1. (a) Read the following data into R in three variables called Group1, Group 2 and Group 3 :

|  |  |  |
| --- | --- | --- |
| Group 1 | Group 2 | Group 3 |
| 22 | 21 | 48 |
| 22 | 23 | 45 |
| 28 | 21 | 51 |
| 23 | 24 | 29 |
| 45 | 22 | 38 |
| 21 | 20 | 40 |
| 22 | 21 | 38 |

(b) Use the function summary to calculate some basic statistics about these variables. Which variable has the highest measures of location?

(c) Find the variable with the highest variability

(d) Create a boxplot for each of the three variables (a side-by-side plot with all three boxplots in one

graphic). Analyse the outputs of the boxplot and the statistical output

(e) Create a variable with the name total, which consists of the three variables Group1, Group2 and

Group3 appended in this order

(f) Create a histogram of total

(g) Construct a variable called total\_big which contains the elements of total that are greater than

or equal to 25. Find the number of elements in big total

2. (a) Create a vector called d, consisting of the numbers 0 to 99.

(b) Create a vector called e, consisting of the 100 randomly chosen numbers from a normal distribution with expectation 3 and standard deviation 4

(c) Create a vector f which is the addition of vectors d and f

(d) Create a histogram each for d, e and f

(e) Extract the 65th element of f

3. Use the file “Clothing Store” for this exercise

(a) Find the variable names in the dataset. Count the number of observations and variable names in the dataset

(b) Randomly select 70% into a dataset and the remaining 30% to a dataset

(c) Find the variable with 2 unique levels. Also, find the variable with maximum levels

(d) The target variable is the variable RESP. Construct a histogram to see the distribution of this variable

(e) Remove the variables HHKEY, ZIP\_CODE