ISTA 116: Lab Assignment #6 (50 pts)

Due Monday, November 14th, 11:59 P.M.

Problem 1:

(32 pts)

A robot steps back and forth randomly along a line, but never ventures more than two steps away from the center point. After ten minutes of wandering, its position measured in number of steps from the center is given by the random variable X. Negative numbers indicate that the robot is to the left of the center point, and positive numbers indicate that it is to the right.

The distribution of X is characterized by the following probability mass function:

x	-2	-1	0	1	2
$P(X = x)$	0.05	0.2	0.5	0.2	0.05

a. (4 pts) Produce a "spike plot" based on this distribution.

b. (4 pts) Create a table showing the CDF for the distribution.

c. (6 pts) Compute the mean and standard deviation of this distribution.

d. (6 pts) Produce 3 random samples (*with replacement*) from this distribution: one of size 20, one of size 200, and one of size 2000. Use `sample()` in `R`, and store the result of each sample as a variable.

e. (6 pts) For each sample, produce a table and a spike plot showing the proportion of draws at each value (you'll need to use `table()` followed by `prop.table()` for the table, followed by `plot()` for the plot). If you should end up with a sample where not every possible value shows up, you may need to use the following trick after creating your sample and before creating your table to get 0s to appear in the table, rather than leaving out cells entirely:

```
mySample <- sample(some code here)
mySample <- factor(mySample, levels = x)
```

where `x` is the vector of the possible values.

Show both the tables and plots, and clearly label which corresponds to which sample size.

f. (4 pts) Compare the sample spike plots in (e) to the distribution spike plot in (a). Comment on the similarities and differences.

g. (4 pts) Repeat parts (d) and (e), generating new samples. Which sample has changed the most?

Problem 2:

(18 pts)

The following can all be answered either with simple counting or with one or two line computations in `R`. (Hint: for each one, consider a random variable that captures the quantity of interest, and think about whether it is characterized by a named distribution)

a. (3 pts) What's the probability of requiring all 7 flips to decide the winner in a "best of 7" coin toss game, assuming the coin is fair?

b. (3 pts) Which is more likely: rolling two dice 24 times and getting at least one double-sixes, or rolling one die four times and getting at least one six?

c. (3 pts) A multiple choice test has 20 questions, each with 3 answers, exactly one of which is correct. What's the probability of scoring 65% or better by completely random guessing?

d. (3 pts) A town has 1000 births in a particular year. If the probability of a boy for any particular birth is 51.3%, what is the probability that between 49.3% and 53.3% (inclusive) of the town's births are male, assuming all births are independent?

e. (3 pts) A city has 10000 births in a particular year. If the probability of a boy for any particular birth is 51.3%, what is the probability that between 49.3% and 53.3% (inclusive) of the city's births are male, assuming all births are independent?

f. (3 pts) Suppose you roll a ten-sided die, numbered 0 through 9 to select digits for a random number. The first roll is the ones digit, and subsequent rolls generate successive decimal places. What's the probability that after 10 rolls, you have an integer? After 100 rolls? 1000?