Lab 2 - Data Handling in R

Contents

Lab Objectives

- Use a range of functions in R to calculate descriptive statistics, such as mean, median, and standard deviation.
- Use R to produce a properly formatted bar chart.
- Use R to produce a frequency table.

A sample of useful functions available in R: Review the functions in the following table.

Function	What it does	Example code
+	Adds values	X+Y
-	Subtracts values	Y-X
*	Multiplies values	X*Y
/	Divides values	Y/X
^	Raises a value to an exponent	X^Y
sqrt	Calculates the square root	sqrt(Y)
sum	Calculates the sum of a dataset	sum(X,Y)
mean	Average of a dataset	mean(Z)
median	Median (middle) of a dataset	median(Z)
sd	Standard deviation of a dataset	sd(Z)
length	Counts the size of dataset	length(Z)
min	Returns the minimum number of a dataset	min(Z)
max	Returns the maximum number of a dataset	max(Z)
rank	Returns the rank value all numbers in a dataset	rank(Z)

Replace the letters X and Y and Z with actual numbers and try the various functions out in the console window of RStudio:

```
X<-5
Y<-7
Z<-c(3,5,6,8,9,2,12)
```

A useful reference list of functions available in R can be found here: http://www.sr.bham.ac.uk/~ajrs/R/r-function_list.html

---

Exercise 1: Data Entry

Student	Midterm	Final	Labs
A	67	54	91
B	70	57	87
C	77	65	77
D	88	66	74
E	56	67	56
F	90	89	96
G	88	80	76

You are provided with the data above and you want to enter this into R. The easier approach would be to

save it in a spreadsheet and export that spreadsheet as a file, usually a comma-delimited file (.csv). But we will do this manually here by creating 4 variables and bringing those variables together as a table.

Each variable is "built" by concatenating them together using the **c()** function.

Then we bring all variables together into a table. In R, this will be called a data.frame:

```
Student<-c("A", "B", "C", "D", "E", "F", "G")
Midterm<-c(67,70,77,88,56,90,88)
Final<-c(54,57,65,66,67,89,80)
Labs<-c(91,87,77,74,56,96,76)
d<-data.frame(Student, Midterm, Final, Labs)
d
```

```
##   Student Midterm Final Labs
## 1       A      67    54   91
## 2       B      70    57   87
## 3       C      77    65   77
## 4       D      88    66   74
## 5       E      56    67   56
## 6       F      90    89   96
## 7       G      88    80   76
```

Exercise X: Sorting values and manually determining median

Given the following data:

```
x<-c(7, 36, 45, 9, 23, 25, 39, 12, 19, 21)
```

Sort them to a new variable, y, using the **sort()** function:

```
##  [1]  7  9 12 19 21 23 25 36 39 45
```

Use the appropriate formulas from class to calculate which observation (or average observations) corresponds to the median, first quartile, and third quartile.
