Homework #1

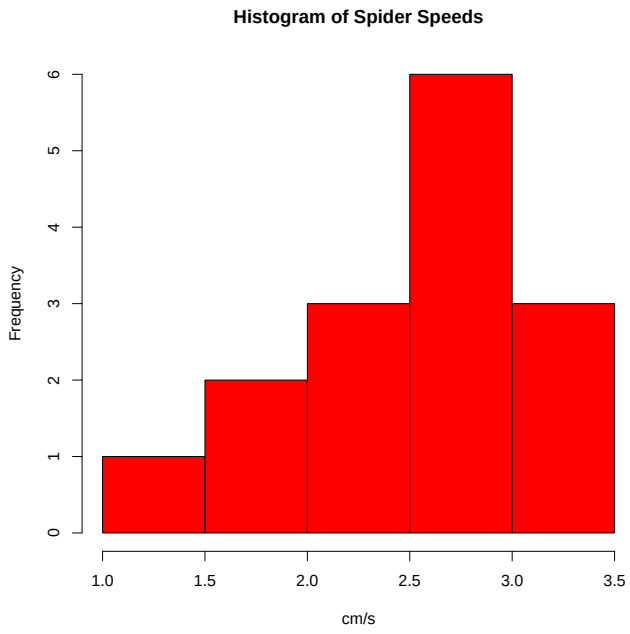
For a problem that asks you to use R, include a copy of the code and output.

1. Consider the following data on the running speed (cm/s) of a certain species of spider:

   1.25, 1.64, 1.91, 2.31, 2.37, 2.38, 2.84, 2.87, 2.93, 2.94, 2.98, 3.00, 3.09, 3.22, 3.41

   (a) Use R to create a frequency histogram of the data.

   ANSWER:
   One possible histogram is below:



   Here is R code to create this histogram:

```
#enter data#
speeds <- c(1.25, 1.64, 1.91, 2.31, 2.37, 2.38, 2.84, 2.87, 2.93, 2.94, 2.98,
3.00, 3.09, 3.22, 3.41)

#make histogram#
hist(speeds, col = 2, main = "Histogram of Spider Speeds", xlab = "cm/s")
```

   The default number of bins seems ok.

(b) Use a calculator to computer the mean and standard deviation for the data. Make sure to include units. Say what these values mean. Then use R to compute them. Do the answers agree?

ANSWER:

The mean and sd, as computed by R, are 2.609 cm/s and 0.618 cm/s, respectively. Here is R code to compute them and its output:

```
> mean(speeds)
[1] 2.609333

> sd(speeds)
[1] 0.6178473

> summary(speeds) # Here's another way to see the mean.
   Min.  1st Qu.  Median  Mean   3rd Qu.   Max.
  1.250   2.340   2.870   2.609   2.990   3.410
```

The mean indicates the center (or balance point) of the data. The sd indicates the spread.

(c) Find the three quartiles for the data. Then use R to compute them. Do the answers agree?

ANSWER:

There are $n = 15$ speeds. Note that they are already sorted. Since n is odd, the median, M (which is the same as Q_2), is at position $\frac{15+1}{2} = 8$ with value `speeds[8]` = 2.87 cm/s, which agrees with R (using code as for problem (1b)).

