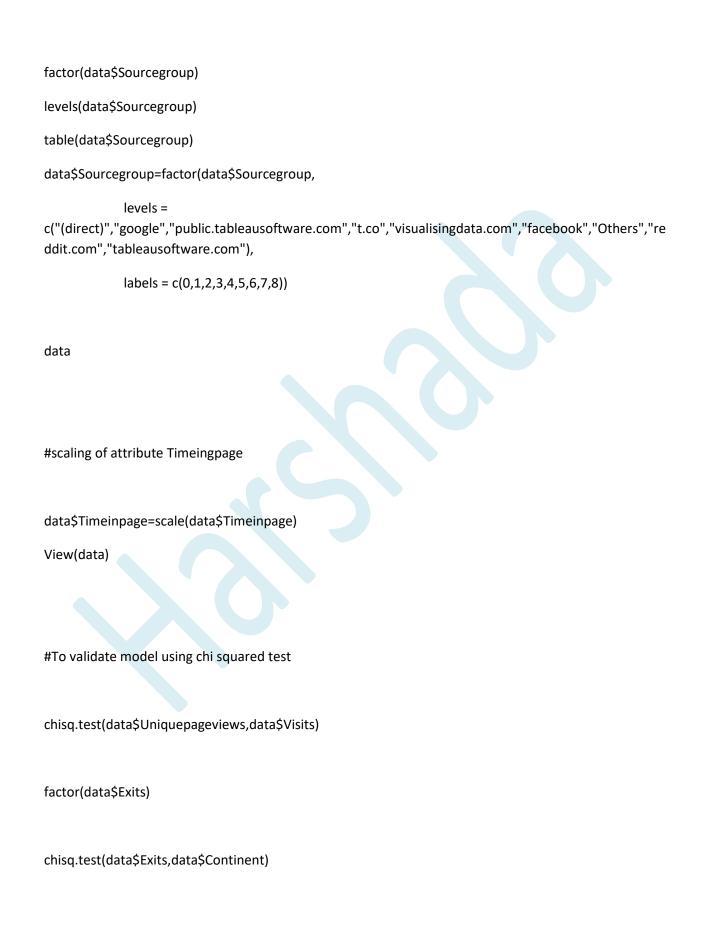# Web Analysis

#Que 1

#The team wants to analyze each variable of the data collected

#through data summarization to get a basic understanding of the

#dataset and to prepare for further analysis.

```
library(readxl)

data<-read_xlsx("C:/Users/Harshada/data/1555058318_internet_dataset.xlsx")

View(data)

head(data)

str(data)

summary(data)

is.na(data)

#data<-na.omit(data)
```

