(10 points)

(A) We are all molistic in a way

Imagine that you have heard these sentences:

Jane is molistic and slatty.
Jennifer is cluvious and brastic.
Molly and Kyle are slatty but danty.
The teacher is danty and cloovy.
Mary is blitty but cloovy.
Jeremiah is not only sloshful but also weasy.
Even though frumsy, Jim is sloshful.
Strungy and struffy, Diane was a pleasure to watch.
Even though weasy, John is strungy.
Carla is blitty but struffy.
The salespeople were cluvious and not slatty.

A1. Then which of the following would you be likely to hear?

a. Meredith is blitty and brastic.
b. The singer was not only molistic but also cluvious.
c. May found a dog that was danty but sloshful.

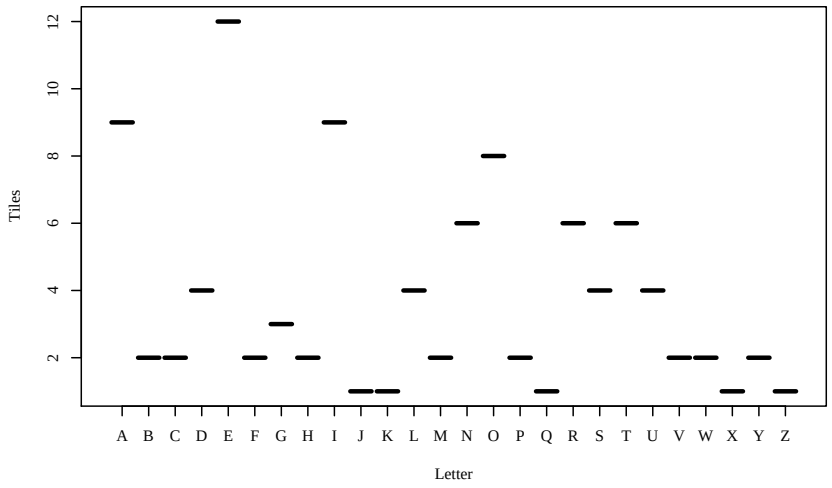
A2. What quality or qualities would you be looking for in a person?

a. blitty
b. weasy
c. sloshful
d. frumsy

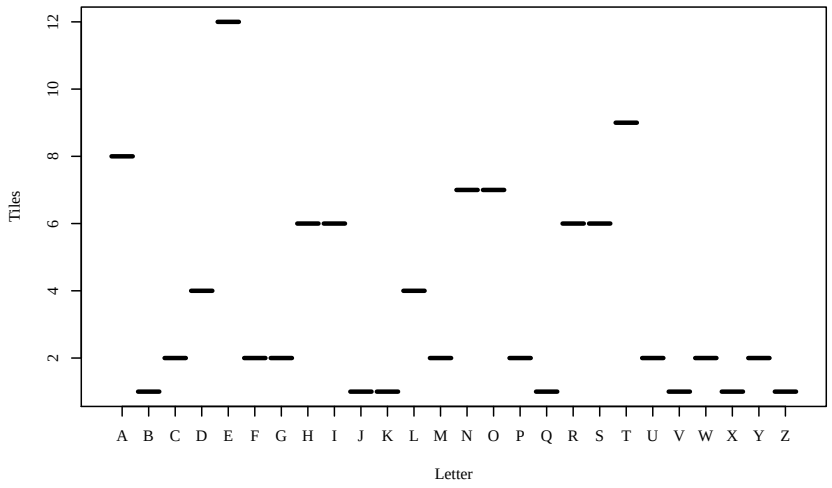
A3. Explain all your answers. (Hint: The sounds of the words are not relevant to their meanings)

Scrabble letters

When Alfred Mosher Butts developed Scrabble beginning in 1933, he chose the distribution of letters after long and careful consideration. He ultimately decided there should be 100 tiles, with two blanks, and the other 98 divided among the letters like so:



If Butts had simply counted the letters on the front page of *The New York Times* as is commonly believed, his letter distribution would have been more like this:



For the questions that follow, let's assume that the actual Scrabble distribution is perfect and that the alternate distribution is wrong.