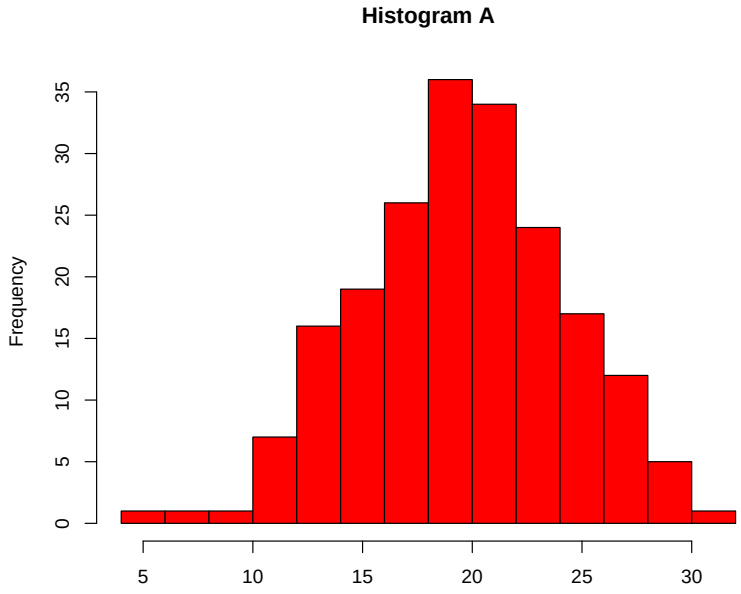
When n is odd, we compute Q_1 as the median of the first half of the data, including the median. This is the average of the 4th and 5th observations, or 2.34 cm/s, agreeing with R. Q_3 is the median of the second half, including the median. This is the average of the 11th and 12th observations, or 2.99 cm/s, agreeing with R. (With other data sets, by-hand computations of Q_1 and Q_3 by our rules may not agree with R. Try, for example, `data = c(1, 2); summary(data)`. There are many defintions of the quartiles. They all approximately agree on large data sets, but they differ in their details.)

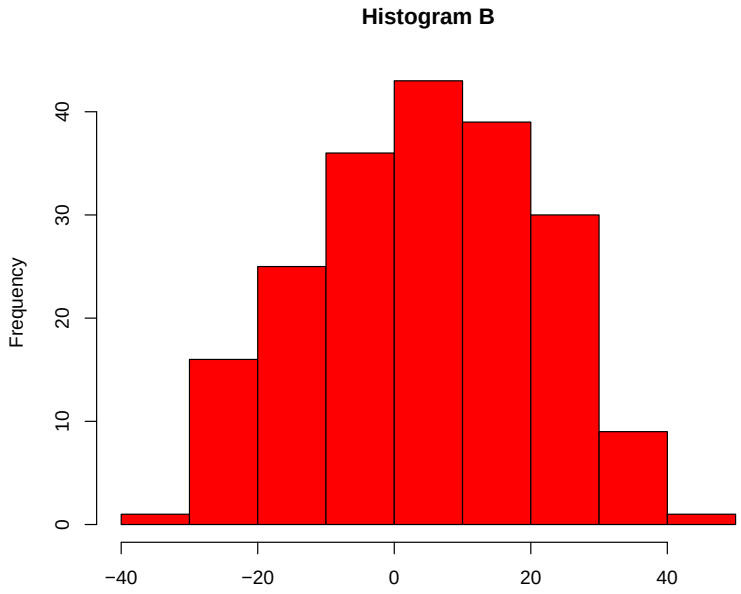
(d) Use R to compute the range and IQR for the data.

ANSWER:

From the output of `summary()` in problem (1b), R computes the maximum of 3.41 cm/s and minimum of 1.25 cm/s, thus the range is $3.41 - 1.25 = 2.16$ cm/s. The IQR is the third quartile minus the first quartile, or $2.99 - 2.34 = 0.65$ cm/s.

2. Consider the following two histograms:





(a) Say which of the two histograms has a larger mean, or if they are about the same. Explain your answer.

ANSWER:
Histogram A has the larger mean. The center (or balance point) of A is about 20, while the center of B is about 5.

(b) Say which of the two histograms has a larger standard deviation, or if they are about the same. Explain your answer.

ANSWER:

Histogram B has the larger standard deviation. A's data ranges from 5 to 30, while B's data ranges from -40 to 40; so there is much more spread in B. This problem shows the importance of including an x-axis on a histogram, and choosing the scale carefully. The mean and sd would have been easier to compare if the scales had been the same, as shown below: