childsmath $\int_M d\omega = \int_{\partial M} \omega$

R Tutorials

Stem and Leaf Plots on R

We will use the `weight` variable in the databank data set (which is given in **pounds**). We will construct a stem-and-leaf plot of the `weight` variable in **kilograms**. We first convert to kilograms by multiplying the weight in pounds by 0.4536:

1. Open the `databank.RData` workspace that you created in the introductory lab using `File->Load Workspace` (on a **Mac:** `workspace > load workspace file`).
2. We first convert the weight column to kilograms:

```
weight<-databank$weight
weight_kg<-weight*0.4536
```

3. To display the stem-and-leaf plot, we use the `stem.leaf` function, which is part of the `aplpack` package. (**Note:** If the below command doesn't work, then it could be because you don't have the `aplpack` package installed. Go back to the introductory lab and install it.)

```
library("aplpack")
stem.leaf(weight_kg,na.rm=T,trim.outliers=F)

1 | 2: represents 12
 leaf unit: 1
            n: 100
    1     4* | 4
    8     4. | 5888999
   21     5* | 0000002333444
   36     5. | 555667888899999
   45     6* | 011444444
  (11)    6. | 56777788889
   44     7* | 00012233333344
   30     7. | 555677889
   21     8* | 111233344
   12     8. | 5666778
    5     9* | 03
    3     9. | 67
         10* |
    1    10. | 6
```

Note that `stem.leaf` has used the intervals 40-45, 45-50, 50-55, 55-60, etc. i.e., each stem has been separated into `m` = 2 parts. We have used the `na.rm=T` option so that `NA's` are removed from the data, and we have used the `trim.outliers=F` option so that R does not remove outliers when creating the stem and leaf plot.

4. If we want to use intervals of length 10, we have to split each stem into `m` = 1 part:

```
stem.leaf(weight_kg,na.rm=T,trim.outliers=F,m=1)

1 | 2: represents 12
 leaf unit: 1
            n: 100
    8     4 | 45888999
   36     5 | 00000023334445556678888999999
  (20)    6 | 01144444456777788889
   44     7 | 000122333333445556778889
   21     8 | 1112333445666778
    5     9 | 0367
```

```
      1    10 | 6
```

Explanation of Depths: The column of numbers at the left of the plot are called **depths**. The brackets around `(20)` mean two things. First of all, it means that there are 20 observations in that row, i.e., 20 observations between 60 and 70 (not including 70). Secondly, it means that the median is in that row, i.e., the median is somewhere between 60 and 70 (not including 70). The numbers **above** `(20)`, i.e., `36` and `8` give the number of observations that are **less than or equal to** the numbers in that row. So, for example, the `36` means that there are 36 observations that are less than or equal to 59. i.e., there are a total of 36 observations in the **first two rows combined**. The numbers **below** `(20)`, i.e., `44`, `21`, `5`, and `1` give the number of observations that are **greater than or equal to** the numbers in that row. So, for example there are 44 observations that are greater than or equal to 70. i.e., there are a total of 44 observations in the **last four rows combined**.