

Data Analysis Assignment 4

Aarushi Verma

05/11/2021

Missing Data Mechanics

Question 1

Create a dataset with 30% of the age values missing completely at random, leaving all values of diameter observed. Report the R commands you used to make the dataset. Also report the dataset values after you made the ages missing. (This is so we can tell which cases you made missing.)

First we read in the Treeage dataset into R. We looked at the dimension, structure and summary of the data to understand the columns as well as number of rows present in the data. In order to create a new dataset with 30% of the age values missing completely at random, the following code was used -

```
set.seed(578)
treeage$age[sample(1:nrow(treeage), 0.3*nrow(treeage))] = NA
treeage
```

```
##      number diameter age
## 1         1      12.0 125
## 2         2      11.4 119
## 3         3       7.9  83
## 4         4       9.0  85
## 5         5      10.5  99
## 6         6       7.9 117
## 7         7       7.3  NA
## 8         8      10.2 133
## 9         9      11.7  NA
## 10        10      11.3 168
## 11        11       5.7  NA
## 12        12       8.0  80
## 13        13      10.3 114
## 14        14      12.0 147
## 15        15       9.2 122
## 16        16       8.5  NA
## 17        17       7.0  82
## 18        18      10.7  88
## 19        19       9.3  NA
## 20        20       8.2  NA
```