

Back in HS you learned to plot things in the Cartesian plane

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x,y) \mid x \in \mathbb{R} \quad y \in \mathbb{R}\}$$

$$A \times B = \{(x,y) \mid x \in A \text{ and } y \in B\}$$

e.g. $A = \{1, 2, 3\}$

$B = \{a, b\}$

$$A \times B = \{(1,a), (1,b), (2,a), (2,b), (3,a), (3,b)\}$$

Note that $(1,a) \neq (a,1)$. Order matters.

Similarly,

$$B \times A = \{(a,1), (a,2), (a,3), (b,1), (b,2), (b,3)\}$$

$$\therefore \quad \boxed{A \times B \neq B \times A}$$

b	(1,b)	(2,b)	(3,b)
a	(1,a)	(2,a)	(3,a)
	1	2	3

$$|A \times B| = |A| \cdot |B|$$

Generalize to multiple sets:

$$A \times B \times C = \{(x, y, z) \mid x \in A, y \in B, z \in C\}$$

$$|A \times B \times C| = |A| \cdot |B| \cdot |C|$$

We might have: $A_1 \times A_2 \times \cdots \times A_n$

$$\Downarrow$$

$$(a_1, a_2, \ldots a_n)$$

"n-tuple" - an ordered collection of n elements.

OPTIONAL

A <u>Relation</u> is a subset of a cartesian product.

$$R \subseteq A \times B$$

For example: $ID = \{x \mid x \in \mathbb{Z}^+\} = \{1, 2, 3, \ldots\}$

NAME = {name | name is a character sequence with 1..20 characters}.

{a, b, c, ..., aa, ab, ... joe ... , ... mary ..... }

DOB = {dob | dob is a valid date}

= {2001-Feb-1, 2008-Dec-25, ... }

So $ID \times NAME \times DOB = \{ (a_1, b_1, c_1), (a_2, b_2, c_2), \vdots (a_n, b_n, c_n) \}$

A Possible Relation: on $ID \times NAME \times DOB$

$\{ (1, joe, 1995\text{-}Mar\text{-}25), (2, jack, 2003\text{-}Oct\text{-}8), (3, abby, 2006\text{-}Apr\text{-}22) \}$

$\Downarrow$

ID	Name	Dob
1	joe	1995-Mar-25
2	jack	2003-Oct-8
3	abby	2006-Apr-22
4		

Relation = Table

Tuple $\Rightarrow$ Record.

OPTIONAL

9 Representation of Sets in Computers.

$U = \{1, 2, 3, \ldots 10\}$

$A = \{2, 4, 6, 8\}$ evens

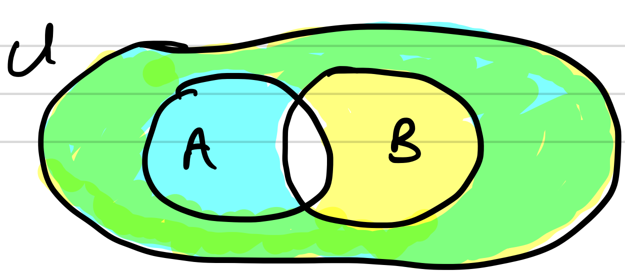
$B = \{1, 2, 3, 4, 5\}$.

We could have a bit string of length equal to the size of the universe.

U	1	2	3	4	5	6	7	8	9	10
A	0	1	0	1	0	1	0	1	0	1
B	1	1	1	1	1	0	0	0	0	0
$A \cup B$ Bitwise OR	1	1	1	1	1	1	0	1	0	1
$A \cap B$ Bitwise AND	0	1	0	1	0	0	0	0	0	0
$A \Delta B$ Bitwise XOR	1	0	1	0	1	1	0	1	0	1
$\overline{B}$ Bitwise NOT	0	0	0	0	0	1	1	1	1	1

$\overline{A \cap B}$ analogous to $\neg(A \wedge B) = \neg A \vee \neg B$

$= \overline{A} \cup \overline{B}$



OPTIONAL

Sets & Counting / Combinatorics.

① Intro

There are lots of things in CS we want to count. E.g. steps in an algorithm - is one algorithm for sorting better than another.

A motivating example: The security of passwords. How many passwords are their given different rules.

An ATM card might have 4-digits. How many pins? 10,000 = 10^4.

Each digit has 10 possibilities and there are 4 digits.

4-digits: 10,000 = 10^4

4 - lower case: 26^4 = 456,976

4 - upper/lower 52^4 = 7,311,616

More passwords ⇒ more security. There is a <u>reason</u> why you can't have a 2-digit pin!!

"Brute force" attacks (attempting all possibilities) should be hard.

NSF Rules:

7-10 characters
uppercase, lowercase, digits
≥ 2 letters
≥ 2 digits

less obvious how to count the size of this search space.

The restrictions <u>decrease</u> size of the PW space.

Why? To combat human nature: short passwords are easy to remember

(10) Sum and Product Rule

Pants or Shorts 8 pants, 6 pairs of shorts

I can wear one or another. so I have $8+6=14$ options.

Pants and Shirts 8 pants, 12 shirts

$8 \times 12 \Rightarrow 96$ options.

A) **Product Rule:** If A & B are finite sets then the number of ways to pick an $a \in A$ and an element $b \in B$ is:

$$|A \times B| = |A| \times |B|$$

more generally, $A_1, A_2, \ldots A_n$

$$|A_1 \times A_2 \times \cdots A_n| = |A_1| \times |A_2| \times \cdots \times |A_n|$$

4 char passwords: upper/lower/digit.
How many pw. do you have?