#Factorize the attribute Continentss

```
data$Continent=factor(data$Continent,

        levels = c("AF","AS","EU","N.America","OC","SA"),

        labels = c(0,1,2,3,4,5))

data
```
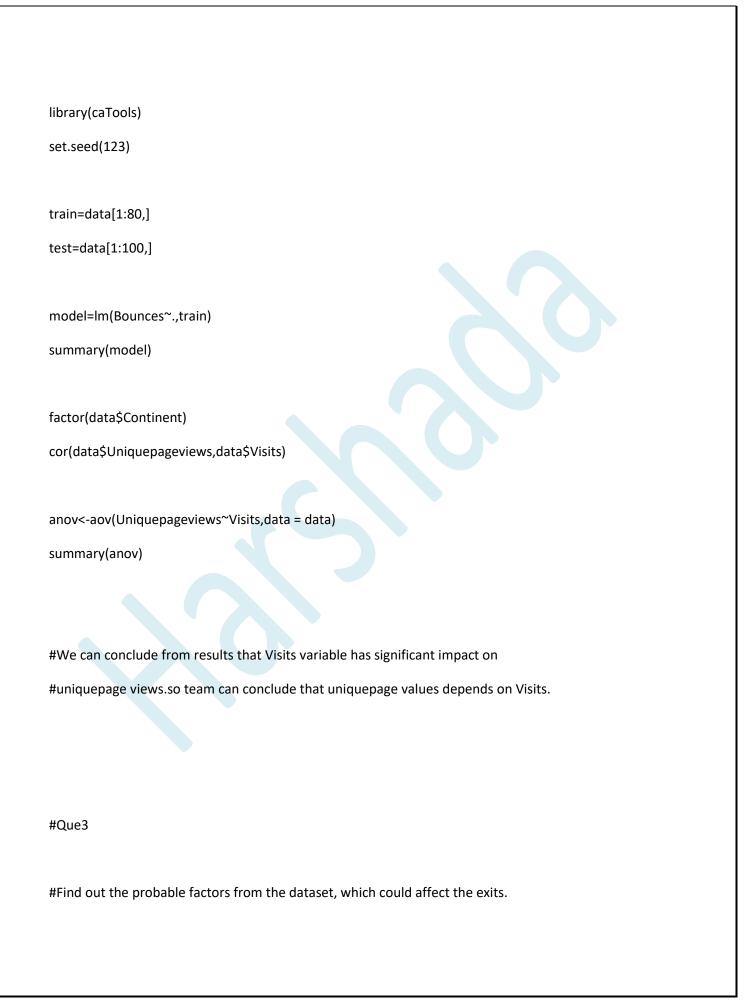
#Factorize the attribute Sourcegroup

```r
factor(data$Sourcegroup)

levels(data$Sourcegroup)

table(data$Sourcegroup)

data$Sourcegroup=factor(data$Sourcegroup,

            levels =
c("(direct)","google","public.tableausoftware.com","t.co","visualisingdata.com","facebook","Others","re
ddit.com","tableausoftware.com"),

            labels = c(0,1,2,3,4,5,6,7,8))


data


#scaling of attribute Timeingpage


data$Timeinpage=scale(data$Timeinpage)

View(data)


#To validate model using chi squared test


chisq.test(data$Uniquepageviews,data$Visits)


factor(data$Exits)


chisq.test(data$Exits,data$Continent)
```

```r
chisq.test(data$Exits,data$Sourcegroup)

chisq.test(data$Exits,data$Timeinpage)

chisq.test(data$Exits,data$Uniquepageviews)

chisq.test(data$Exits,data$Visits)

chisq.test(data$Exits,data$Bounces)


#bounces min=0,max=30

#exit min=0 max=36

#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 A


#ques2

#As mentioned earlier,a unique page view represents the number of sessions during which that

#page was viewed one or more times. A visit counts all instances, no matter

#how many times the same visitor may have been to your site. So the team needs

#to know whether the unique page view value depends on visits.


library(ggplot2)

ggplot(data,aes(x=Bounces,y=Visits))+geom_point(color="red",shape=3)
```

```
library(caTools)

set.seed(123)


train=data[1:80,]

test=data[1:100,]


model=lm(Bounces~.,train)

summary(model)


factor(data$Continent)

cor(data$Uniquepageviews,data$Visits)


anov<-aov(Uniquepageviews~Visits,data = data)

summary(anov)



#We can conclude from results that Visits variable has significant impact on

#uniquepage views.so team can conclude that uniquepage values depends on Visits.



#Que3


#Find out the probable factors from the dataset, which could affect the exits.
```
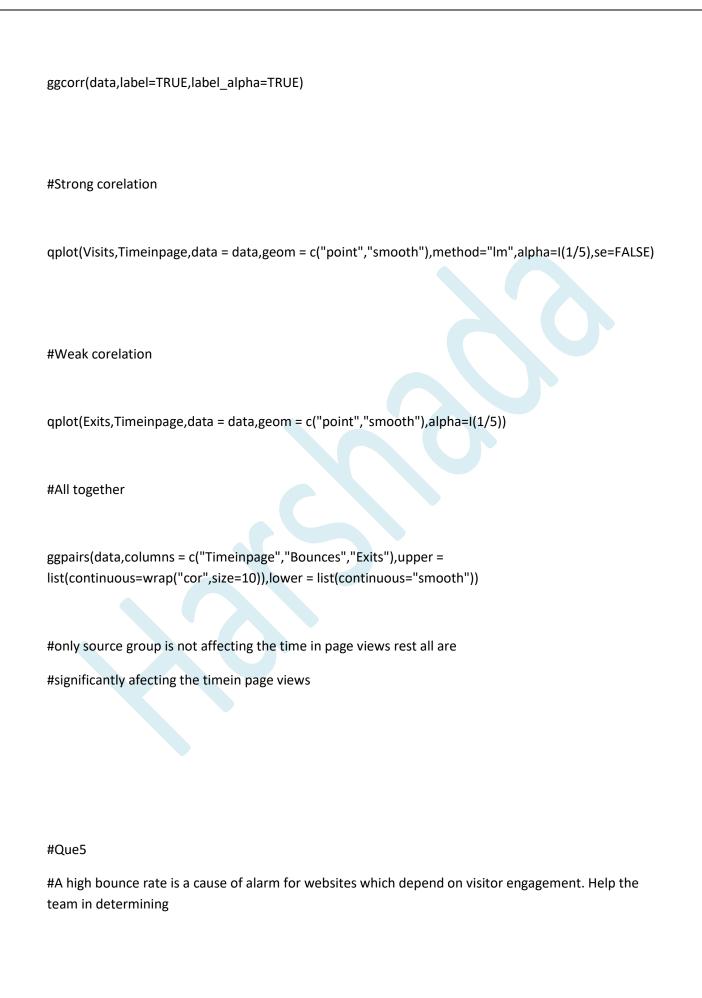
```
#Exit Page Analysis is usually required to get an idea about why a user leaves

#the website for a session and moves on to another one.

#Please keep in mind that exits should not be confused with bounces.


anoo<-aov(Exits~.,data = data)

summary(anoo)


newModel=lm(Exits~Bounces+Sourcegroup+Timeinpage+Visits,train)

summary(newModel)


predExits=predict(newModel,test)

View(predExits)

predExits=round(predExits)

View(predExits)

factor(predExits)


final_data=cbind(test,predExits)

View(final_data)


#Plot of affects of Bounces and timeingpage on Exits


ggplot(data,aes(x=Bounces+Timeinpage,y=Exits))+geom_point(color="red",shape=3)


#From the result of ANOVA given here, we can see that source.group, bounces,

#and unique.pageviews have more significance. Visits have comparatively less significance.
```

