#

#

# Create a 5 by 3 sample data frame

#

aa<-sample(1:5)

bb<-sample(1:5)

cc<-sample(1:5)

df<-data.frame(aa,bb,cc)

df

str(df)

#

names(df)<-c("col1","col2","col3") # assign col names

df

#

# randomly assign NA values

# NA is generally interpreted as a missing value

#

df[2,2]<-NA

df[4,3]<-NA

df

#

mean(df$col2) # take the mean of col2,, should get an error

mean(df$col2,na.rm=TRUE) # take the mean of col2 and 'manage' the NA's

#

# let's omit NA's

#

df # look at df

df1<-na.omit(df) # omit the NA's, store in a new df, review the impact

df1 # what do you see, so what ??

#

#

df2<-df # lets's create anothe df utilizing our original df with NA's

df2

length(df2$col2) # check the lenght of col2

count <-length(df$col2[df$col2=='NA']) # how many NA's in col2

count

count2<-length(df2[df2=='NA']) # test to see how many NA's in entire df

count2

##### more NA examples

# create a sample data frame with random NA's

#

m <- matrix(sample(c(NA, 1:10), 100, replace = TRUE), 10)

d <- as.data.frame(m)

d

any(is.na(d)) # do any NA's exist, returns T or F

length(d[d=='NA']) # how many NA's

#

#

if (any(is.na(d))) # if function to count NA's if they exist

{

count<-length(d[d=='NA']) # how many NA's

print(c("NA Count ", count))

} else

{

print("no NA's")

}

d[is.na(d)] <- 0 # replace NA's with zero, entire df

d

#

m <- matrix(sample(c(NA, 1:10), 100, replace = TRUE), 10) # re-create a matrix with random NA's

d <- as.data.frame(m)

d

d$V1[is.na(d$V1)] <- mean(d$V1,na.rm=TRUE) # replce NA's in col with mean of the col

d

##

##############

##

any(is.null(d))

sum(is.null(d))

#

if (any(is.null(d)))

{

count<-sum(is.null(d)) # how many NULL's

print(c("NULL Count ", count))

} else

{

print("no NULL's")

}

#If you are looking for NA counts for each column in a dataframe then:

# na\_count <-sapply(x, function(y) sum(length(which(is.na(y)))))

#should give you a list with the counts for each column.

#na\_count <- data.frame(na\_count)

#Should output the data nicely in a dataframe like:

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

| row.names | na\_count

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

| column\_1 | count

###### exploring NULL's

#

# Vectors (which is essentially a column in a data.frame) cannot

# contain a mixture of observed values and NULL values.

# In R, we prefer NA to NULL to represent missing data

#

# NA vs NULL https://www.r-bloggers.com/r-na-vs-null/

#

x<-NULL

y<-NULL

is.null(x)

z<-c(x,y)

is.null(z)

z<-c(1,x,2,y)

any(is.null(z))

z

length(z)

length(x)

sum(is.null(z))

l<-list(1,x,2,y)

length(l)

any(is.null(l))

is.null(l)

l

is.null(l[[2]])

l[2]

numRows<-length(d)

for(i in 1:numRows) {

record <- d[[i]]

record[sapply(record, is.null)] <- NA

d[[i]] <- record

}

d

ct<-0

if(is.null(d[[2]])){

ct<-ct+1

}

ct