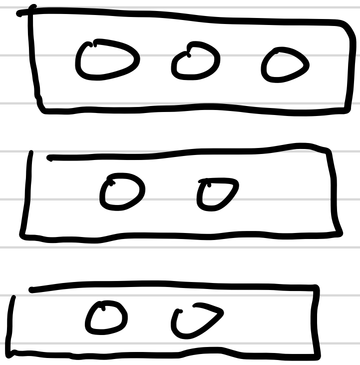
$A = \{a, b, c, \ldots z, A, B, C, \ldots Z, 0 \ldots 9\}$

$A \times A \times A \times A$ $\qquad |A| = 62$

$62 \times 62 \times 62 \times 62 = 62^4 = 14,776,336$

"For each position I have 62 choices"
62 choices for position 1 <u>AND</u> position 2 <u>AND</u> ....

Pick Representative from each of three tables



$$3 \times 2 \times 2 = 12 \text{ possibilities}$$

Think about the problem in English:

"AND" $\Rightarrow$ multiplication / product

"OR" $\Rightarrow$ addition / sum.

B) <u>Sum Rule</u> "Pick this or that"

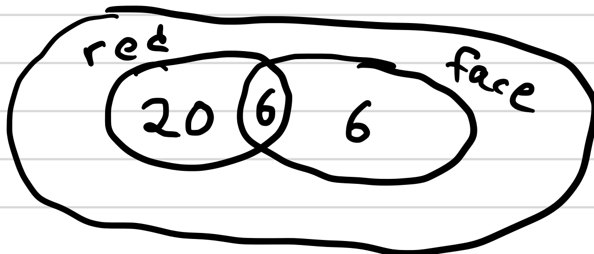
If A & B are <u>disjoint</u> finite sets, then the # ways of picking an object from A <u>OR</u> B is $|A| + |B|$

Generally $A_1, A_2, \ldots A_n$ all <u>mutually disjoint</u> then

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = |A_1| + |A_2| + \cdots + |A_n|$$

c) <u>Principle of Inclusion/Exclusion:</u>

What if the sets aren't disjoint?



deck of 52 cards

R : set of red cards = 26
F : set of face cards = 12

How many ways can I pick a card that is red <u>or</u> a face card?

<u>NOT</u> $26 + 12 = 38$ ✗ (I'm double counting red face cards)

Rather: $26 + 12 - 6 = 32$ ✓

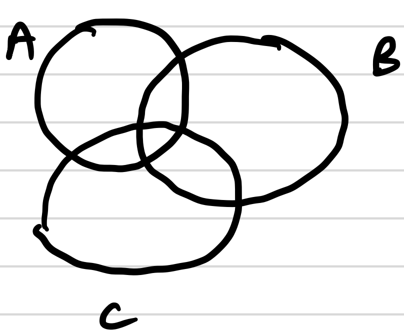
Principle of Inclusion/Exclusion says:

2 sets A, B

$$|A \cup B| = |A| + |B| - |A \cap B|$$

It still works when the sets are disjoint because then $|A \cap B| = 0$

<u>Three Sets</u>



"Pick an element from A or B or C"

$$\boxed{\begin{aligned}|A \cup B \cup C| &= |A| + |B| + |C| \\ &- |A \cap B| - |A \cap C| - |B \cap C| \\ &+ |A \cap B \cap C|\end{aligned}}$$

For larger sets, its a continuing series of inclusions & exclusions

Examples

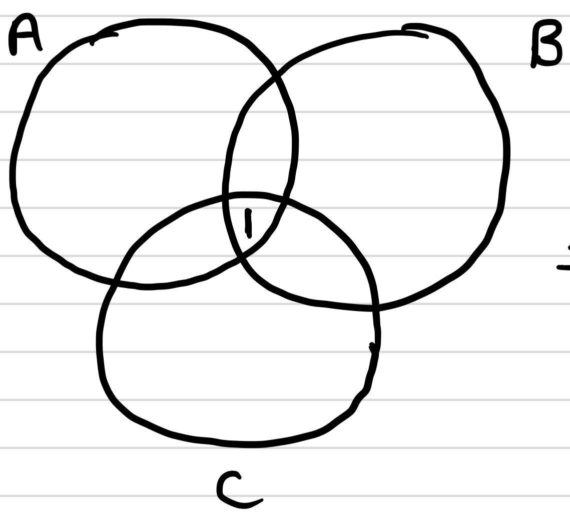
41 students total

Three courses: Algebra, Biology, Chemistry

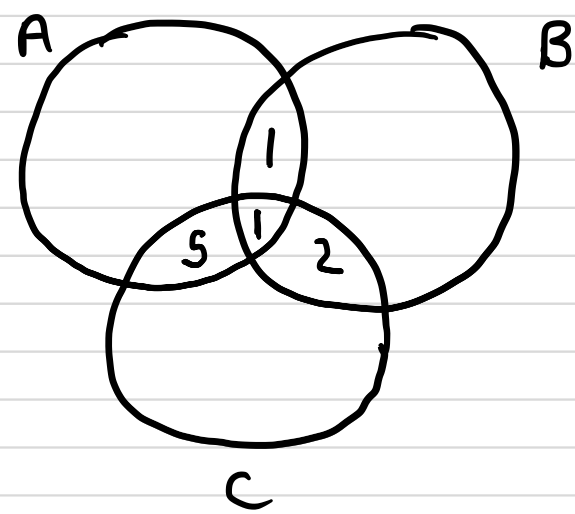
Subject	A	B	C	AB	AC	BC	ABC
Failed	12	5	8	2	6	3	1

How many students failed at least one course?

How many passed all three courses?



$\Longrightarrow$



$\Downarrow$

15 failed 1 or more

26 failed 0

41 students.