Exercise 1: There is a dataset called banknote_authentication that is in the package nonet, from Dr. Volker Lohweg. It is also available at the UCI machine learning repository
The variables are:

1. variance of wavelet transformed image
2. skewness of wavelet transformed image
3. kurtosis of wavelet transformed image
4. entropy of image
5. class (0 is genuine, 1 is forgery).

This research started with a classical approach to image analysis. You take the image, convert it to grey scale, then apply a mathematical transform. In this case the transform is a wavelet, other options include Fourier transforms of various types. They encode information about similarity of close-together pixels. To give you an idea of how wavelet transforms work (in the simplest possible case), I'll have you do a 1-D wavelet transform of a particularly simple kind - the Haar transform. Consider the following series of values:

$$3, 4, 3, 2, 1, 5, 6, 6$$

And the Haar weights:

```
s2 <- sqrt(2)
s4 <- sqrt(4)
HaarWt <- matrix(ncol=8, byrow=TRUE,
          c(1, 1, 1, 1, 1, 1, 1, 1,
            1, 1, 1, 1, -1, -1, -1, -1,
            s2, s2, -s2, -s2, 0, 0, 0, 0,
            0, 0, 0, 0, s2, s2, -s2, -s2,
            s4, -s4, 0, 0, 0, 0, 0, 0,
            0, 0, s4, -s4, 0, 0, 0, 0,
            0, 0, 0, 0, s4, -s4, 0, 0,
            0, 0, 0, 0, 0, 0, s4, -s4))
x <- c(5,4,5,3,1,1,2,1)
Haar_transform <- HaarWt%*%x
Haar_transform
```

```
##          [,1]
## [1,] 22.000000
## [2,] 12.000000
## [3,] 1.414214
## [4,] -1.414214
## [5,] 2.000000
## [6,] 4.000000
## [7,] 0.000000
## [8,] 2.000000
```

Essentially each row of the matrix is multiplied by x and the values added up (i.e. a dot product). Notice the large value for the 'large wavelength' second wavelet value. This is because there are large values clumped in the first half of x and smaller values in the second half of x. This is typical of wavelets, which have a frequency (how often they vary) and are local (based on only a small chunk of the data). Note that the first transform isn't really worth looking at, as it's just an sum over everything.

a Try to explain the pattern if you can guess it, or look it up?

Solution:

b Repeat the Haar transform for the following dataset and interpret the wavelets:

```
x <- c(5,6,1,2,8,9,2,2)
```

You'll notice that (ignoring the first one) two of the transform values are particularly large. What in the data are they detecting?

Solution: