**Statistics 787** **Assignment 2** **Due: September 14**

1. Write an R function which will draw population pyramids (back-to-back histograms). Use the following data set (available on the class website) to test your function. The data set gives the number of people in New Zealand in 1996, broken down by age group and gender.

| | *Gender* | |
|---|---|---|
| *Age Group* | *Female* | *Male* |
| 0-4 | 135489 | 144111 |
| 5-9 | 140571 | 147723 |
| 10-14 | 128520 | 135663 |
| 15-19 | 129405 | 133572 |
| 20-24 | 136926 | 134832 |
| 25-29 | 140850 | 132453 |
| 30-34 | 151032 | 142452 |
| 35-39 | 145923 | 139293 |
| 40-44 | 129597 | 125439 |
| 45-49 | 120942 | 120249 |
| 50-54 | 93366 | 93351 |
| 55-59 | 79821 | 78783 |
| 60-64 | 67845 | 67419 |
| 65-69 | 67788 | 65187 |
| 70-74 | 61902 | 51762 |
| 75-79 | 48726 | 33561 |
| 80-84 | 34872 | 20409 |
| 85-89 | 18645 | 8385 |
| 90-94 | 6840 | 2385 |
| 95+ | 1782 | 426 |

   Your function should work for this and for similar data sets. The function should take the cell breakpoints and bars heights for the two populations being compared as arguments. There should also be arguments which allow customisation of the axis annotation and general labelling.

   (Hint: using `layout` will be helpful.)

2. Write an R function which produces superimposed histograms. The function should take as arguments: a vector of cell breakpoints, a matrix of bar heights (each column corresponding to a different population group) and arguments which make it possible to customise the appearance of the plot. (You don't need `layout` for this.

3. Write an R function which operates like that in question 2, but produces a smoothed version of the histogams. You will probably find the `splinefun` function useful. The expression

```
s = splinefun(x, y)
```

obtains the function which interpolates the values in `x` and `y`. The derivative of `s` can be plotted as follows:

```
x = seq(min(x), max(x), length = 1001)
y = s(x, deriv = 1)
plot(x, y, type = "l")
```

You can produce the smoothed histograms by computing the corresponding distribution function at the cell breakpoints, fitting a spline through this to obtain a smooth distribution function and then plotting the derivative of this distribution function.