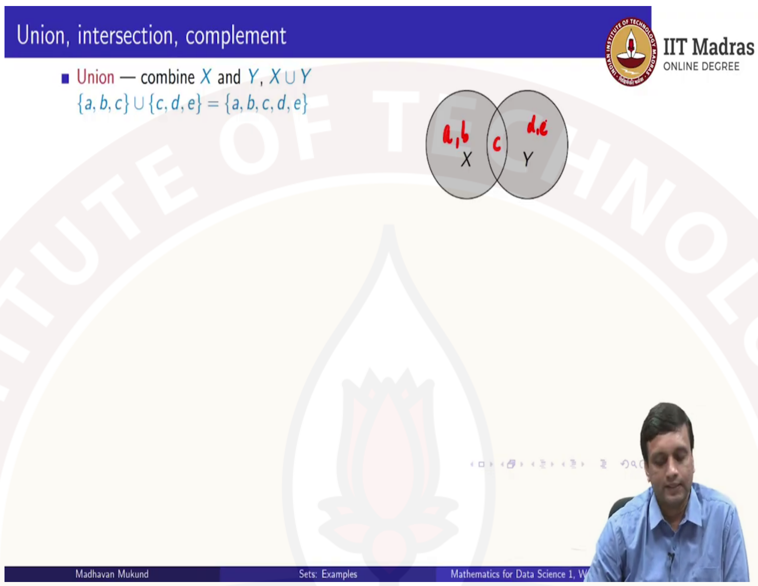
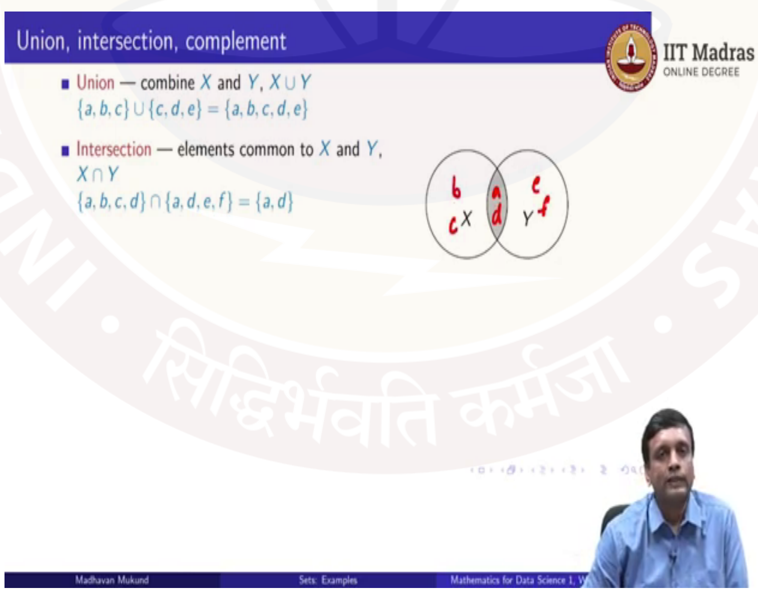
IIT Madras

