22BCE0476

Aman Chauhan

```
library(dplyr)
library(tidyr)
library(ggplot2)

#1. Load the fruits.csv file into the R environment.
fruit<- read.csv("C:/Users/batch1/Downloads/fruits.csv")
print(fruits)

#2. For the existing data frame insert the first record as 'Your name, Rollno (last 4 digits), roll no, roll no, roll no, NaN)
details<-c('Aman Chauhan',0476,0476,0476,0476,NaN)
rbind(fruit,details)

#3.  Check if there exists any NA. (is, any, are)
#is
any_na<-any(is.na(fruits))
print(any_na)

#4. Finding the missing summaries of the dataset (miss_var_summary, miss_var_table). load the package "naniar" if required
miss_summary <-colSums(is.na(fruits))
print(miss_summary)

miss_table<-data.frame(Variable=names(fruits),MissingCount=colSums(is.na(fruits)))
print(miss_table)

#5.5.      Find the total number of NA in the existing dataset (n miss, colSums)
total_na<-sum(is.na(fruits))
print(total_na)
```

```
col_na_sums<-colSums(is.na(fruits))
print(col_na_sums)
```

#6. Count the total number of complete values in Store4 and Store5 (n complete)

```
n_complete_stored<-sum(!is.na(fruits$V5))
n_complete_store5<-sum(!is.na(fruit$V6))
print(n_complete_stored)
print(n_complete_store5)
```

#7. Count the total number of missing values in each of the variables w.r.t to proportion (prop miss, prop complete)

```
prop_miss<-colMeans(is.na(fruits))
print(prop_miss)
prop_complete<- 1-prop_miss
print(prop_complete)
```

#8. Displaying the missing values per column, for each observation. (missing plot).

```
missing_plot<-function(df){
  missing_data<-is.na(df)
  missing_count<-rowSums(missing_data)
  plot_data<-data.frame(Row=1:nrow(df),MissigCount=missing_count)
  ggplot(plot_data,aes(x=Row,y=MissingCount))+geom_bar(stat="identify")+labs(title="missing values per row 22bce0476",x="row number 22bce0476",y="col number 22bce0476")

}
missing_plot(fruits)
```

#9. Performing row-wise deletion. (na omit)

```
fruits_cleaned<-na.omit(fruits)
print(fruits_cleaned)
```

