Lab 1 – a trip to the zoo

Ms. Livingston's 9th grade biology class visited the zoo yesterday. The students collected several data sets during their visit.

SAS
The students recorded the height of each giraffe living at the zoo and rounded off to the nearest tenth of a meter. Here is a SAS program that creates a data set named giraffes. It is available on Canvas as the file giraffes.sas. Another option is to type it into the Program Editor window yourself.

```
data giraffes;
input height @@;
datalines;
5.4 4.7 4.2 5.1 2.4 3.3 4.8 5.0 3.5 3.7 2.8 5.5 1.2 4.1 5.8 1.5 4.5 5.7
5.2 3.1
;
proc print data=giraffes;
run;
```

Run the program once to confirm it works. You should see the data set in the output window. Now, add lines to the SAS program that will enable you to answer the following questions. The procs you need can be found on the SAS 2 handout.

How many giraffes are there? n = _____

mean height = _____

median height = _____

standard deviation = _____

75th percentile = _____

range = max – min = _____

R

Now let's work with the same data set within R. Create a vector named giraffes using the c function as seen below. You can type it in or copy/paste it from a file available on Canvas named giraffes.txt.

```
> giraffes=c(5.4,4.7,4.2,5.1,2.4,3.3,4.8,5.0,3.5,3.7,
2.8,5.5,1.2,4.1,5.8,1.5,4.5,5.7,5.2,3.1)
```

Confirm your answers on the previous page using built-in R functions. The functions you need can be found on the R 2 handout. <u>Warning</u>: SAS and R may disagree on the values of the 25th and 75th percentile!

Have R draw a histogram and stem-and-leaf plot.

Describe their shape: _____

What value is larger, the mean or the median? Does your answer agree with the shape of the histogram and stem-and-leaf plot?

****

Several of Ms. Livingston's students decided to purchase a t-shirt from the zoo gift shop. Here are the sizes and colors of the t-shirts that were sold:

L	blue
S	red
L	blue
M	red
L	red
L	red
S	red
M	blue
L	blue
M	red
M	blue
M	blue
L	red
L	blue

SAS

Write your own SAS program that creates a data set named tshirts, inputs character variables named size and color, then reads in the data using the datalines command.

Choose appropriate procs (look at the SAS 2 handout) to add to your SAS program and answer the following questions.

How many of each size t-shirt was sold?

S	M	L	total

How many of each color t-shirt was sold?

red	blue	total

Have SAS draw pie charts for both variables.

What proc did you use? _____

Fill in the two-way frequency table with counts:

	S	M	L	
red				
blue				

****

Every person who visits the zoo fills out a survey before they leave. One question on the survey is: "Before today, how many times have you visited our zoo?". Ms. Livingston's class is given the last 100 responses to this question as a data set.

R
Generate the data set at random using the following code. Name the data visits. It will be 100 observations of a Poisson random variable with mean = 3.

```
> visits=rpois(100,3)
```

To peek at the beginning few observations (or last few observations) try the head or tail function.

```
> head(visits)
> tail(visits)
```

Using built-in R functions (look at the R 2 handout) answer the following questions about the data set:

mean # of visits = _____

(Hint: this value should be close to what number? _____)

max # of visits = _____

Have R draw a histogram and stem-and-leaf plot.

Describe their shape: _____

Execute the following code:

```
> table(visits)
```

How many visitors had never been to the zoo before? _____

How many visitors had been to the zoo exactly once before? _____

****

After you have completed this handout, complete the Canvas quiz titled:
Lab 01 – a trip to the zoo