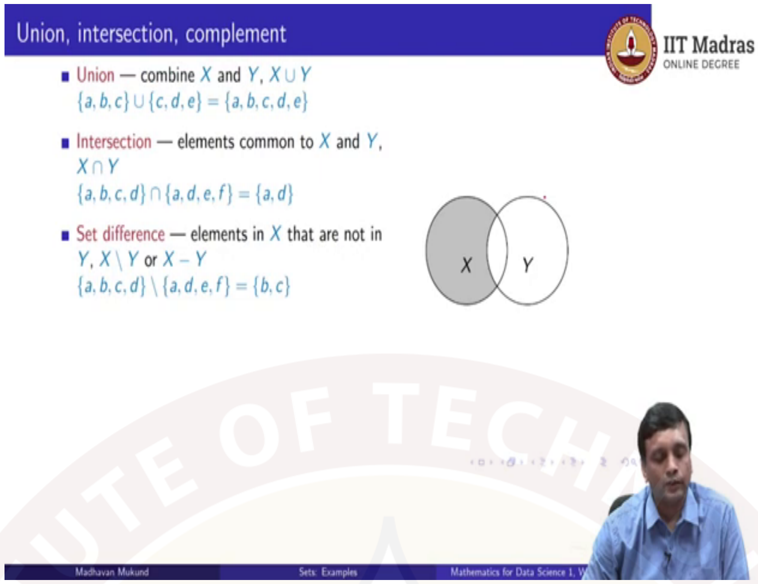
ONLINE DEGREE

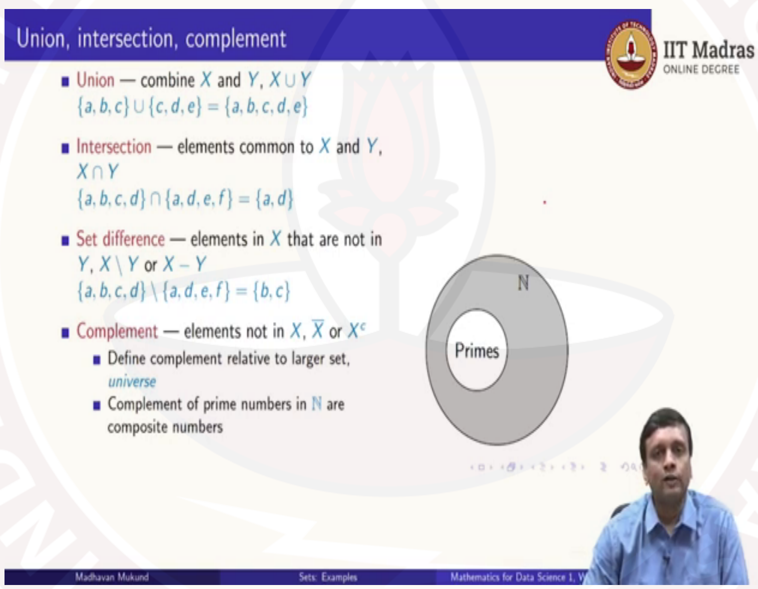
Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute
Lecture- 5B
Examples of Set Operations and Counting Problems

(Refer Slide Time: 00:15)









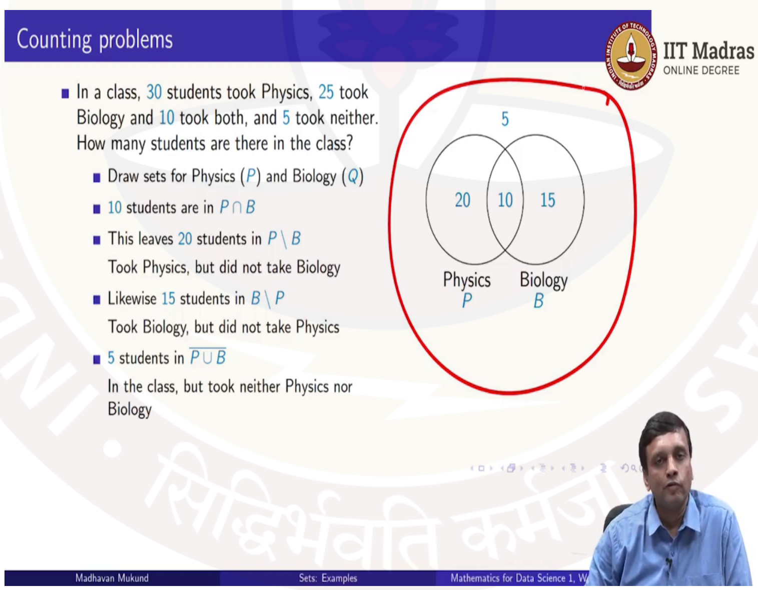
So, the other operations that we saw on sets are union, intersection, and complement, which we represented using Venn diagrams as shown here. So, the union takes two sets and combines them and removes the duplicates. So, the overlapping part between the two diagrams represents the common element. So, in this case, we would have this common element c over here, and then we have had a and b over here, and we would have d and e over here because d and e belongs only to Y, a, b belongs only to X.

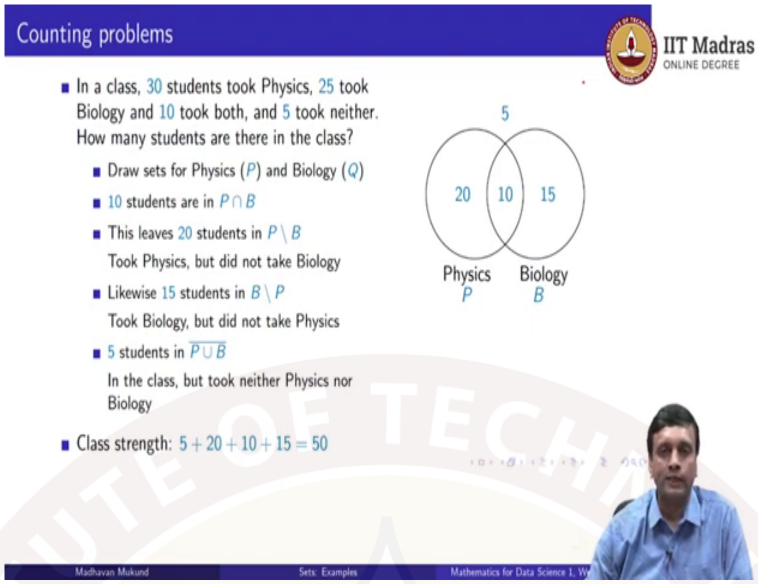
Conversely, we can take only those things which are common to the two and in this case, we have a and d over here, and then we know that b and c are only on the left and e and f are only on the right. So, the intersection tells us the elements which are common to the two sets. Set difference tells us what is on the left but not on the right.