#Hence we can say that exit from the site is affected by the factors of source group,

#bounces, and unique.pageviews.

#Que 4

#Every site wants to increase the time on page for a visitor. This increases

#the chances of the visitor understanding the site content better and hence

there are more chances of a transaction taking place. Find the variables

#which possibly have an effect on the time on page.


```
anooo<-aov(Timeinpage~.,data = data)

summary(anooo)

library(psych)

describe(data)


summary(data)
```


#find corelation between variables

#one way annova of bounces

```
aov1<-aov(Timeinpage~Bounces,data = data)

summary(aov1)

print(model.tables(aov1,"means"),digits = 3)
```

```r
#one way annova of Continent

aov2<-aov(Timeinpage~Continent,data = data)

summary(aov2)

print(model.tables(aov2,"means"),digits = 3)



#Two way annova


aov3<-aov(Timeinpage~Bounces*Continent,data = data)

summary(aov3)

print(model.tables(aov3,"means"),digits = 3)



#Corelation

library(ggplot2)

library(GGally)


cor(data$Timeinpage,data$Bounces)

cor(data[,5:6])



#Corelation coefficient is between -1 to 1 hence graph shows strong and weak

#corelation between variables.
```

ggcorr(data,label=TRUE,label_alpha=TRUE)

#Strong corelation

qplot(Visits,Timeinpage,data = data,geom = c("point","smooth"),method="lm",alpha=I(1/5),se=FALSE)

#Weak corelation

qplot(Exits,Timeinpage,data = data,geom = c("point","smooth"),alpha=I(1/5))

#All together

ggpairs(data,columns = c("Timeinpage","Bounces","Exits"),upper =
list(continuous=wrap("cor",size=10)),lower = list(continuous="smooth"))

#only source group is not affecting the time in page views rest all are

#significantly afecting the timein page views

#Que5

#A high bounce rate is a cause of alarm for websites which depend on visitor engagement. Help the team in determining

#the factors that are impacting the bounce.

data$Bounces=data$Bounces*0.01

rmm<-glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits,data = data,family = "binomial")

summary(rmm)

#As can be inferred from the result shown, the BouncesNew, Unique.Pageviews and visits are the variables that

#impact the target variable bounces

#it has greater significance.

