

```
1 # STAT 3093 Assignment 1
2 # Question 4 (Chapter 1, Question 52)
3 # Albert Lockett (3254354, k44if@unb.ca)
4
5 q4_data <- c(
6   389, 356, 359, 363, 375, 424, 325, 394, 402,
7   373, 373, 370, 364, 366, 364, 325, 339, 393,
8   392, 369, 374, 359, 356, 403, 334, 397
9 )
10 q4_data <- sort(q4_data)
11 n <- length(q4_data)
12
13 # Part A)
14 lower_fourth <- median(q4_data[1:(n/2)])
15 upper_fourth <- median(q4_data[(n/2 + 1):n])
16 fourth_spread <- upper_fourth - lower_fourth
17
18 print('part a')
19 print(sprintf('The lower forth is: %.1f', lower_fourth))
20 print(sprintf('The upper forth is: %.1f', upper_fourth))
21 print(sprintf('The fourth spread is: %.1f', fourth_spread))
22 # [1] "part a)"
23 # [1] "The lower forth is: 359.0"
24 # [1] "The upper forth is: 392.0"
25 # [1] "The fourth spread is: 33.0"
26
27
28 # Part B)
29 upper_outlier_val <- upper_fourth + 1.5 * fourth_spread
30 lower_outlier_val <- lower_fourth - 1.5 * fourth_spread
31 upper_outliers <- q4_data[q4_data > upper_outlier_val]
32 lower_outliers <- q4_data[q4_data < lower_outlier_val]
33 num_outliers <- length(upper_outliers) + length(lower_outliers)
34
35 upper_extreme_outlier_val <- upper_fourth + 3 * fourth_spread
36 lower_extreme_outlier_val <- lower_fourth - 3 * fourth_spread
37 upper_extreme_outliers <- q4_data[q4_data > upper_extreme_outlier_val]
38 lower_extreme_outliers <- q4_data[q4_data < lower_extreme_outlier_val]
39 num_extreme_outliers <- length(upper_extreme_outliers) +
40   length(lower_extreme_outliers)
41
42 print('part b')
43 print(
44   sprintf(
45     'outliers will have value > %.1f or < %.1f',
46     upper_outlier_val,
47     lower_outlier_val
48   )
49 )
50 print(sprintf('There are %d outliers:', num_outliers))
51 print(c(upper_outliers, lower_outliers))
52
53 print(
54   sprintf(
55     'extreme outliers will have value > %.1f or < %.1f',
56     upper_extreme_outlier_val,
57     lower_extreme_outlier_val
58   )
59 )
60 print(sprintf('There are %d extreme outliers:', num_extreme_outliers))
61 print(c(upper_extreme_outliers, lower_extreme_outliers))
```

```
60
61 # [1] "part b)"
62 # [1] "outliers will have value > 441.5 or < 309.5"
63 # [1] "There are 0 outliers:"
64 # [1] "extreme outliers will have value > 491.0 or < 260.0"
65 # [1] "There are 0 extreme outliers:"
66
67
68 # Part C
69 boxplot(q4_data)
70 # output / comments on next page
71
72 # Part D
73
74 # The fourth spread will not be affected by decreasing the largest
75 # observation until the value has decreased to a value less than the
76 # upper fourth
77
78 # the value will not be affected
79 q4_data[length(q4_data)] <- upper_fourth # largest value is at end of list
80 new_upper_fourth <- median(q4_data[(n/2 + 1):n])
81 print(new_upper_fourth == upper_fourth)
82 # [1] TRUE
83
84 # the value will be affected
85 q4_data[length(q4_data)] <- upper_fourth - 1
86 new_upper_fourth <- median(q4_data[(n/2 + 1):n])
87 print(new_upper_fourth == upper_fourth)
88 # [1] FALSE
89
```