And finally, the complement can be taken if we have an overall universe that is a full set to talk about. And with respect to that set, we can ask which elements are not in the set that we are looking at. So for instance, if we are looking at the natural numbers as a whole, the primes are a subset of the natural numbers, the complement of the primes are all those natural numbers that are not primes.

Now, remember that the complement matters, because if we take the complement of the primes, for example, with respect to the real numbers, we will get all sorts of other numbers which are not even integers. So, whenever we define the complement, we need to define the universe that we are talking about.

(Refer Slide Time: 01:37)





So, this leads us to a class of problems that you might come across, which can be solved nicely using these Venn diagrams. So, these Venn diagrams are not just pretty pictures, they are actually useful ways to reason about these problems. So, here is a typical problem that you could come across. So, you have a class in which 30 students have taken physics, and 25 students have taken biology, but 10 have actually taken both physics and biology, but there are also 5 who have taken neither of these two subjects.

So, these are the facts that are given to you. There are 30 students taking physics, 25 taken biology, 10 take both, 5 take neither, the question is how many students are there in the class. So, using Venn diagram notation, you can represent the fact that there are two sets of students, those who take physics and those who take biology by representing them by two sets, say P and Q. And we know that some take both, so, there is an intersection so these two sets overlap.

Now, from the data that we are given, we know that the overlap has 10 students, so we can write a number 10 in the intersection to indicate that there are 10 students who take physics and take biology. Now, we know that 30 students took physics overall and we have already accounted for 10 of them because they have all taken both physics and biology. So, there are 20 students who have taken physics, but have not taken biology.

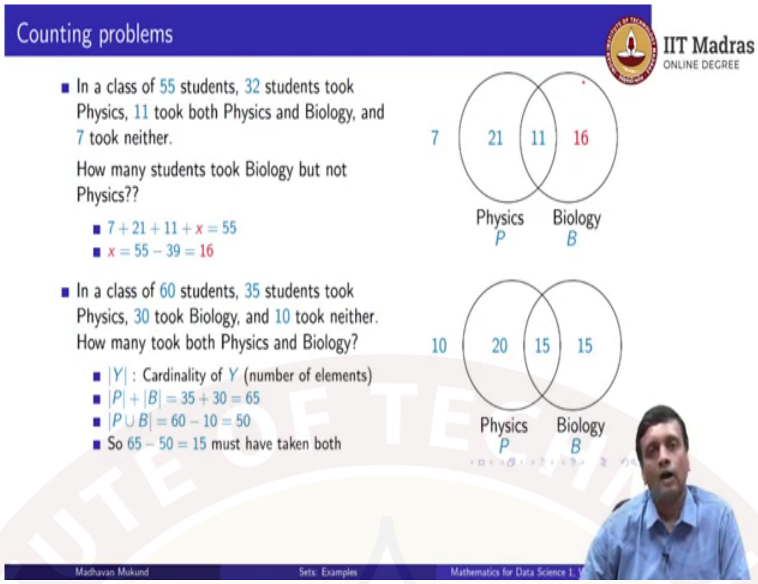
So, this in our set notation is the set difference, it is the difference between P and B, how many elements are in P which are not in B, how many students have taken physics who have not taken biology. And we have a symmetric thing on the right hand side. So, we know that there are 10 students who have taken both but 25 students take biology. So, there must be 15 students who are in $B \setminus P$, these are students who took biology and did not take physics.

So, in this way, we can populate the three regions of the Venn diagram with numbers indicating how many students are in each of these regions at 10 in the intersection, 20 on the left hand side, 15 on the right hand side. But, this is not the entire class because with respect to the entire class we have to take the number who are in the complement, those who have taken neither physics nor biology, and these are 5 students who are outside $P \cup B$.

Now, technically one should draw outside this the complement to indicate the entire class but just for convenience, I have not done that, but this entire complement outside this contains 5 elements. So, totally from this, we can see that there are 4 regions of interest. We have the $P \setminus B$ region physics but not biology, we have the $B \setminus P$ region, biology but not physics, we have the $P \cap B$ region taking both, and we have the complement, taking neither, and these are all disjoint from each other.

