Using R to Plot Histograms (II)

The `barplot` and `hist` commands in `R` can be used to plot histograms. The latter is used when the raw data is available. `barplot` is used when we already have the data organized into a table.

1. The distribution of families by income in the U.S. in 1973 is tabulated below.

Income level	Percent
\$0–\$1,000	1
\$1,000–\$2,000	2
\$2,000–\$3,000	3
\$3,000–\$4,000	4
\$4,000–\$5,000	5
\$5,000–\$5,000	5
\$6,000–\$7,000	5
\$7,000–\$10,000	15
\$10,000–\$15,000	26
\$15,000–\$25,000	26
\$25,000–\$50,000	8
\$50,000 and over	1

The percentages give the areas of the boxes of a histogram. The class intervals (income level intervals) give the widths. `> incomePercents <- c(1,2,3,4,5,5,5,15,26,26,8)`

Let's enter the widths (in units of thousands of dollars) of the class intervals.

```
> intervalWidths <- c(1,1,1,1,1,1,1,3,5,10,25)
```

The heights of the boxes in the histogram should be measured in 'percent per thousand dollars' so that areas are percentages. We compute these by dividing the percent by the width of the corresponding class interval. We do this as one (vector) step in the first argument to `barplot` below.

```
> barplot(incomePercents/intervalWidths, intervalWidths, space=0,
   xlab="Income (thousands of dollars)", ylab="Percent per Thousand Dollars", cex.axis=.75)
> axis(1, seq(0,50,by=5), cex.axis=0.75)
```

The output produced is shown below in Figure 1. Let's describe the parameters passed to `barplot`.

- `incomePercents/intervalWidths` heights of the boxes
- `intervalWidths` widths of the class intervals
- `space=0` space (none) included between the bars
- `xlab="Income (thousands of dollars)"` label to put below the x axis
- `ylab="Percent per Thousand Dollars"` label to put left of the y axis
- `cex.axis=.75` reduce the size of numbers along the axes by 25%

After we use `barplot`, we use the `axis` command to put a income measurements along the x-axis. The parameters passed to `axis` are described below.

- `1` apply this to the x-axis
- `seq(0,50,by=5)` put tick marks from 0 to 50 in intervals of 5 units
- `cex.axis=.75` reduce the size of numbers along the axes by 25%

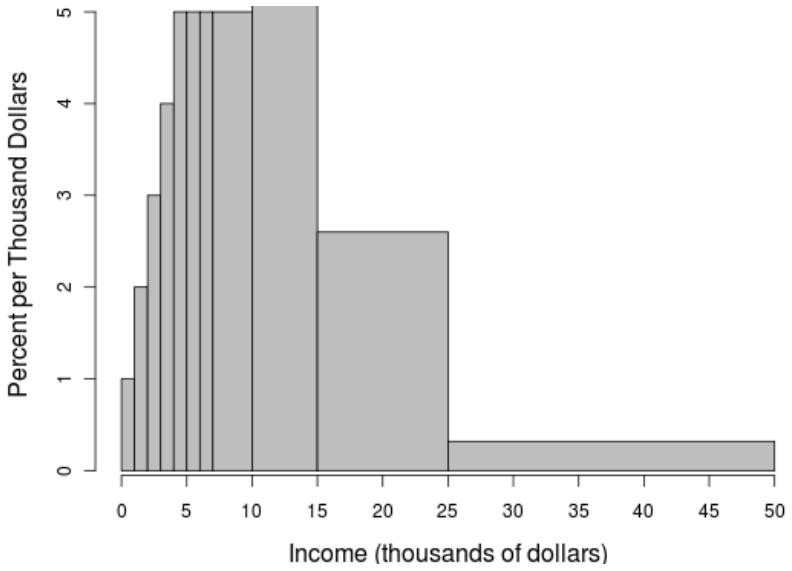

Figure 1: Family income distribution histogram from Figure 4 on page 37 of our text

2. The table below gives the distribution of educational level for persons age 25 and over in the U.S. in 1960, 1970 and 1971.

Educational level (years of schooling)	1960	1970	1991
0–5	8	6	2
5–8	14	10	4
8–9	18	13	4
9–12	19	19	11
12–13	25	31	39
13–16	9	11	18
16 or more	8	11	21

a) Make a histogram for the 1960 data.

```
> educationLevels <- c(2, 4, __, __, __, __, __)
> intervalWidths <- c(5, 3, __, __, __, __, 1)
> barplot(educationLevels/intervalWidths, _____ , space=0,
   xlab="Education Level (years)", ylab="Percent per _____", cex.axis=.75)
> axis(1, seq(0,17,by=1), cex.axis=0.75)
```

b) Repeat a) for the 1970 and 1991 data.

c) Discuss any interesting features that you see in the histograms.