1. Sometimes it is difficult to understand data if you do not know what the numbers represent. Provide short definitions of two words: sepal, and petal (be sure to cite your sources even if you paraphrase):

sepal: **each of the parts of the calyx of a flower, enclosing the petals and typically green and leaflike.**

petal: **each of the segments of the corolla of a flower, which are modified leaves and are typically colored.**

2. There is a cumulative relative frequency table printed above for petal lengths (using rounded values for petal length). Below the number 3 in that table is the number .35. What does .35 represent? (multiple choice)

a. Three of the flowers had petal length of 0.35.

b. There were 0.35 observations that had petal length of 3 (after rounding the petal lengths).

c. Of all the flowers measured in this sample 35% had a petal length of 3 (after rounding the petal lengths).

d. Of all the flowers measured in this sample 35% had a petal length of 3 or less (after rounding the petal lengths).

e. A study of all flowers on the planet would show that about 35% had petal lengths of 3 or less (after rounding the petal lengths).

3. Using only the cumulative relative frequency table printed above combined with some simple paper-and- pencil calculations, which petal length occurs most frequently ? **4 and 5**

4. Describe how you determined your answer to the previous question (describe the calculations that you used). Do not show R code for this task--it will not be counted as an answer.

There are 150 flowers in the database

Find the relative frequency (subtract each cumulative frequency with the previous value. Example: The relative frequency for petal of length 1 is 0.16. So its frequency is 0.16x150. The relative frequency for petal of length 2 is 0.33-0.16 =0.17 So its frequency is 0.17*150) and the multiply by 150 to get frequency

---

5. Assuming that you read the flowers.csv file into an R object called flower.data, run the following R code (do not paste the ">" character into R) and paste both the command and the output into your answer (you should see five names, each of which should be enclosed in quotes--if you do not see this, try again or contact your instructor):

```
> names(flower.data)

[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
```

6. The number of observations in the "flower.data" data frame is: .

```
> dim(dat)[1]
[1] 150
```

7. List the variables in the data frame (you can do this by entering the name of the R object that holds that data that you read using the read.csv command--you should have called it flower.data). If you do not see five columns of data, then there was a problem reading the input file--try again or contact your instructor. For each variable identify the type of the variable (factor or numeric).

The name and type of the 1st variable:	**Sepal.Length**	**type:num**
The name and type of the 2nd variable:	**Sepal.Width**	**type:num**
The name and type of the 3nd variable:	**Petal.Length**	**type:num**
The name and type of the 4nd variable:	**Petal.Width**	**type:num**
The name and type of the 5nd variable:	**Species**	**type: Factor**

```
> str(dat)
'data.frame': 150 obs. of 5 variables:
$ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
$ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
$ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
$ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
$ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

8. Round the data for the variable Sepal.Length so that it contains integers, then find the frequency of the value 7 (not the relative frequency): **24**

```
dat<-read.csv("flowers.csv")

sepal=round(dat$Sepal.Length,0)

table(sepal)
```

Assuming that you read the flowers.csv file into an R object called flower.data, run the following R code (do not paste the ">" character into R). Note that we are not rounding the numbers here. Use the output for the next five tasks:

```
> table(flower.data$Sepal.Width)

> plot(table(flower.data$Sepal.Width))
```

9. What is the sum of the first three frequencies in the frequency table for sepal width? **8**

10. What does your answer to the previous question represent (in terms of sepal width and frequency and the percentage of all sepal measurements)

It is the cumulative frequency telling that there are 8 sepal with width less than or equal to 2.3

11. What is the sum of the last three frequencies in the frequency table for sepal width? **3**

12. How many flowers in the sample had sepal widths less than 4 (do NOT round the sepal width numbers for this, but you can round your final answer to 3 decimal places)? **146**

```
freq1 <- table(dat$Sepal.Width)

cumsum(freq1)
```

13. What does the tallest bar in the plot represent? **mode**

14. Create a frequency table that shows the frequencies for each species of flower in the sample. Paste your R command and output into your answer (do NOT display data from a data frame, display data using the table() command)

```
> table(dat$Species)

    setosa versicolor  virginica
        50         50         50
```

15. Explain two things about the table that you created for the previous task:

Why did the frequency table for flower species contain words in the first row as opposed to numbers? **The type of this variable is factor**

What is the meaning of the numbers in the second row of the table? **It is the frequency of each species in the sample**