....OUTPUT...

library(readxl)

> data<-read_xlsx("C:/Users/Harshada/data/1555058318_internet_dataset.xlsx")

> head(data)

A tibble: 6 x 8

Bounces	Exits	Continent	Sourcegroup	Timeinpage	Uniquepageviews	Visits	BouncesNew
<dbl>	<dbl>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 0	0	OC	(direct)	18	1	0	0
2 0	0	N.America	(direct)	4	1	0	0
3 0	0	N.America	Others	35	1	0	0
4 0	0	N.America	public.table~	70	1	0	0
5 0	0	N.America	public.table~	81	1	0	0
6 0	0	N.America	public.table~	75	1	0	0

```
> str(data)

tibble [32,109 x 8] (S3: tbl_df/tbl/data.frame)

 $ Bounces       : num [1:32109] 0 0 0 0 0 0 0 0 0 0 ...

 $ Exits         : num [1:32109] 0 0 0 0 0 0 0 0 0 0 ...

 $ Continent     : chr [1:32109] "OC" "N.America" "N.America" "N.America" ...

 $ Sourcegroup   : chr [1:32109] "(direct)" "(direct)" "Others" "public.tableausoftware.com" ...

 $ Timeinpage    : num [1:32109] 18 4 35 70 81 75 186 710 712 344 ...

 $ Uniquepageviews: num [1:32109] 1 1 1 1 1 1 1 1 1 1 ...

 $ Visits        : num [1:32109] 0 0 0 0 0 0 0 0 1 1 ...

 $ BouncesNew    : num [1:32109] 0 0 0 0 0 0 0 0 0 0 ...

> summary(data)

   Bounces         Exits        Continent         Sourcegroup

 Min.   : 0.000  Min.   : 0.000  Length:32109       Length:32109

 1st Qu.: 0.000  1st Qu.: 1.000  Class :character  Class :character

 Median : 1.000  Median : 1.000  Mode  :character  Mode  :character

 Mean   : 0.713  Mean   : 0.906

 3rd Qu.: 1.000  3rd Qu.: 1.000

 Max.   :30.000  Max.   :36.000

   Timeinpage     Uniquepageviews     Visits        BouncesNew

 Min.   :  0.00  Min.   : 1.000  Min.   : 0.000  Min.   :0.00000

 1st Qu.:  0.00  1st Qu.: 1.000  1st Qu.: 1.000  1st Qu.:0.00000

 Median :  0.00  Median : 1.000  Median : 1.000  Median :0.01000

 Mean   : 73.18  Mean   : 1.114  Mean   : 0.906  Mean   :0.00713

 3rd Qu.: 10.00  3rd Qu.: 1.000  3rd Qu.: 1.000  3rd Qu.:0.01000
```

```
 Max.  :46745.00  Max.  :45.000  Max.  :45.000  Max.  :0.30000
> is.na(data)
```

```
      Bounces Exits Continent Sourcegroup Timeinpage Uniquepageviews Visits

 [1,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [2,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [3,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [4,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [5,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [6,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [7,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [8,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

 [9,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[10,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[11,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[12,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[13,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[14,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[15,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[16,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[17,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[18,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[19,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[20,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE

[21,]  FALSE FALSE    FALSE      FALSE      FALSE        FALSE  FALSE
```

```
[22,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[23,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[24,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[25,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[26,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[27,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[28,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[29,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[30,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[31,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[32,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[33,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[34,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[35,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[36,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[37,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[38,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[39,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[40,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[41,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[42,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[43,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[44,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[45,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
```

```
[46,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[47,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[48,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[49,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[50,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[51,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[52,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[53,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[54,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[55,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[56,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[57,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[58,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[59,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[60,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[61,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[62,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[63,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[64,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[65,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[66,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[67,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[68,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
[69,] FALSE FALSE   FALSE   FALSE   FALSE   FALSE FALSE
```

```
[70,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[71,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[72,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[73,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[74,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[75,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[76,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[77,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[78,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[79,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[80,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[81,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[82,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[83,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[84,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[85,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[86,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[87,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[88,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[89,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[90,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[91,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[92,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE

[93,] FALSE FALSE    FALSE    FALSE    FALSE    FALSE FALSE
```