(a) Which five letters' counts change the most between the two distributions?

(b) The Brown Corpus consists of over one million words of text taken from a variety of sources and genres. We will pretend it is a reasonable approximation to the front page of *The New York Times*.

The twenty words that occur most frequently in the Brown Corpus are *the*, *of*, *and*, *to*, *a*, *in*, *that*, *is*, *was*, *he*, *for*, *it*, *with*, *as*, *his*, *on*, *be*, *at*, *by*, and *I*, in that order. These twenty words comprise about 31% of the word tokens in the corpus. Here *token* refers to an instance of a word in the text.

For which of the letters in your answer to (a) does this list help explain why the two distributions assign it a different number of tiles?

(c) The words in (b) are not equally frequent. Rather, frequency decreases rapidly with rank:

	Word	Frequency (%)
1.	the	6.8872
2.	of	3.5839
3.	and	2.8401
4.	to	2.5744
5.	a	2.2996
6.	in	2.1010
7.	that	1.0428
8.	is	0.9943
9.	was	0.9661
10.	he	0.9392
11.	for	0.9340
12.	it	0.8623
13.	with	0.7176
14.	as	0.7137
15.	his	0.6886
16.	on	0.6636
17.	be	0.6276
18.	at	0.5293
19.	by	0.5224
20.	I	0.5099

How does this new information change your answer to part (b)?

(10 points)

(B) Pooh's encyclopedia

Once upon a time, a very long time ago, Winnie-the-Pooh and his friends bought an electronic encyclopedia, and tried to find answers to several important questions:

Winnie-the-Pooh:
Where should a bear stock his jars of honey?
How much honey should a bear store for the winter?

Eeyore:
Where should I look for my lost tail?
Which animals sleep during the winter?

Christopher Robin:
What is the shortest way from my place to the house of Winnie-the-Pooh?
Who wrote the books about Pooh Bear?

The encyclopedia's search engine identified a number of articles related to their questions; for example, it returned the following matches:

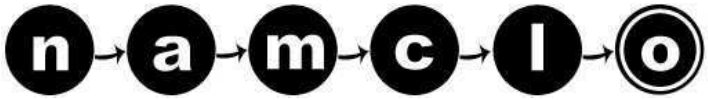
Winter food storage (for Winnie-the-Pooh)
Sleep patterns in mammals and other animals (for Eeyore)
Short stories and movies about Winnie-the-Pooh (for Christopher Robin)
Writers of children's books (for Christopher Robin)

On the other hand, the search engine missed several other relevant articles; in particular, it did *not* retrieve the following articles:

Planning of food supplies
Lost-and-found agencies
Finding shortest paths on a map
Biography of A.A. Milne, the author of Winnie-the-Pooh

Your task is to determine who received each of the following matches; two of these matches were for Winnie-the-Pooh, two for Eeyore, and two for Christopher Robin. Explain why!

Books about care and feeding of bears
Effects of honey on the sleep quality of humans and animals
Lost tales of "Bulls vs. Bears" stock trading
Ways to look for lost things
Ways to store food in the house
Winter hibernation of bears and rodents



Kernel Sentences and Complex Sentences

In the 1950's, the linguist Zellig Harris proposed that complex sentences could be derived from "kernel" sentences. Some examples of kernel sentences and complex sentences are shown below.

Kernel sentences

- The bear ate a sandwich.
- The bear yawned.
- It was likely that S (where S is a sentence)
- X persuaded Y that S (where X and Y are sentient beings and S is a sentence)

Complex sentences

- A sandwich was eaten by the bear.
- The bear that ate the sandwich yawned.
- It was likely that the bear ate a sandwich.
- The bear was likely to eat a sandwich.

In order to turn kernel sentences into complex sentences, there must be a set of operations for combining and transforming sentences. Following are some operations that you can practice.

