2.12 School absences. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.

A few notes: The groups are disjoint, given that they are mutually exclusive. Also, we keep in mind that the outcome we are measuring, days missed due to sickness, will be whole numbers greater than or equal to zero.

(a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?

> Given that P(0) + P(1) + P(2) + P(3+) encompasses all possibilities (= 1.0 or 100%), and that the sets are disjoint, I can determine P(0) as equal to 100% - (P(1) + P(2) + P(3+), or 100% -(25% + 15% + 28%). P(0) = 32%.

(b) What is the probability that a student chosen at random misses no more than one day?

> P(0 or 1) = P(0) + P(1), since they are disjoint. P(0 or 1) = 32% + 25%, or 57%.

(c) What is the probability that a student chosen at random misses at least one day?

> P(1 or 2 or 3+) = P(1) + P(2) + P(3+), since they are disjoint. P(1 or 2 or 3+) = 25% + 15% + 28%. P(1 or 2 or 3+) = 68%. Note that we could have also calculated this as the complement of P(0), calculated in (a) above.

(d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

> Assuming that each child's attendance is an independent event, we can multiply probabilities such that P(0) for 2 children = P(0) * P(0), or .32 * .32. Under this assumption, the probability that neither child will miss days due to illness is .1024, or 10.24%

(e) If a parent has two kids at a DeKalb County elementary school, what is the probability that that both kids will miss some school, i.e. at least one day? Note any assumption you make.

> Assuming that each child's attendance is an independent event, we can multiply probabilities such that P(1 or 2 or 3+) for 2 children = P(1 or 2 or 3+) * P(1 or 2 or 3+), or .68 * .68. Under this assumption, the probability that neither child will miss days due to illness is .4624, or 46.24%

(f) If you made an assumption in part (d) or (e), do you think it was reasonable? If you didn't make any assumptions, double check your earlier answers.

> I do not believe the assumptions are reasonable, given that many childhood illnesses that would keep a child out of school are communicable and would spread within a family, such that the attendance of each child would no longer be an independent event.

2.14 Weight and health coverage, Part I. The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table summarizes two variables for the respondents: weight status using body mass index (BMI) and health coverage, which describes whether each respondent had health insurance.

		Weight Status			
		Neither overweight nor obese (BMI < 25)	**Overweight (25 ≤ BMI < 30)**	**Obese (BMI ≥ 30)**	**Total**
Health Coverage	**Yes**	134,801	141,699	107,301	383,801
	No	15,098	15,327	14,412	44,837
	Total	149,899	157,026	121,713	428,638

(a) If we draw one individual at random, what is the probability that the respondent is overweight and doesn't have health coverage?

Given that weight status and health coverage are two distinct events for which I have detailed data, I can determine the probability that the respondent is overweight and doesn't have health coverage as a joint probability.

I see that "overweight" and "no health coverage" line up in my table with a number of 15,327. I can divide that by my total sample size (428,638) to come up with a value of 0.0358, or 3.58%.

I can also calculate this probability using the general multiplication rule. Let P(overweight) = P(A) and P(no health coverage) = P(B). The general multiplication rule tells us that P(A and B) = P(A | B) * P(B). In our case, P(A | B), which means P(overweight | no health coverage) = (15,327 / 44,837) and P(B), which resolves to P(no health coverage) = 44,837 / 428,638). This evaluates to the same value we obtained with the previous approach.

(b) If we draw one individual at random, what is the probability that the respondent is overweight or doesn't have health coverage?

These two events, being overweight and not having health coverage, are not disjoint, so we use the general addition rule to calculate this probability. The general addition rule tells us that P(overweight or no health coverage) = P(overweight) + P(no health coverage) - P(overweight and no health coverage). This prevents double counting in the overlap of overweight and health coverage.

P(overweight) = 157,026 / 428,638.
P(no health coverage) = 44,837 / 428,638
P(overweight and no health coverage) = 15,327 / 428,638

P(overweight) + P(no health coverage) - P(overweight and no health coverage) = 0.4352, or 43.52%.

2.28 Socks in a drawer. In your sock drawer you have 4 blue, 5 gray, and 3 black socks. Half asleep one morning you grab 2 socks at random and put them on. Find the probability you end up wearing

(a) 2 blue socks

> The probability that the first sock I select will be blue is 4 (number of total blue socks) / 12 (total number of socks). The probability that the second sock I select will be blue is 3 (number of remaining blue socks) / 11 (number of remaining socks). Multiplying these probabilities (since they are independent events) gives me a probability of 0.091.

(b) no gray socks

> The probability that the first sock I select will be not grey is 7/12 (total non-grey socks / total socks). The probability that the second sock I select will be non grey is 6/11 (remaining non-grey socks / remaining socks). Multiplying these gives me a probability of 0.32.

(c) at least 1 black sock

> The probability of our first attempt turning out to be a black sock is simple: 3/12, or .25. We need not determine the probability of our second attempt if our first attempt to get a black sock is successful — regardless of the color of the second sock, the condition has been met. We will add this probability to the probability of failing on our first attempt but succeeding on our second.
>
> If we assume that we've failed on our first attempt, we still have 3 black socks, but 11 remaining socks, so the probability of success is 3/11.
>
> Our total probability is 3/12 + 3/11, or .523.

(d) a green sock

> There are no green socks, so the probability is 0.

(e) matching socks

> "Matching" here implies 2 blue, 2 grey, or 2 black socks. The probability of 2 blue socks has been calculated in part (a) to be 0.091 (or more precisely, 12/132). We can apply the same logic as we did in part (a) to calculate the probability of 2 grey socks (5/12 * 4/11) and 2 black socks (3/12 * 2/11).

Adding these probabilities, we get 38/132, or .288.

2.30 Books on a bookshelf. The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

		Format		
		Hardcover	***Paperback***	***Total***
Type	Fiction	13	59	72
	Nonfiction	15	8	23
	Total	28	67	95

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

The probability of drawing a hardcover book first, P(B), is 28/95 (total number of hardcovers divided by the total number of books). We multiply that probability by P(A l B), the probability of drawing a paperback fiction book second, assuming the first draw is of a hardcover book. This second probability is 59/94. P(A and B) is 0.18, or 18%.

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.

A tree diagram can help us here, since our conditional probability events are not disjoint.

First we determine P(A) = P(fiction book) as our initial fork:

Fiction drawn first?

Yes: 72/95

No: 23/95

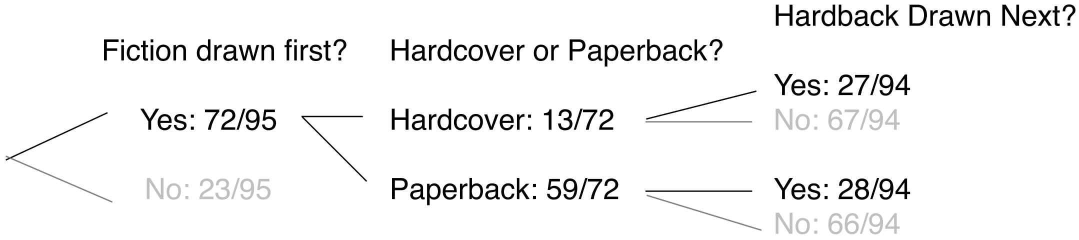
Then we add a fork for the two circumstances of "Yes, fiction was drawn first," namely, that fiction hardcover was drawn and that fiction paperback was drawn. We don't bother fleshing out the "No" possibility, since it doesn't apply to what we're looking for.

Fiction drawn first? Hardcover or Paperback?

Yes: 72/95 Hardcover: 13/72

No: 23/95 Paperback: 59/72

Next, we calculate the probability that a hardcover book was drawn second, given the two scenarios that matter to us: that a hardcover fiction was drawn first, and that a paperback fiction work was drawn first. I grey out the results that don't matter to the question at hand.



Now we have two distinct paths in our tree:

The possibility that a hardcover fiction book is selected first and followed by a hardback book second can be calculated by following the upper branch and multiplying (72/95) * (13/72) * (27/94), which results in an upper branch probability of 0.039

The possibility that a paperback fiction book is selected first and followed by a hardcover book can be calculated by following the lower branch and multiplying (72/95) * (59/72) * (28/94), which results in a lower branch probability of 0.18.

Adding these two possibilities together, we get 0.22, or a 22% chance that we draw a fiction book first and then a hardcover book second.

(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.

In this case, there is no need for a tree diagram, because we are dealing with independent events. We can simply multiply the possibility of getting a fiction book by the probability of getting a hardcover book:

(72/95) * (28/95), or 0.22.

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

Because of rounding, the answers in (b) and (c) look identical. It is not surprising that they are quite close, because the numbers of books are high enough that replacement or lack thereof has only a small effect on overall probability. If our n had been significantly lower (say, dealing with a library of 20 books), the difference between (b) and (c) would have been more noticable.