[94,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[95,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[96,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[97,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[98,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[99,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[100,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[101,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[102,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[103,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[104,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[105,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[106,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[107,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[108,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[109,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[110,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[111,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[112,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[113,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[114,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[115,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[116,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE
[117,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE FALSE

[118,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[119,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[120,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[121,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[122,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[123,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[124,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[125,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

BouncesNew

[1,]	FALSE
[2,]	FALSE
[3,]	FALSE
[4,]	FALSE
[5,]	FALSE
[6,]	FALSE
[7,]	FALSE
[8,]	FALSE
[9,]	FALSE
[10,]	FALSE
[11,]	FALSE
[12,]	FALSE
[13,]	FALSE
[14,]	FALSE
[15,]	FALSE

[16,]   FALSE

[17,]   FALSE

[18,]   FALSE

[19,]   FALSE

[20,]   FALSE

[21,]   FALSE

[22,]   FALSE

[23,]   FALSE

[24,]   FALSE

[25,]   FALSE

[26,]   FALSE

[27,]   FALSE

[28,]   FALSE

[29,]   FALSE

[30,]   FALSE

[31,]   FALSE

[32,]   FALSE

[33,]   FALSE

[34,]   FALSE

[35,]   FALSE

[36,]   FALSE

[37,]   FALSE

[38,]   FALSE

[39,]   FALSE

[40,]   FALSE

[41,]   FALSE

[42,]   FALSE

[43,]   FALSE

[44,]   FALSE

[45,]   FALSE

[46,]   FALSE

[47,]   FALSE

[48,]   FALSE

[49,]   FALSE

[50,]   FALSE

[51,]   FALSE

[52,]   FALSE

[53,]   FALSE

[54,]   FALSE

[55,]   FALSE

[56,]   FALSE

[57,]   FALSE

[58,]   FALSE

[59,]   FALSE

[60,]   FALSE

[61,]   FALSE

[62,]   FALSE

[63,]   FALSE

[64,]   FALSE

[65,]   FALSE

[66,]   FALSE

[67,]   FALSE

[68,]   FALSE

[69,]   FALSE

[70,]   FALSE

[71,]   FALSE

[72,]   FALSE

[73,]   FALSE

[74,]   FALSE

[75,]   FALSE

[76,]   FALSE

[77,]   FALSE

[78,]   FALSE

[79,]   FALSE

[80,]   FALSE

[81,]   FALSE

[82,]   FALSE

[83,]   FALSE

[84,]   FALSE

[85,]   FALSE

[86,]   FALSE

[87,]   FALSE

[88,]   FALSE

[89,]   FALSE

[90,]   FALSE

[91,]   FALSE

[92,]   FALSE

[93,]   FALSE

[94,]   FALSE

[95,]   FALSE

[96,]   FALSE

[97,]   FALSE

[98,]   FALSE

[99,]   FALSE

[100,]   FALSE

[101,]   FALSE

[102,]   FALSE

[103,]   FALSE

[104,]   FALSE

[105,]   FALSE

[106,]   FALSE

[107,]   FALSE

[108,]   FALSE

[109,]   FALSE

[110,]   FALSE

[111,]   FALSE

```
[112,]    FALSE

[113,]    FALSE

[114,]    FALSE

[115,]    FALSE

[116,]    FALSE

[117,]    FALSE

[118,]    FALSE

[119,]    FALSE

[120,]    FALSE

[121,]    FALSE

[122,]    FALSE

[123,]    FALSE

[124,]    FALSE

[125,]    FALSE
```

 [reached getOption("max.print") -- omitted 31984 rows]

[reached getOption("max.print") -- omitted 31984 rows]

> data$Continent=factor(data$Continent,

+   levels = c("AF","AS","EU","N.America","OC","SA"),

+   labels = c(0,1,2,3,4,5))

