**Question 1**

**Code:**

setwd(choose.dir())

getwd()

Int\_Data <- read.csv("1555058318\_internet\_datasetz.csv")

summary(Int\_Data)

**Result:**

**From the result of summarized dataset, it is observed that the numerical data**

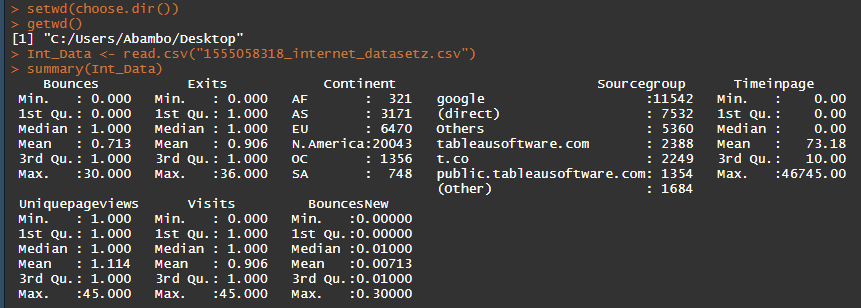
**includes information related to the maximum, minimum, and mean data. The categorical**

**data like continent includes the data of the number of times the category has**

**been repeated in the dataset. We can see that there is a maximum value of 30**

**bounces for the website. This site was accessed maximum number of times by visitors**

**from North America.**



**Question 2**

**Code:**

**Web\_Anov <- aov(Visits ~ Uniquepageviews, data = Int\_Data)**

**summary(Web\_Anov)**

**Alpha <- 0.05**

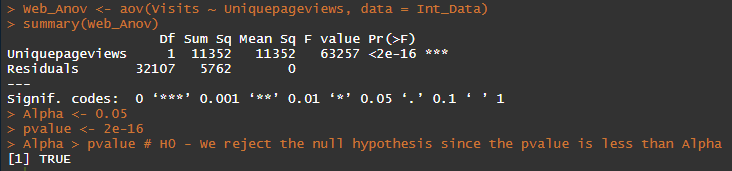
**pvalue <- 2e-16**

**Alpha > pvalue**

**Result:**

**H0 - We reject the null hypothesis since the pvalue is less than Alpha**

**H0 – The unique page view value does not depends on visits**



**Question 3**

**Code:**

**Exits\_Anov <- aov(Exits ~., data = Int\_Data)**

**summary(Exits\_Anov)**

**Result:**

**Exits has the following probable factors:**

**Pr(>F)**

**Bounces 2e-16 \*\*\***

**Continent 1.62e-05 \*\*\***

**Sourcegroup 4.89e-12 \*\*\***

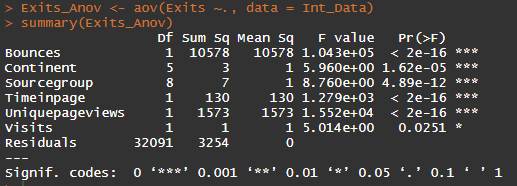
**Timeinpage 2e-16 \*\*\***

**Uniquepageviews 2e-16 \*\*\***

**Exit is dependent to the above factors since the alpha is greater than pvalue of all the above factors**

**H0 : Therefore we reject the null hypothesis from all the above factors**

**H1: We accept the Alternative hypothesis of all the above factors**



**Question 4**

**Code:**

Time\_Anov <- aov(Timeinpage ~., data = Int\_Data)

summary(Time\_Anov)

**Result:**

The following variables that have effect on Timeonpage:

Pr(>F)

Bounces 2e-16 \*\*\*

Exits 2e-16 \*\*\*

Continent 2.51e-06 \*\*\*

Sourcegroup 0.202

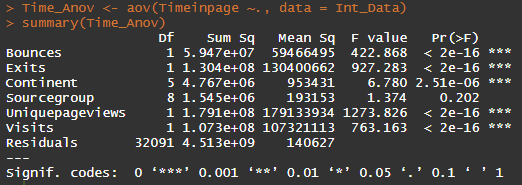
Uniquepageviews 2e-16 \*\*\*

Visits 2e-16 \*\*\*

Timeonpage is dependent in the above variables since the alpha is greater than their pvalue, accept the Sourcegroup variable, it pvalue is greater than alpha.

H0: we reject the Null hypothesis, Because the pvalue is less than alpha

H1: we accept the Alternative hypothesis, since the Sourcegroup is greater than the alpha.



**Question 5**

**Code:**

str(Int\_Data)

for(i in c("Bounces", "Exits","Timeinpage","Uniquepageviews","Visits"))

{

Int\_Data[,i] = as.factor(Int\_Data[,i])

}

Bounce\_Log <- glm(Bounces\*0.01 ~.,

family = binomial(link = 'logit'),

data = Int\_Data)

summary(Bounce\_Log)

**Result:**

**I used the Logistic Regression to determine the factors that impact the bouncing,**

**See the screenshot below for the output:**

