Data Science for Business With R - CH01

Creating and using vectors

This is a vector

```
c(43,42,12,8,5)
```

```
## [1] 43 42 12  8  5
```

Assign the vector to a variable

```
myFamilyAges <- c(43,42,12,8,5)
```

Find the sum of the values in the vector

```
sum(myFamilyAges)
```

```
## [1] 110
```

Find the range of ages in the vector

```
myRange <- range(myFamilyAges)

myRange
```

```
## [1]  5 43
```

Subsetting Vectors

Select specific values in the vector by using the index of each value.

```
myFamilyAges[ c(3,2,5)]
```

```
## [1] 12 42  5
```

Select a sequential range of values on the vestor

```
myFamilyAges[ c(3:5) ]
```

```
## [1] 12  8  5
```

Remove elements of the vector with negative indexes

```
myFamilyAges[ c(-3:-5) ]
```

```
## [1] 43 42
```

Elements of a vector can also be select using TRUE and FALSE

```
myFamilyAges[ c(TRUE, FALSE, TRUE, FALSE, FALSE) ]
```

```
## [1] 43 12
```

TRUEs and FALSEs can be loaded into a variable

```
selectedFamily <- c(TRUE, FALSE, TRUE, FALSE, FALSE)
myFamilyAges[ selectedFamily ]
```

```
## [1] 43 12
```

TRUEs and FALSEs can be input from a conditional evaluation

```
myFamilyAges > 21
```

```
## [1]  TRUE  TRUE FALSE FALSE FALSE
```

```
selectedFamily <- myFamilyAges > 21

myFamilyAges[ selectedFamily ]
```

```
## [1] 43 42
```

The conditional evaluation can also be done inside the square brackets

```
myFamilyAges[ myFamilyAges == 12 ]
```

```
## [1] 12
```

```
myFamilyAges[ myFamilyAges != 12 ]
```

```
## [1] 43 42  8  5
```

```
myFamilyAges[ !(myFamilyAges == 12) ]
```

```
## [1] 43 42  8  5
```

Multiple conditional evaluations can be done within one expression

```
myFamilyAges[ myFamilyAges > 6 &  myFamilyAges < 20 ]
```

```
## [1] 12  8
```

CASE STUDY: Calculating NPS (Net Promoter Score)

- case key points
  - Define a vector that represents likelihood to recommend
  - Calculate the number of promoters and detractors
  - Calculate the net promoter score (NPS)

```
#--- define a test vector
ltr <- c(9,8,3,9,7,8,9,6,7,8,9)

#--- what is the range of the ltr vector
range(ltr)
```

```
## [1] 3 9
```

```
#--- create a new vector with just the promoters
#--- then calculate the length of the promoters vector
promoters <- ltr[ltr>8]
numPromoters <- length(promoters)

#--- Calculate the number of detractors by summing the
#--- elements that are less than 7
detractorsTrueFalse <- ltr < 7
numDetractors <- sum(detractorsTrueFalse)

#--- Calculate NPS, based on the length of the ltr vector
#--- and the number of promoters and detractors
total <- length(ltr)
nps <- (numPromoters/total - numDetractors/total)*100

#--- output
nps
```

```
## [1] 18.18182
```

Chapter Challenges

1. Use c() to ad another family member's age to the end of the myFamilyAges vector

```
myFamilyAges <- c(myFamilyAges, 17)
```

2. Use square bracket subsetting to the first element of the myFamilyAges vector

```r
myFamilyAges[1]
```

```
## [1] 43
```

3. Use square bracket subsetting to show the odd numbered elements

```r
myFamilyAges[c(TRUE,FALSE)]
```

```
## [1] 43 12  5
```

4. Create a conditional expression that outputs a set of TRUEs ans FALSEs. The expression should show TRUE when an element of myFamilyAges = 12

```r
myFamilyAges == 12
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE FALSE
```

5. put an exclamation point in front of the expression from challenge 4

```r
!myFamilyAges == 12
```

```
## [1]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
```

6. Using the previous expression select the values of myFamilyAges that != 12

```r
myFamilyAges[!myFamilyAges == 12]
```

```
## [1] 43 42  8  5 17
```

7. Using the Nile dataset, create a conditional expression that shows TRUE where the Nile observation is greater than 900. Then use sum() to count the number of these observances.

```r
myNile <- Nile
nileHighTF <- myNile > 900

nileHigh <- sum(nileHighTF)
nileHigh
```

```
## [1] 49
```

```r
nileLow <- sum(!nileHighTF)
nileLow
```

```
## [1] 51
```