Standard and Non Standard Missing Values

In this document, we will go over standard missing values that R recognizes, and how to handle non standard missing values that R may not recognize.

We will use the dataset missing.csv as a working example. Download the dataset from Collab and read it into R, as well as load the `tidyverse` package

```
library(tidyverse)
Data<-read.csv("missing.csv", header=TRUE)
```

Let us take a look at this dataframe

```
Data
```

```
##   Height Weight
## 1     62    135
## 2     66    190
## 3     70    230
## 4     65    130
## 5     NA    260
## 6    NaN    250
## 7     70
## 8     72     na
## 9     63    N/A
```

Visually, it appears that

- observations 5 and 6 have missing values for Height,
- observations 7, 8, and 9 have missing values for Weight.

However, when we apply the `is.na()` function to the dataframe, we get the following

```
##see which entries are viewed as missing
is.na(Data)
```

```
##       Height Weight
##  [1,]  FALSE  FALSE
##  [2,]  FALSE  FALSE
##  [3,]  FALSE  FALSE
##  [4,]  FALSE  FALSE
##  [5,]   TRUE  FALSE
##  [6,]   TRUE  FALSE
##  [7,]  FALSE  FALSE
##  [8,]  FALSE  FALSE
##  [9,]  FALSE  FALSE
```

R has only identified the entries with `NA` and `NaN` as missing values. Indeed, these are the standard missing values for R which R recognizes. Other ways of expressing missing values do not get recognized.

Strictly speaking, `NaN` is meant to represent an undefined number, while `NA` is meant to represent a missing value. However, `is.na()` recognizes both of these as missing.

We can convert the non standard missing values to 'standard missing values'NA' using the `replace()` function within `mutate()`

```
Data<-Data%>%
  mutate(Weight = replace(Weight, Weight == "na", NA))%>%
  mutate(Weight = replace(Weight, Weight == "N/A", NA))%>%
  mutate(Weight = replace(Weight, Weight == "", NA))

is.na(Data)
```

```
##       Height Weight
##  [1,]  FALSE  FALSE
##  [2,]  FALSE  FALSE
##  [3,]  FALSE  FALSE
##  [4,]  FALSE  FALSE
##  [5,]   TRUE  FALSE
##  [6,]   TRUE  FALSE
##  [7,]  FALSE   TRUE
##  [8,]  FALSE   TRUE
##  [9,]  FALSE   TRUE
```