Answers to Challenge #1

1. ATGTCATTTCCAAG

2. CTTGGAAATGACAT

3. P(A) = 4/14 P(T) = 5/14 P(G) = 2 / 14 P(C) = 3/14

4. P(AGCTCG) = 4/14 * 2/14 * 3/14 * 5/14 * 3/14 * 2 / 14 = 0.000096

5. Lets enumerate all of the 2-mers that start with each character below. Note there will only be 13 since the last G does not count for this:

AT	TG	CA	GT
AT	TC	CC	
AA	TT	CA	
AG	TT		
	TC		

   As an example, for P(T|A) we compute #AT / all 2 bases strings of size 2 that start with an A, or 2/4. It follows that:

   P(T|A) = 2/4 P(A|A) = ¼ P(G|A) = ¼
   P(T|T) = 2/5 P(C|T) = 2/5 P(G|T) = 1/5
   P(C|C) = 1/3 P(A|C) = 2/3
   P(T|G) = 1

   All other (n = 7) probabilities would be 0 given my training string above

6. Here is one alignment of the two strings from the lecture/text

   VIVALA S VEGAS
   VIVADA - VI - - S

   Note that you can align the "I" in DAVIS also to the G or A and get the same score under a constant gap penalty scheme

7. Here is a simple alignment of the two DNA strings:

   CATC–A–C
   C–TCCAGC

   1 -1 1 1 -1 1 -1 1 = score of 2