> data

A tibble: 32,109 x 8

  Bounces Exits Continent Sourcegroup Timeinpage Uniquepageviews Visits BouncesNew

   <dbl> <dbl> <fct>   <chr>   <dbl>   <dbl> <dbl>   <dbl>

 1   0   0 4   (direct)   18   1   0   0

```
2     0    0 3      (direct)        4          1    0       0

3     0    0 3      Others          35         1    0       0

4     0    0 3      public.tabl~    70         1    0       0

5     0    0 3      public.tabl~    81         1    0       0

6     0    0 3      public.tabl~    75         1    0       0

7     0    0 3      public.tabl~    186        1    0       0

8     0    0 3      (direct)        710        1    0       0

9     0    0 4      (direct)        712        1    1       0

10    0    0 1      Others          344        1    1       0
# ... with 32,099 more rows
> factor(data$Sourcegroup)
 [1] (direct)              (direct)

 [3] Others                public.tableausoftware.com

 [5] public.tableausoftware.com public.tableausoftware.com

 [7] public.tableausoftware.com (direct)

 [9] (direct)              Others

[11] Others                visualisingdata.com

[13] Others                public.tableausoftware.com

[15] (direct)              (direct)

[17] Others                google

[19] google                (direct)

[21] (direct)              Others

[23] Others                visualisingdata.com

[25] (direct)              Others
```

[27] public.tableausoftware.com Others

[29] google   google

[31] google   (direct)

[33] google   google

[35] google   google

[37] (direct)   Others

[39] Others   (direct)

[41] Others   Others

[43] google   google

[45] google   google

[47] google   google

[49] Others   tableausoftware.com

[51] Others   Others

[53] google   Others

[55] (direct)   Others

[57] Others   Others

[59] Others   (direct)

[61] (direct)   (direct)

[63] (direct)   (direct)

[65] (direct)   t.co

[67] t.co   t.co

[69] Others   Others

[71] (direct)   visualisingdata.com

[73] visualisingdata.com   t.co

[75] t.co   (direct)

 [77] t.co   Others

 [79] (direct)   visualisingdata.com

 [81] t.co   t.co

 [83] t.co   t.co

 [85] (direct)   google

 [87] (direct)   visualisingdata.com

 [89] visualisingdata.com   (direct)

 [91] (direct)   t.co

 [93] t.co   visualisingdata.com

 [95] google   (direct)

 [97] (direct)   google

 [99] google   visualisingdata.com

[101] (direct)   Others

[103] google   (direct)

[105] google   (direct)

[107] (direct)   Others

[109] google   visualisingdata.com

[111] google   Others

[113] Others   Others

[115] (direct)   (direct)

[117] Others   Others

[119] google   Others

[121] google   google

[123] google   google

[125] google   google

[127] google   google

[129] google   public.tableausoftware.com

[131] visualisingdata.com   visualisingdata.com

[133] visualisingdata.com   visualisingdata.com

[135] Others   t.co

[137] t.co   t.co

[139] t.co   t.co

[141] t.co   t.co

[143] t.co   t.co

[145] t.co   visualisingdata.com

[147] (direct)   t.co

[149] t.co   google

[151] t.co   t.co

[153] Others   Others

[155] (direct)   public.tableausoftware.com

[157] tableausoftware.com   Others

[159] Others   public.tableausoftware.com

[161] (direct)   public.tableausoftware.com

[163] Others   tableausoftware.com

[165] (direct)   (direct)

[167] google   google

[169] Others   google

[171] google   tableausoftware.com

[173] visualisingdata.com   (direct)

[175] (direct)   Others

[177] google   google

[179] google   google

[181] google   google

[183] google   google

[185] google   google

[187] google   visualisingdata.com

[189] Others   Others

[191] Others   Others

[193] t.co   (direct)

[195] Others   google

[197] tableausoftware.com   tableausoftware.com

[199] tableausoftware.com   tableausoftware.com

[201] tableausoftware.com   (direct)

[203] (direct)   visualisingdata.com

[205] google   t.co

[207] t.co   t.co

[209] t.co   t.co

[211] tableausoftware.com   tableausoftware.com

[213] tableausoftware.com   google

[215] Others   t.co

[217] t.co   visualisingdata.com

[219] (direct)   google

[221] visualisingdata.com   visualisingdata.com

[223] Others   t.co

[225] tableausoftware.com   (direct)

[227] (direct)   (direct)

[229] google   t.co

[231] Others   google

[233] (direct)   google

[235] Others   tableausoftware.com

[237] (direct)   Others

[239] google   google

[241] (direct)   google

[243] google   public.tableausoftware.com

[245] public.tableausoftware.com tableausoftware.com

[247] (direct)   (direct)

