For my interest driven problem, I chose to use baseball statistics that apply to certain players to find the probability of many different situations that could happen against a certain pitcher. To begin my problem, I chose three batters on the Yankees and used their stats against the Red Sox starting pitcher, Clay Buchholz. The three batters I chose were Jacoby Ellsbury, Brett Gardner, and Alex Rodriguez. In the lineup that I chose, Ellsbury bats first, Gardner second, and a-rod bats third. To get everything straight, the abbreviation, OBP, stands for on base percentage. This abbreviation is very important because OBP means something a little bit different than AVG (batting average - the chance that the batter has of getting a hit based on the amount of hits they have during that season - in this case, the AVG and the OBP are based on the batter's stats against Clay Buchholz in their entire career). OBP is the chance that the batter gets on base. Getting on base includes getting walked, a hit by pitch, or any time that the batter gets on base without getting a hit. Usually, a batter's OBP is a little bit higher than his AVG. From there, my job was to find the probability that the hitters get singles, doubles, triples, or home runs and then find the probability that the first batter gets a single, the second batter gets a double, and the third batter gets a home run, etc.. The list of situations goes on and on. For some batters, they have never gotten a triple or a home run, so based on the stats, that batter has a 0% chance of getting to that base. Obviously, there is a possibility that the batter gets a triple or a home run but based on what I know, it is impossible for him to get to that base.

To find the probability of these situations, I needed to start off by finding the probability that each individual batter got to each different base and the probability that the batters got on base without getting a hit. To find the probability that the hitter got on without a hit, I took the OBP and then subtracted the AVG from it. By doing this, I am taking away the probability that the batter gets a hit from the OBP and leaving just the probability that the batter gets on without a hit. After this, I found the probability that the batters get to each base. From the data I had, all I knew was how many singles, doubles, triples, or home runs the batter got against Clay Buchholz in their career. I knew the AVG and the OBP for every batter and I also knew the amount of at bats the batter had and the amount of hits they had. With this information, I took the batter's average and divided it by the number of hits that the batter got. I then multiply that number by the number of singles, doubles, triples, or home runs the batter got and get the probability that that batter gets to the base I was focusing on. At first, I just found these probabilities for Ellsbury and Gardner because I needed to find the probability that Ellsbury got a single, and Gardner got a double, etc., to then find the probability that Ellsbury gets a single, Gardner gets a single, and a-rod gets a home run etc.. When I found all of the probabilities for Gardner and Ellsbury, I used a visual model to represent the probability in a different way than numbers. I drew a ten by ten square and put Ellsbury on the left side, and Gardner on the top. Each of these sides represented 100%. If Ellsbury got 10/10 singles on Buchholz, his whole side would be the probability that he gets a single, 100% chance. With each side being 100%, this then meant that every individual box represented 10%. Ellsbury had a .200 average so from the bottom up, I find 20% and then draw a line horizontally across the square. With the .073 chance that Ellsbury gets on without a hit, I then draw a line at 27.3% because .273 is 27.3%. Using this method, you could physically see the probability that Ellsbury got a hit (he hasn't got anything but singles against Buchholz so the probability that he gets a hit is the probability that he gets a single), which really helped me understand how small or big the chance really was. With Gardner on the top line, from right to left I did the same thing but this time for Gardner getting on without a hit, Gardner getting a single, and Gardner getting a double. When these lines cross in the bottom right hand corner of the square, you could see the probability that Ellsbury got on without a hit, and Gardner got a double etc.. There are many different situations that you could see the probability of and to find the probability of these situations in numbers rather than the size of a box, you would multiply the two probabilities together. If you wanted to find the probability that both Ellsbury and Gardner got on base, you would multiply .273 (Ellsbury's on base percentage) by .290 (Gardner's on base percentage) and you would get .07917. After I found the probabilities between Ellsbury and Gardner, I needed to find the probabilities of the different situations between Ellsbury/Gardner and a-rod. To do this, I made a one hundred by one hundred square box and put Ellsbury/Gardner on the left side while putting a-rod on the top. The reason that I made the square one hundred by one hundred is because this time for Ellsbury and Gardner, I drew a line at .07917 to represent that both the runners got on base and then split that up into the different situations in which they both got on. With a ten by ten square, 10% is represented by one tiny box so I couldn't fit all of the different outcomes in one box and with a one hundred by one hundred, 10% is ten boxes so it was much easier to fit all of the lines in ten boxes rather than one. After this, I did exactly what I did when I used just Ellsbury and Gardner, and got probabilities like a .00413 chance that Ellsbury got a single, Gardner got a double, and a-rod got a home run.

To end my interest driven problem, I found a bunch of probabilities for many different situations. With the probabilities between just Ellsbury and Gardner, I then used those numbers and multiplied them by (for example) the probability that a-rod gets a home run, single, double, or triple. With all of this information, I have access to the probabilities of about 100 different situations that could come with Ellsbury, Gardner, and Texeira. A couple of the most important probabilities that I got were a .00815 chance that all three runs score. I then wanted to find the probability that any runs score but it ended up being too hard to calculate. To find the chance that any runs score, I needed to add up all of the situations in which a run scores. There were way too many situations to add up so I just decided to estimate this probability. In the end, my estimate was .23 which basically means that there is a 23% chance that any runs score. In easier words, 23 out of 100 games when Ellsbury bats first, Gardner second, and a-rod third, at least one run will score not including any of the batters after a-rod.