So, now if we add up the students across these, we get the exact number of students. And in this case it is $5 + 20 + 10 + 15 = 50$. So, there are actually 55 students taking physics and biology together, but the total class strength is only 50. And actually only 45 students are taking these subjects because 5 are not taken either of them.

(Refer Slide Time: 04:43)



So, here is a variation where the data for the problem is given in a different way. So now, you are told the class strength 55, you are told that 32 students took physics and of them 11 took physics and biology and you are also told that 7 took neither. So, the question is how many took biology but not physics. So again, we draw a Venn diagram and from the previous question, we know that we can put 11 in the intersection, because that is the number who took both.

And since there are 32 who took physics, we can subtract out these 11 and say that $P\setminus B$ is 21 and in the complement, we have 7. So, the question now is how many are in $B\setminus P$, which I have marked by x, but now we know the total. So, we know that the four numbers together, add up to the total which is 55. So, $7 + 21 + 11 + x = 55$. So, if we solve for x, we get that $x = 16$. So, we can deduce that 16 students have taken biology but not physics in this situation.

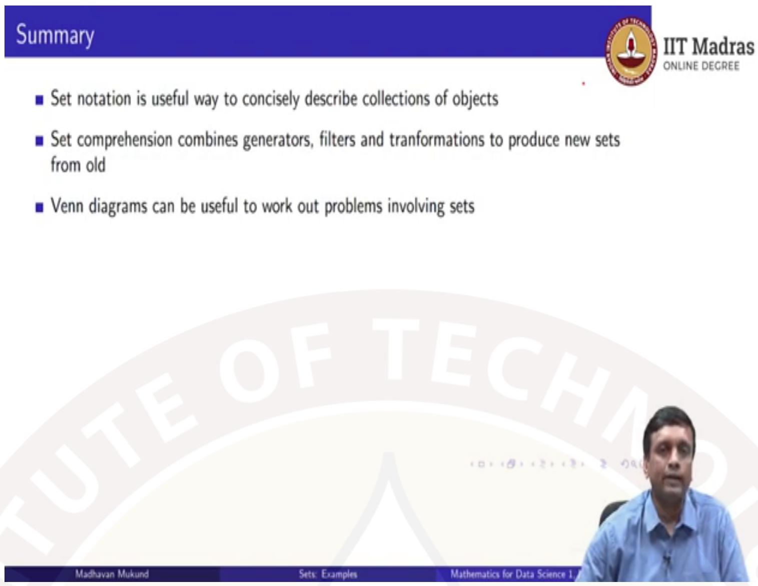
So here is yet another version of this. So, we have 60 students in the class. So again, we know the total number of students in the class, we are told that 35 students took biology, 35 students took physics, and 30 took biology, and 10 took neither. So now, we are trying to calculate the intersection, how many people took both subjects. So again, let us use this notation which we introduced when we first introduced sets.

So, this perpendicular bar on the side of a set indicates the size of the set. So, this is the cardinality of a set, cardinality is the number of elements, so the cardinality of Y is denoted by putting Y inside these bars. So, what we are told is that the set P has cardinality 35. That is a set of students who have taken physics overall, including those who have taken both, set B has 30 and 35 plus 30, there are 65 students who have taken in the union, of these I mean, have taken these together.

But we also know that there are 60 students in the class of whom 10 have taken neither. So, the actual union has only 50 elements. So, there are totally 65 people who are taking either physics or biology or both, but this total number actually spans only 50 students, so some of them must be taking both and are being counted twice. So, this must be the difference of the two.

So, 15 of these people must be counted twice, otherwise we would not have this mismatch. So, if we draw the diagram for this, this is how it comes out. We have 15, that we calculated for the intersection by taking the total number, realizing the 10 have taken neither, and then computing the difference between the number who should have taken both the subjects from those who are actually registered for either one or both of the subjects. So, these are three different examples using Venn diagrams to indicate how you can solve these kinds of counting problems.

(Refer Slide Time: 07:37)



So, to summarize, we use set notation because it is a very useful and precise way to talk about collections of objects. And if we use it nicely, it is also a concise way sometimes instead of writing out a long sequence of values, we can actually describe it using a condition. So, this is typically where we use set comprehension.

So, remember that set comprehension has three parts, some of which may not be used. So, you always have a generator, a basic set from which you are creating new sets, you may have a filter which takes out some elements from the generated set and throws them away and keeps only those that satisfy the condition.

And finally, you may have a transformation which takes these filtered elements and does something to make them into the elements that you want, for example, the squares of the even numbers. And then we also saw that Venn diagrams are not just simple doodles that you draw to indicate sets, Venn diagrams can actually be very useful for calculating properties about sets, especially numerical problems about sets. So, it is important to be able to draw the proper Venn diagram to indicate which groups of sets overlap, how they overlap, and which parts are empty, and so on.