Operation 1: Passive Voice

```
Input: S1: X verb Y rest-of-S1
Output: Y was verb-past-participle by X rest-of-S1

Example:
Input: The bear (X) ate (verb) a sandwich (Y) in the park (rest-of-S1).
Output: A sandwich was eaten by the bear in the park.
```

Task 1: Perform the Passive Voice Operation on the sentence: *Five students took the test on Tuesday.*

1. In the input to the operation, what is X?
2. What is Y?
3. What is the verb?
4. What is rest-of-S1?
5. What is the output of the Passive Voice Operation?

Operation 2: Relative Clause

```
Input: S1: X verb-1 rest-of-S1
       S2: X verb-2 rest-of-S2
Output: X that verb-2 rest-of-S2 verb-1 rest-of-S1

Example:
Input: S1: The bear (X) ate (verb-1) a sandwich (rest-of-S1)
       S2: The bear (X) yawned (verb-2)
Output: The bear (X) that yawned (verb-2) ate (verb-1) a sandwich (rest-of-S1).
        (Note that rest-of-S2 is empty because S2 ends with verb-2.)
```

Task 2: Perform the Relative Clause Operation where S1 is *The student passed the test* and S2 is: *The student saw a movie*.

1. In the input to the operation, what is X?
2. What is verb-1?
3. What is rest-of-S1?
4. What is verb-2?
5. What is rest-of-S2?
6. What is the output of the Relative Clause Operation?

Operation 3: substitution

```
Example:  Input: X persuaded Y that S1
                 X=Pat
                 Y=Kim
                 S1=the sandwich rotted.

           Output: Pat persuaded the bear that the sandwich rotted.


Example: Input: It was likely that S1
                S1=The bear yawned.
         Output: It was likely that the bear yawned.
```

Task 3: Perform substitution operations:

1. What is the output of this substitution:

   ```
   Input: X persuaded Y that S1
          X=Chris
          Y=the teacher
          S1=the student passed the test
   ```

2. What is the output of this substitution:

   ```
   Input: It was likely that S1
          S1=the student passed the test
   ```

Operation 4: Infinitive Operation for "likely"

```
Input: It was likely that S1
       S1: X verb-1 rest-of-S1
Output: X was likely to verb-1-infinitive

Example:
Input: It was likely that S1
       S1: The bear (X) ate (verb-1) a sandwich (rest-of-S1)
Output: The bear (X) was likely to eat (verb-1-infinitive) a sandwich(rest-of-S1).
```

Operation 5: Infinitive Operation for "persuade"

```
Input: X persuaded Y that S1
       S1: Y verb rest-of-S1.
Output: X persuaded Y to verb-infinitive rest-of-S1.

Example:
Input: X persuaded Y that S1
       X=the bear
       Y=the butterfly
       S1: The butterfly flies away.
Output: The bear (X) persuaded the butterfly (Y) to fly
        (verb-infinitive) away (rest-of-S1)

        (Note that this operation slightly changed the meaning of the
        kernel sentences.)
```

Task 4: Perform the following infinitive operations:

1. Infinitive Operation for "likely" where S1 is *The student passed the test*.
   (a) In the input to the infinitive operation, what is X?
   (b) What is verb-1?
   (c) What is rest-of-S1?
   (d) What is the output of the infinitive operation?
2. Infinitive Operation for "persuade" where X is *the teacher*, Y is *the student* and S1 is *The student does the homework*.
   (a) What is the output of the operation?

Combining Operations: The following sentences can be derived from kernel sentences by a sequence of the operations described above. For each sentence give the kernel sentences that it is derived from and list the operations that apply. One example is done for you: *The bear was persuaded by the students to yawn.*

```
Kernel Sentences: S1: The bear yawned.
                  X persuaded Y that S1

Operations:
1.  Infinitive Operation for ‘‘persuade’’
    Input: The students persuaded the bear that the bear yawned.
    Output: The students persuaded the bear to yawn.

2.  Pasive Operation:
    Input: The students (X) persuaded (verb) the bear (Y) to yawn (rest-of-S).
    Output: The bear (Y) was persuaded (verb-past-participle)
            by the students (X) to yawn (rest-of-S).
```

Task 5: Provide derivations for the following sentences including kernel sentences and operations.

1. The bear that was likely to eat a sandwich yawned.
2. The sandwich was eaten by the bear that was persuaded to yawn.