**Tutorial 4: Code Book**  
Name: Nur Aliah Binti Hasshim  
Matrix No: S2031689  
  
**1. Codebook using standard R API functionality**  
  
**Step 1:** Install “memisc” package  
> install.packages("memisc")   
  
**Step 2:** Loading listing of packages

> library(memisc)  
  
**Step 3:** Assign “faithful” dataset in variable “x” using “data.set” command from “memisc”

> x <- data.set(faithful)  
  
**Step 4:** Check the typeof variable x

> typeof(x)

[1] "list"  
  
**Step 5:** Check number of lines and variables in the dataset with the first 25 lines printed out

> x

Data set with 272 observations and 2 variables

faithful.eruptions faithful.waiting

1 3.600 79

2 1.800 54

3 3.333 74

4 2.283 62

5 4.533 85

6 2.883 55

7 4.700 88

8 3.600 85

9 1.950 51

10 4.350 85

11 1.833 54

12 3.917 84

13 4.200 78

14 1.750 47

15 4.700 83

16 2.167 52

17 1.750 62

18 4.800 84

19 1.600 52

20 4.250 79

21 1.800 51

22 1.750 47

23 3.450 78

24 3.067 69

25 4.533 74

.. .................. ................

(25 of 272 observations shown)  
  
**Step 6:** Use “Codebook” command from “memisc” package. This command will show the Minimum, Maximum, Mean and Standard Deviation value for each variable. In “faithful” dataset, there are only 2 variables.

> codebook(x)

faithful.eruptions

------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Storage mode: double

Measurement: interval

Min: 1.600

Max: 5.100

Mean: 3.488

Std.Dev.: 1.139

==========================================================================================================================================================================================

faithful.waiting

------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Storage mode: double

Measurement: interval

Min: 43.000

Max: 96.000

Mean: 70.897

Std.Dev.: 13.570

**Step 7:** Check the summary of the dataset. “Summary” command will show the Minimum, 1st Quadrant, Median, Mean, 3rd Quadrant and Maximum value for each variables in the dataset.

> summary(x)

faithful.eruptions faithful.waiting

Min. :1.600 Min. :43.0

1st Qu.:2.163 1st Qu.:58.0

Median :4.000 Median :76.0

Mean :3.488 Mean :70.9

3rd Qu.:4.454 3rd Qu.:82.0

Max. :5.100 Max. :96.0

**2. Manually create a custom codebook**  
**Step 1:** Check the class of the dataset  
> class(faithful)

[1] "data.frame"  
  
**Step 2:** Use “sapply” command and “class” function to check the class of each variable.

> sapply(faithful, class)

eruptions waiting

"numeric" "numeric"   
  
**Step 3:** Use “sapply” command and “min” function to check the minimum value of each variable.

> sapply(faithful, min)

eruptions waiting

1.6 43.0   
  
**Step 4:** Use “sapply” command and “max” function to check the maximum value of each variable.

> sapply(faithful, max)

eruptions waiting

5.1 96.0   
  
**Step 5:** Use “sapply” command and “range” function to check the range value of each variable.

> sapply(faithful, range)

eruptions waiting

[1,] 1.6 43

[2,] 5.1 96

**Step 6:** Check the summary of the dataset. “Summary” command will show the Minimum, 1st Quadrant, Median, Mean, 3rd Quadrant and Maximum value for each variables in the dataset.

> summary(faithful)

eruptions waiting

Min. :1.600 Min. :43.0

1st Qu.:2.163 1st Qu.:58.0

Median :4.000 Median :76.0

Mean :3.488 Mean :70.9

3rd Qu.:4.454 3rd Qu.:82.0

Max. :5.100 Max. :96.0  
  
**Conclusions**There are two different ways to create the codebook for any dataset using R. Using “memisc” packages, fewer commands is needed to get the descriptions of the dataset compared to the manual codebook.