[249] Others   Others

[251] Others   Others

[253] google   google

[255] google   google

[257] google   public.tableausoftware.com

[259] tableausoftware.com   tableausoftware.com

[261] tableausoftware.com   visualisingdata.com

[263] t.co   t.co

[265] t.co   t.co

[267] (direct)   t.co

[269] google   Others

[271] (direct)   google

[273] (direct)   (direct)

[275] google   Others

[277] google   google

[279] google   Others

[281] Others   t.co

[283] t.co   t.co

[285] t.co   t.co

[287] t.co   t.co

[289] t.co   t.co

[291] t.co   Others

[293] (direct)   (direct)

[295] google   google

[297] google   Others

[299] t.co   t.co

[301] t.co   t.co

[303] (direct)   (direct)

[305] (direct)   (direct)

[307] (direct)   (direct)

[309] google   google

[311] google   Others

[313] t.co   t.co

[315] t.co   t.co

[317] (direct)   (direct)

[319] google   google

[321] Others   Others

[323] t.co   (direct)

[325] Others   Others

[327] t.co   t.co

[329] google   visualisingdata.com

[331] (direct)   (direct)

[333] google   Others

[335] visualisingdata.com   Others

[337] google   (direct)

[339] google   (direct)

[341] visualisingdata.com   google

[343] google   google

[345] Others   Others

[347] google   Others

[349] (direct)   (direct)

[351] (direct)   (direct)

[353] (direct)   (direct)

[355] (direct)   (direct)

[357] Others   Others

[359] google   google

[361] google   google

[363] google   google

[365] google   google

[367] google   google

[369] Others   tableausoftware.com

[371] tableausoftware.com   visualisingdata.com

[373] Others   t.co

[375] t.co   t.co

[377] t.co   t.co

[379] t.co   t.co

[381] t.co   t.co

[383] t.co   t.co

[385] t.co   t.co

[387] t.co   t.co

[389] tableausoftware.com   t.co

[391] t.co   Others

[393] t.co   t.co

[395] t.co   t.co

[397] t.co   t.co

[399] t.co   t.co

[401] t.co   Others

[403] Others   Others

[405] Others   Others

[407] google   t.co

[409] t.co   Others

[411] google   t.co

[413] t.co   (direct)

[415] google   (direct)

[417] google   Others

[419] Others   Others

[421] google   google

[423] (direct)   Others

[425] public.tableausoftware.com Others

[427] (direct)   Others

[429] google   (direct)

[431] (direct)   (direct)

[433] (direct)   (direct)

[435] Others   Others

[437] google   google

[439] google   google

[441] google   Others

[443] visualisingdata.com Others

[445] Others   Others

[447] Others   t.co

[449] t.co   visualisingdata.com

[451] (direct)   Others

[453] Others   t.co

[455] (direct)   Others

[457] google   tableausoftware.com

[459] Others   tableausoftware.com

[461] t.co   tableausoftware.com

[463] Others   tableausoftware.com

[465] t.co   google

[467] google   google

[469] t.co   Others

[471] (direct)   visualisingdata.com

[473] (direct)   Others

[475] google   (direct)

[477] (direct)   (direct)

[479] (direct)   Others

[481] Others   google

[483] google   google

[485] google   google

[487] google   google

[489] tableausoftware.com   Others

[491] t.co   t.co

[493] tableausoftware.com   (direct)

[495] t.co   (direct)

[497] Others   Others

[499] visualisingdata.com   visualisingdata.com

[501] visualisingdata.com   Others

[503] Others   t.co

[505] Others   Others

[507] t.co   t.co

[509] Others   google

[511] Others   visualisingdata.com

[513] visualisingdata.com   google

[515] google   google

[517] (direct)   visualisingdata.com

[519] (direct)   (direct)

[521] (direct)   google

[523] Others   visualisingdata.com

[525] google   google

[527] visualisingdata.com   Others

[529] (direct)   (direct)

[531] google   google

[533] google   google

[535] Others   visualisingdata.com

[537] Others   Others

[539] t.co   Others

[541] t.co   google

[543] (direct)   (direct)

[545] (direct)   (direct)

[547] (direct)   (direct)

