Most teams did fine with #1, 3, 4, and 5. Any mistakes should be cleared up by the comments I made in your evaluation email.

The one that several teams struggled with was #2. It turns out that describing a process (or in computer science, an *algorithm*) is not always easy to do. Following are two excellent answers to #2. They are general enough to apply to any number of missing 7-mers. Elizabeth and Will's is very succinct, but Peter and Anders' has the benefit of examples. Great job, both teams!

Elizabeth & Will Kuchinski

2) We started out writing into my $regex a single letter to see if all possibilities of words starting with that letter were present in the text. Since there are 7 positions, with one fixed position, each letter should have printed out 4^6 different words if all possibilities were in the text. We did this until we found a letter(s) that did not produce 4^6 possibilities. Thus we continued using the letter(s) that did not have 4^6 words and tried the four different letters on the 2nd position to see if they produced all 4^5 possibilities. We continued in this manner adding on the letters that did not produce 2^(7-n) possibilities, where n is the position we were changing. Finally we were able to pinpoint the exact missing 7-mer.

Peter & Anders:

(Question 2) Describe a procedure for finding all missing 7-mers in a list.

We can use the number of possible combinations for each given selection to limit the search and make it simpler. A sample procedure is as follows:

First count the total number of 7-mers. If the number matches the total number possible, 4^7, there are no 7-mers missing – we are done. If the number is less than 4^7 then continue searching. Save the number missing in order to know how many we have to find.

Then we will search for one missing7-mer at a time. We will stop when we have found all of the missing 7-mers.

To start the procedure, we have to define a search position. We will start from 'A'. Thus count every 7-mer that starts with 'A'.
If that number is 4^6 then start searching 7-mers beginning with other characters. Find all 7-mers that starts with 'C' and repeat the procedure.
If that number is not 4^6 then continue searching for all missing 7-mers starting with 'A'.

Count every 7-mer that starts with ‘AA’. If that number is 4^5 then search ‘AT’, then ‘AC’, then ‘AG’ until there is a number that is less than 4^5. Then continue searching given the missing initial combination (e.g. AC).

Count every 7-mer that starts with the two letters selected above plus another letter (example => ‘AAC’). If that number is 4^4 then search ‘AAA’, ‘AAT’, and ‘AAG’ until there is a number less that 4^4. Then continue searching.

ETC….

Once the missing 7-mer has been found through this winnowing process, it will be printed. The number of missing 7-mers is decreased by one, it is checked if it is 0. If no more are missing, we are done.
If more are missing, we have to start searching from where we left off – which is saved through the position of this search procedure.