Assignment 1

Chang An Le Harry Jr

1/19/2022

Question 1: Basics in R programming

1. Create a vector of all integers from 2 to 100, and save it as x1.

```
x1 <- c(2:100)
x1
```

```
##  [1]   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
## [20]  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39
## [39]  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58
## [58]  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77
## [77]  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96
## [96]  97  98  99 100
```

2. Create a vector of all even integers from 2 to 100, and save it as x2.

```
x2 <- seq(2,100,2)
x2
```

```
##  [1]   2   4   6   8  10  12  14  16  18  20  22  24  26  28  30  32  34  36  38
## [20]  40  42  44  46  48  50  52  54  56  58  60  62  64  66  68  70  72  74  76
## [39]  78  80  82  84  86  88  90  92  94  96  98 100
```

3. Compute the sum of x1 and x2.

```
x1 + x2
```

```
##  [1]   4   7  10  13  16  19  22  25  28  31  34  37  40  43  46  49  52  55  58
## [20]  61  64  67  70  73  76  79  82  85  88  91  94  97 100 103 106 109 112 115
## [39] 118 121 124 127 130 133 136 139 142 145 148 151  54  57  60  63  66  69  72
## [58]  75  78  81  84  87  90  93  96  99 102 105 108 111 114 117 120 123 126 129
## [77] 132 135 138 141 144 147 150 153 156 159 162 165 168 171 174 177 180 183 186
## [96] 189 192 195 198
```

4. What do the commands sum(x1) and length(x1) do? Use these commands to compute the average of all values in x1.

```
sum(x1)
```

```
## [1] 5049
```

```
length(x1)
```

```
## [1] 99
```

- sum(x1) computes the sum of all integers in the vector x1 (i.e $2 + 3 + 4 + \ldots + 99 + 100 = 5049$)
- length(x1) prints the total number of integers in the vector x1 (99)

```
ans <- sum(x1) / length(x1)
ans
```

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

This mean value can also be verified using the code below:

```
mean(x1)
```

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

5. The formula for the sum of the first n positive integers is n(n+1)/2. Compute the sum of all integers from 1 to 2101 using this formula.

```
2101*(2101+1)/2
```

```
## [1] 2208151
```

Question 2: Revisit the cars data

1. Read in the dataset cars and name it as data1. Read the data help files and briefly describe the definition of each variable.

```
data1 <- cars
?cars
```

The dataset has 2 variables: 'speed' and 'dist'.

- 'speed' refers to the speed of the car in each observation, measured in mph
- 'dist' refers to the stopping distance that the car requires while travelling at a particular speed, measured in ft

2. How many observations and how many variables are there in this dataset? Write code to answer this question.

```
dim(data1)
```

```
## [1] 50  2
```

3. Compute the summary statistics of all variables in the dataset using the summary() function. Based on the statistics, determine whether the mean or the median is a better measure of the center for each variable.

```
summary(data1)
```

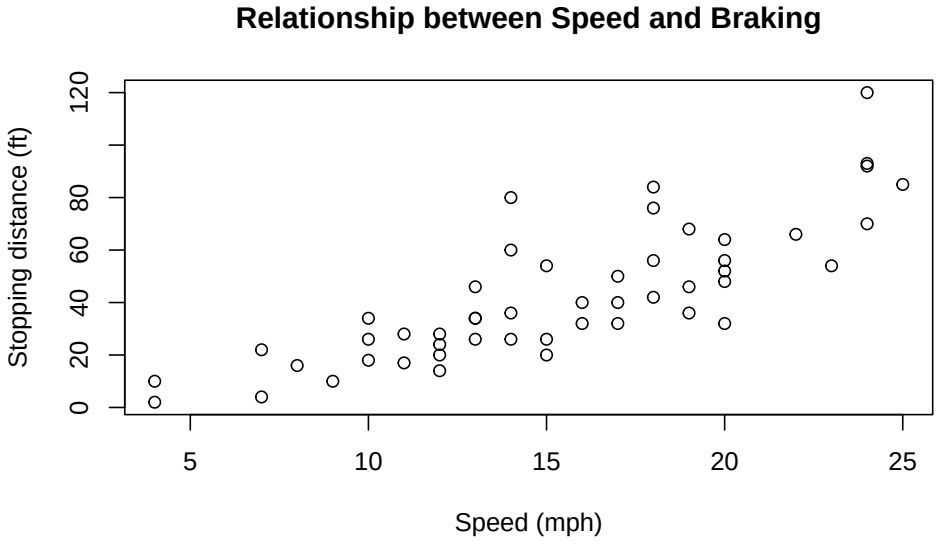
```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

- Mean is a better measure for the centre of the 'speed' variable as the data points are generally more evenly distributed amongst each other.

- Median is a better measure for the centre of the 'dist' variable due to the outlier value of 120 which may greatly affect the value of the mean. Thus in this case median would be the more appropriate measure.

4. Create a scatterplot to visualize the relationship between the variables in the dataset. Add appropriate titles to your chart and axes. Briefly interpret the scatterplot (e.g trends, relationships, etc)

```
plot(data1$speed, data1$dist,
     xlab = "Speed (mph)", ylab = "Stopping distance (ft)",
     main = "Relationship between Speed and Braking")
```



From the scatterplot, it seems like there's generally a positive relationship between speed and stopping distance. When the speed of the car increases, the stopping distance required by the car to stop increases.