Output

```
library(dplyr)
> library(tidyr)
> library(ggplot2)
> ?naniar
No documentation for ‘naniar’ in specified packages and libraries:
you could try ‘??naniar’
> #1. Load the fruits.csv file into the R environment.
> fruit<- read.csv("C:/Users/batch1/Downloads/fruits.csv")
> print(fruit)
      fruits store1 store2 store3 store4 store5
1      apple     15     16     17     20    NaN
2     banana     18     19     20    NaN    NaN
3       kiwi     21     22     23    NaN    NaN
4     grapes     24     25     26    NaN    NaN
5      mango     27     28     29    NaN    NaN
6 watermelon     15     16     17     18    NaN
7    oranges    NaN    NaN    NaN    NaN    NaN
8  pineapple     12     13     NA    NaN    NaN
>
> #2. For the existing data frame insert the first record as ‘Your name, Rollno (last
> details<-c('Aman Chauhan',0476,0476,0476,0476,NaN)
> rbind(fruit,details)
         fruits store1 store2 store3 store4 store5
1        apple     15     16     17     20    NaN
2       banana     18     19     20    NaN    NaN
3         kiwi     21     22     23    NaN    NaN
4       grapes     24     25     26    NaN    NaN
5        mango     27     28     29    NaN    NaN
6   watermelon     15     16     17     18    NaN
7      oranges    NaN    NaN    NaN    NaN    NaN
8    pineapple     12     13   <NA>    NaN    NaN
9 Aman Chauhan    476    476    476    476    NaN
>
> #3.   Check if there exists any NA. (is, any, are)
> #is
> any_na<-any(is.na(fruits))
> print(any_na)
[1] TRUE
>
> #4. Finding the missing summaries of the dataset (miss_var_summary, miss_var_table).
> miss_summary <-colSums(is.na(fruits))
> print(miss_summary)
V1 V2 V3 V4 V5 V6
 0  0  0  1  0  0
>
> miss_table<-data.frame(Variable=names(fruits),MissingCount=colSums(is.na(fruits)))
> print(miss_table)
   Variable MissingCount
V1       V1            0
V2       V2            0
V3       V3            0
V4       V4            1
V5       V5            0
V6       V6            0
>
> #5.5.        Find the total number of NA in the existing dataset (n miss, colSums)
> total_na<-sum(is.na(fruits))
> print(total_na)
[1] 1
>
> col_na_sums<-colSums(is.na(fruits))
> print(col_na_sums)
V1 V2 V3 V4 V5 V6
 0  0  0  1  0  0
```

```
>
> #6.        Count the total number of complete values in Store4 and Store5 (n complet
>
> n_complete_stored<-sum(!is.na(fruits$V5))
> n_complete_store5<-sum(!is.na(fruit$V6))
> print(n_complete_stored)
[1] 9
> print(n_complete_store5)
[1] 0
>
> #7.     Count the total number of missing values in each of the variables w.r.t to p
> prop_miss<-colMeans(is.na(fruits))
> print(prop_miss)
       V1        V2        V3        V4        V5        V6
0.0000000 0.0000000 0.0000000 0.1111111 0.0000000 0.0000000
> prop_complete<- 1-prop_miss
> print(prop_complete)
       V1        V2        V3        V4        V5        V6
1.0000000 1.0000000 1.0000000 0.8888889 1.0000000 1.0000000
>
> #8.   Displaying the missing values per column, for each observation. (missing plot)
> missing_plot<-function(df){
+   missing_data<-is.na(df)
+   missing_count<-rowSums(missing_data)
+   plot_data<-data.frame(Row=1:nrow(df),MissigCount=missing_count)
+   ggplot(plot_data,aes(x=Row,y=MissingCount))+geom_bar(stat="identify")+labs(title="
0476",y="col number 22bce0476")
+
+ }
> missing_plot(fruits)
Error in `geom_bar()`:
! Can't find stat called "identify"
Run `rlang::last_trace()` to see where the error occurred.
> library(dplyr)
> library(tidyr)
> library(ggplot2)
> ?naniar
No documentation for ‘naniar’ in specified packages and libraries:
you could try ‘??naniar’
> #1. Load the fruits.csv file into the R environment.
> fruit<- read.csv("C:/Users/batch1/Downloads/fruits.csv")
>
>
> #2. For the existing data frame insert the first record as ‘Your name, Rollno (last
> details<-c('Aman Chauhan',0476,0476,0476,0476,NaN)
> rbind(fruit,details)
          fruits store1 store2 store3 store4 store5
1          apple     15     16     17     20    NaN
2         banana     18     19     20    NaN    NaN
3           kiwi     21     22     23    NaN    NaN
4         grapes     24     25     26    NaN    NaN
5          mango     27     28     29    NaN    NaN
6   watermelon       15     16     17     18    NaN
7       oranges     NaN    NaN    NaN    NaN    NaN
8     pineapple      12     13   <NA>     NaN    NaN
9 Aman Chauhan     476    476    476    476    NaN
>
> #3.   Check if there exists any NA. (is, any, are)
> #is
> any_na<-any(is.na(fruits))
> print(any_na)
[1] TRUE
>
> #4. Finding the missing summaries of the dataset (miss_var_summary, miss_var_table).
> miss_summary <-colSums(is.na(fruits))
> print(miss_summary)
V1 V2 V3 V4 V5 V6
 0  0  0  1  0  0
>
```

```
> miss_table<-data.frame(Variable=names(fruits),MissingCount=colSums(is.na(fruits)))
> print(miss_table)
   Variable MissingCount
V1       V1            0
V2       V2            0
V3       V3            0
