Intro to Data Science - HW 1


```
# Enter your name here: Chaithra Kopparam Cheluvaiah
```

Attribution statement: (choose only one and delete the rest)

```
# 1. I did this homework by myself, with help from the book and the professor.
```

Define a variable:

```
x <- 280
```

Define the following vectors, which represent the **population** (in thousands) and **number of colleges** in each of the five counties in Central New York (CNY) – **Cayuga**, **Cortland**, **Madison**, **Onondaga**, and **Oswego**, in this order:

```
population <- c(80, 49, 73, 467, 122)
colleges <- c(2, 2, 3, 9, 2)
```

Part 1: Calculating statistics using R

A. Show the number of observations in the **population** vector with the length() function:

```
length(population)
```

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

B. Show the number of observations in the **colleges** vector with the length() function:

```
length(colleges)
```

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

C. Calculate the average CNY population using the mean() function:

```
mean(population)
```

```
## [1] 158.2
```

D. Calculate the average number of colleges in CNY using the mean() function:

```
mean(colleges)
```

```
## [1] 3.6
```

E. Calculate the total CNY population using the sum() function:

```
sum(population)
```

```
## [1] 791
```

F. Calculate the total number of colleges in CNY using the sum() function:

```
sum(colleges)
```

```
## [1] 18
```

G. Calculate the average CNY population again, this time using **the results from steps A & E**:

```
sum(population)/length(population)
```

```
## [1] 158.2
```

H. Calculate the average number of colleges in CNY again, this time using **the results from steps B & F**:

```
sum(colleges)/length(colleges)
```

```
## [1] 3.6
```

Part 2: Using the max/min and range functions in {r}

I. How many colleges does the county with most colleges have? Hint: Use the max() function:

```
max(colleges)
```

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

J. What is the population of the least populous county in CNY? **Hint:** Use the min() function:

```
min(population)
```

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

K. Display the populations of the least populous and most populous county in the dataset together. **Hint:** Use the range() function:

```
range(population)
```

```
## [1]  49 467
```

Part 3: Vector Math

L. Create a new vector called **extraPop**, which is the current population of a county **+ 50** (each county has 50,000 more people):

```
extraPop <- population + 50
extraPop
```

```
## [1] 130  99 123 517 172
```

M. Calculate the average of **extraPop**:

```
mean(extraPop)
```

```
## [1] 208.2
```

N. In a variable called **bigCounties**, store all the population numbers from the original **population** vector which are **greater than 120** (using **subsetting** in R):

```
bigCounties <- population[population > 120]
bigCounties
```

```
## [1] 467 122
```

O. Report the length of **bigCounties**:

```
length(bigCounties)
```

```
## [1] 2
```