Exercises

1. The following ten observations, taken during the years 1970-79, are on October snow cover for Eurasia. (Snow cover is in millions of square kilometers):

```
year  snow.cover
1970     6.5
1971     12.0
1972     14.9
1973     10.0
1974     10.7
1975     7.9
1976     21.9
1977     12.5
1978     14.5
1979     9.2
```

   i. Enter the data into R.

   ii. Plot `snow.cover` versus year.

   iii Use the `hist()` command to plot a histogram of the snow cover values.

   iv. Repeat ii and iii after taking logarithms of snow cover.

2. For each of the following code sequences, predict the result. Then do the computation:

   a)
   ```
   answer <- 0
   for (j in 3:5){ answer <- j+answer }
   ```

   b)
   ```
   answer<- 10
   for (j in 3:5){ answer <- j+answer }
   ```

   c)
   ```
   answer <- 10
   for (j in 3:5){ answer <- j*answer }
   ```

3. Look up the help for the function `prod()`, and use `prod()` to do the calculation in 2(c) above. Alternatively, how would you expect `prod()` to work? Try it!

4. Add up all the numbers from 1 to 100 in two different ways: using `for` and using `sum`. Now apply the function to the sequence 1:100. What is its action?

5. Multiply all the numbers from 1 to 50 in two different ways: using `for` and using `prod`.

6. The volume of a sphere of radius r is given by $4\pi r^3/3$. For spheres having radii 3, 4, 5, …, 20 find the corresponding volumes and print the results out in a table. Use the technique of section 2.1.5 to construct a data frame with columns `radius` and `volume`.

7. Use `sapply()` to apply the function `is.factor` to each column of the supplied data frame `tinting`. For each of the columns that are identified as factors, determine the levels. Which columns are ordered factors? [Use `is.ordered()`].

   [To access this data frame, download it from http://wwwmath.anu.edu.au/~johnm/r/dsets/usingR.RData in the folder of R and specify `attach("usingR.RData")`]

8. Plot the graph of brain weight (brain) versus body weight (body) for the data set Animals from the MASS package. Label the axes appropriately.

   [To access this data frame, specify library(MASS); data(Animals)]

9. Repeat the plot 8, but this time plotting log(brain weight) versus log(body weight). Use the row labels to label the points with the three largest body weight values. Label the axes in untransformed units.

10. Repeat the plots 8 and 9, but this time, place the plots side by side on the one page.

11. Check the distributions of head lengths (`hdlngth`) in the `possum` data set.
    Compare the following forms of display:
    a) a histogram (`hist(possum$hdlngth)`);
    b) a stem and leaf plot (`stem(possum$hdlngth)`);
    c) a normal probability plot (`qqnorm(possum$hdlngth)`); and
    d) a density plot (`plot(density(possum$hdlngth))`).
    What are the advantages and disadvantages of these different forms of display?

    [To access this data frame, download it from http://wwwmath.anu.edu.au/~johnm/r/dsets/usingR.RData in the folder of R and specify `attach("usingR.RData")`]

12. Try `x <- rnorm(10)`. Print out the numbers that you get. Look up the help for `rnorm`. Now generate a sample of size 10 from a normal distribution with mean 170 and standard deviation 4.

13. Use mfrow() to set up the layout for a 3 by 4 array of plots. In the first row, show normal probability plots for four separate 'random' samples of size 10, all from a normal distribution. In the second row, display plots for samples of size 100. In the last row, display plots for samples of size 1000. Comment on how the appearance of the plots changes as the sample size changes.

14. The function `runif()` can be used to generate a sample from a uniform distribution, by default on the interval 0 to 1. Try `x <- runif(10)`, and print out the numbers you get. Then repeat exercise 13 above, but taking samples from a uniform distribution rather than from a normal distribution. What shape do the points follow?

15. If you find exercise 14 interesting, you might like to try it for some further distributions. For example `x <-rchisq(10,1)` will generate 10 random values from a chi-squared distribution with degrees of freedom 1.The statement `x<-rt(10,1)` will generate 10 random values from a t distribution with degrees of freedom one. Make normal probability plots for samples of various sizes from these distributions.

16. For the first two columns of the data frame hills, examine the distribution using:

    (a) histograms

    (b) density plots

    (c) normal probability plots.

    Repeat (a), (b) and (c), now working with the logarithms of the data values.