V4       V4            1
V5       V5            0
V6       V6            0
>
> #5.5.        Find the total number of NA in the existing dataset (n miss, colSums)
> total_na<-sum(is.na(fruits))
> print(total_na)
[1] 1
>
> col_na_sums<-colSums(is.na(fruits))
> print(col_na_sums)
V1 V2 V3 V4 V5 V6
 0  0  0  1  0  0
>
> #6.        Count the total number of complete values in Store4 and Store5 (n complet
>
> n_complete_stored<-sum(!is.na(fruits$V5))
> n_complete_store5<-sum(!is.na(fruit$V6))
> print(n_complete_stored)
[1] 9
> print(n_complete_store5)
[1] 0
>
> #7.     Count the total number of missing values in each of the variables w.r.t to p
> prop_miss<-colMeans(is.na(fruits))
> print(prop_miss)
       V1        V2        V3        V4        V5        V6
0.0000000 0.0000000 0.0000000 0.1111111 0.0000000 0.0000000
> prop_complete<- 1-prop_miss
> print(prop_complete)
       V1        V2        V3        V4        V5        V6
1.0000000 1.0000000 1.0000000 0.8888889 1.0000000 1.0000000
>
> #8.   Displaying the missing values per column, for each observation. (missing plot)
> missing_plot<-function(df){
+   missing_data<-is.na(df)
+   missing_count<-rowSums(missing_data)
+   plot_data<-data.frame(Row=1:nrow(df),MissigCount=missing_count)
+   ggplot(plot_data,aes(x=Row,y=MissingCount))+geom_bar(stat="identify")+labs(title="
0476",y="col number 22bce0476")
+
+ }
> missing_plot(fruits)
Error in `geom_bar()`:
! Can't find stat called "identify"
Run `rlang::last_trace()` to see where the error occurred.
> library(dplyr)
> library(tidyr)
> library(ggplot2)
>
> #1. Load the fruits.csv file into the R environment.
> fruit<- read.csv("C:/Users/batch1/Downloads/fruits.csv")
> print(fruits)
          V1     V2     V3     V4     V5     V6
1     fruits store1 store2 store3 store4 store5
2      apple     15     16     17     20    NaN
3     banana     18     19     20    NaN    NaN
4       kiwi     21     22     23    NaN    NaN
5     grapes     24     25     26    NaN    NaN
6      mango     27     28     29    NaN    NaN
7 watermelon     15     16     17     18    NaN
8    oranges    NaN    NaN    NaN    NaN    NaN
9  pineapple     12     13   <NA>    NaN    NaN
```

```
>
> #2. For the existing data frame insert the first record as 'Your name, Rollno (last
> details<-c('Aman Chauhan',0476,0476,0476,0476,NaN)
> rbind(fruit,details)
       fruits store1 store2 store3 store4 store5
1       apple     15     16     17     20    NaN
2      banana     18     19     20    NaN    NaN
3        kiwi     21     22     23    NaN    NaN
4      grapes     24     25     26    NaN    NaN
5       mango     27     28     29    NaN    NaN
6  watermelon     15     16     17     18    NaN
7     oranges    NaN    NaN    NaN    NaN    NaN
8   pineapple     12     13   <NA>    NaN    NaN
9 Aman Chauhan    476    476    476    476    NaN
>
> #3.   Check if there exists any NA. (is, any, are)
> #is
> any_na<-any(is.na(fruits))
> print(any_na)
[1] TRUE
>
> #4. Finding the missing summaries of the dataset (miss_var_summary, miss_var_table).
> miss_summary <-colSums(is.na(fruits))
> print(miss_summary)
V1 V2 V3 V4 V5 V6
 0  0  0  1  0  0
>
> miss_table<-data.frame(Variable=names(fruits),MissingCount=colSums(is.na(fruits)))
> print(miss_table)
   Variable MissingCount
V1       V1            0
V2       V2            0
V3       V3            0
V4       V4            1
V5       V5            0
V6       V6            0
>
> #5.5.        Find the total number of NA in the existing dataset (n miss, colSums)
> total_na<-sum(is.na(fruits))
> print(total_na)
[1] 1
>
> col_na_sums<-colSums(is.na(fruits))
> print(col_na_sums)
V1 V2 V3 V4 V5 V6
 0  0  0  1  0  0
>
> #6.        Count the total number of complete values in Store4 and Store5 (n complet
>
> n_complete_stored<-sum(!is.na(fruits$V5))
> n_complete_store5<-sum(!is.na(fruit$V6))
> print(n_complete_stored)
[1] 9
> print(n_complete_store5)
[1] 0
>
> #7.     Count the total number of missing values in each of the variables w.r.t to p
> prop_miss<-colMeans(is.na(fruits))
> print(prop_miss)
       V1        V2        V3        V4        V5        V6
0.0000000 0.0000000 0.0000000 0.1111111 0.0000000 0.0000000
> prop_complete<- 1-prop_miss
> print(prop_complete)
       V1        V2        V3        V4        V5        V6
1.0000000 1.0000000 1.0000000 0.8888889 1.0000000 1.0000000
>
> #8.   Displaying the missing values per column, for each observation. (missing plot)
> missing_plot<-function(df){
+   missing_data<-is.na(df)
```

```
+   missing_count<-rowSums(missing_data)
+   plot_data<-data.frame(Row=1:nrow(df),MissigCount=missing_count)
+   ggplot(plot_data,aes(x=Row,y=MissingCount))+geom_bar(stat="identify")+labs(title="
0476",y="col number 22bce0476")
+
+ }
> missing_plot(fruits)
Error in `geom_bar()`:
! Can't find stat called "identify"
Run `rlang::last_trace()` to see where the error occurred.


>
```