Suppose that we have a population of 100 people; 80 of these people are "healthy," and 20 are at risk for ALS (Lou Gehrig's Disease). Fortunately, we also have a medical test to tell us whether or not someone is at risk for ALS (presumably so they can obtain preventative treatment); unfortunately, this test is not completely accurate. We run two experiments, in the sense we've been using in class, and obtain the following results:

Frequency	P = NAR	P = AR
T = NAR	72	2
T = AR	8	18

where T is the result reported by the medical test, P is the actual state of the person, and AR and NAR stand for "at risk" and "not at risk," respectively (note: sure, if we know whether or not a person is at risk running an inaccurate test is ridiculous; but let's ignore real-world complications for now and just go with it).

Since we'd rather know the probability of a given person, test combination falling into a given cell in the frequency table, we divide by the total size of our population (100 people) to obtain the probability:

Probability	P = NAR	P = AR
T = NAR	0.72	0.02
T = AR	0.08	0.18

From this, we ask a few questions:

(a) What is the marginal frequency of a person's type (at risk or not at risk)? Of the test's results?

(b) Conditional on a person being not at risk, what is the probability that the test says that they are not at risk? Conditional on a person being at risk, what is the probability that the test says that they are at risk?

(c) Define an event E which represents, "The test reported correctly."

(d) Conditional on the test reporting correctly, what is the probability that a person is at risk? Are these two events independent?

(e) Conditional on the test reporting that a person is at risk, what is the probability that a person is at risk? Are these two events independent?

(there are no typed answers for these right now; if you'd like to see some, we can probably work that out)

Now, we slightly change the question and the information that it gives. Suppose that now we know that 80% of people are not at risk, that the test for whether or not someone is at risk is correct 90% of the time (or it is 90% accurate), and that its accuracy is not affected by whether or not the person is at risk.

This last statement is code for, "Whether or not the test is correct is independent of whether or not the person is at risk." To see this, it may help to represent it a little more mathematically (for shortness, we will abbreviate "test is correct" by "TIC"):

$$\begin{aligned}
P(\text{TIC}|P=\text{AR}) &= P(\text{TIC}|P=\text{NAR}) \\
\frac{P(P=\text{AR}|\text{TIC})P(\text{TIC})}{P(P=\text{AR})} &= \frac{P(P=\text{NAR}|\text{TIC})P(\text{TIC})}{P(P=\text{NAR})} && \text{(Bayes' rule)} \\
\frac{P(P=\text{AR}|\text{TIC})P(\text{TIC})}{P(P=\text{AR})} &= \frac{\Big(1-P(P=\text{AR}|\text{TIC})\Big)P(\text{TIC})}{P(P=\text{NAR})} && \text{(Complement rule)} \\
P(P=\text{NAR})P(P=\text{AR}|\text{TIC}) &= \Big(1-P(P=\text{AR}|\text{TIC})\Big)P(P=\text{AR}) && \text{(algebra)} \\
P(P=\text{NAR})\frac{P(P=\text{AR},\text{TIC})}{P(\text{TIC})} &= \left(1-\frac{P(P=\text{AR},\text{TIC})}{P(\text{TIC})}\right)P(P=\text{AR}) && \text{(Definition of conditional)} \\
P(P=\text{NAR})P(P=\text{AR},\text{TIC}) &= \left(P(\text{TIC})-P(P=\text{AR},\text{TIC})\right)P(P=\text{AR}) && \text{(algebra)} \\
\Big(P(P=\text{NAR})+P(P=\text{AR})\Big)P(P=\text{AR},\text{TIC}) &= P(\text{TIC})P(P=\text{AR}) && \text{(algebra)} \\
P(P=\text{AR},\text{TIC}) &= P(\text{TIC})P(P=\text{AR}) && \text{(Law of total probability)}
\end{aligned}$$

But this last line is exactly the definition of independence. Then the test's correctness is independent of whether or not a person is at risk.

As one final application of Bayes' rule, we ask the following: what is the probability that someone is at risk, given that the test says that they're at risk?

$$P(P=\text{AR}|T=\text{AR}) = \frac{P(T=\text{AR}|P=\text{AR})P(P=\text{AR})}{P(T=\text{AR})}$$

We know by the question's assumptions that $P(P = \text{AR}) = 0.2$; further, since the test is correct 90% of the time and is independent of a person's actual state, $P(T = \text{AR}|P = \text{AR}) = 0.9$. So now we only need to figure out $P(T = \text{AR})$. This we will do by applying the law of total probability,

$$\begin{aligned}
P(T=\text{AR}) &= P(T=\text{AR},P=\text{AR}) + P(T=\text{AR},P=\text{NAR}) \\
&= P(\text{TIC},P=\text{AR}) + P(\text{test is not correct},P=\text{NAR}) && \text{(think about this)} \\
&= P(\text{TIC})P(P=\text{AR}) + P(\text{test is not correct})P(P=\text{NAR}) && \text{(independence)} \\
&= P(\text{TIC})P(P=\text{AR}) + \Big(1-P(\text{TIC})\Big)P(P=\text{NAR}) && \text{(complement rule)} \\
&= 0.9(0.2) + (1-0.9)(0.8) && \text{(plugging in)} \\
P(T=\text{AR}) &= 0.26
\end{aligned}$$

Then we've found the last term we need in our equation, and we find

$$P(P=\text{AR}|T=\text{AR}) = \frac{0.9(0.2)}{0.26} = 0.6923 \quad \text{(roughly)}$$

This should be exactly what we'd find in the previous example, but we've done it here with far less "information" (in terms of explicit probabilities); Bayes' rule allows us to accomplish this.

As a side note, it can be a little confusing that a test which is 90% accurate yields results that say that a person is at risk only 70% of the time the test tells us so (for even more mind bending, try computing the probability that a person is not at risk, given that the test tells us they are not at risk). On an intuitive level, what is happening is that there are far more people who are not at risk than are at risk, so even though the test is fairly accurate, the number of people misreported as at risk is larger than we'd expect it to be. These misreports outweight the truthfulness of the test, causing it to seem less accurate when it is actually applied. By altering the numbers a little bit, we could make these results even more drastic; if only 1% of the population is at risk but the test is still 90% accurate, there is less than a 10% chance that a person is at risk given that the test has told us so!