[549] (direct)   t.co

[551] (direct)   (direct)

[553] google   Others

[555] Others   (direct)

[557] (direct)   t.co

[559] t.co   google

[561] Others   Others

[563] (direct)   t.co

[565] (direct)   t.co

[567] google   (direct)

[569] google   google

[571] google   google

[573] google   (direct)

[575] Others   google

[577] (direct)   Others

[579] Others   google

[581] (direct)   (direct)

[583] google   (direct)

[585] Others   visualisingdata.com

[587] google   (direct)

[589] google   google

[591] google   (direct)

[593] (direct)   (direct)

[595] (direct)   (direct)

[597] Others   google

[599] google   google

[601] google   google

[603] tableausoftware.com   tableausoftware.com

[605] tableausoftware.com   Others

[607] Others   t.co

[609] t.co   t.co

[611] google   (direct)

[613] google   Others

[615] t.co   t.co

[617] google   (direct)

[619] Others   (direct)

[621] (direct)   t.co

[623] t.co   google

[625] google   google

[627] Others   google

[629] t.co   t.co

[631] (direct)   Others

[633] Others   google

[635] (direct)   (direct)

[637] google   tableausoftware.com

[639] (direct)   Others

[641] google   google

[643] visualisingdata.com   (direct)

[645] google   google

[647] google   google

[649] google   google

[651] google   google

[653] google   google

[655] google   google

[657] google   google

[659] google   google

[661] google   google

[663] google   Others

[665] t.co   (direct)

[667] (direct)   (direct)

[669] (direct)   (direct)

[671] (direct)   (direct)

[673] (direct)   (direct)

[675] (direct)   (direct)

[677] (direct)   google

[679] google   google

[681] google   (direct)

[683] google   google

[685] (direct)   Others

[687] (direct)   (direct)

[689] (direct)   Others

[691] google   google

[693] google   (direct)

[695] Others   (direct)

[697] (direct)   (direct)

[699] (direct)	(direct)
[701] Others	google
[703] google	Others
[705] google	google
[707] google	(direct)
[709] (direct)	(direct)
[711] google	google
[713] google	google
[715] Others	Others
[717] google	t.co
[719] google	(direct)
[721] (direct)	Others
[723] google	google
[725] google	google
[727] google	google
[729] google	Others
[731] (direct)	(direct)
[733] Others	google
[735] google	google
[737] google	google
[739] google	google
[741] google	google
[743] google	google
[745] google	google

[747] Others   (direct)

[749] google   (direct)

[751] google   Others

[753] google   google

[755] Others   google

[757] (direct)   Others

[759] (direct)   (direct)

[761] (direct)   (direct)

[763] Others   (direct)

[765] (direct)   (direct)

[767] google   google

[769] google   Others

[771] Others   google

[773] google   google

[775] (direct)   Others

[777] Others   Others

[779] google   Others

[781] visualisingdata.com   google

[783] google   Others

[785] (direct)   Others

[787] (direct)   (direct)

[789] Others   Others

[791] google   google

[793] google   google

[795] google   google

[797] google   google

[799] Others   Others

[801] google   google

[803] google   Others

[805] google   Others

[807] google   google

[809] Others   (direct)

[811] (direct)   google

[813] google   google

[815] google   google

[817] Others   google

[819] visualisingdata.com   Others

[821] visualisingdata.com   Others

[823] visualisingdata.com   google

[825] google   Others

[827] google   google

[829] google   google

[831] google   (direct)

[833] google   google

[835] (direct)   Others

[837] Others   google

[839] google   google

[841] visualisingdata.com   Others

[843] Others   google

[845] google   google

[847] google   Others

[849] t.co   t.co

[851] google   google

[853] (direct)   Others

[855] google   t.co

[857] google   google

[859] google   google

[861] (direct)   Others

[863] google   t.co

[865] google   google

[867] google   Others

[869] google   google

[871] (direct)   google

[873] (direct)   google

[875] (direct)   Others

[877] google   Others

[879] (direct)   Others

[881] google   google

[883] google   google

[885] google   google

[887] google   google

[889] google   google

[891] Others   google

[893] Others   (direct)

[895] tableausoftware.com   google

[897] google   google

[899] google   google

[901] google   google

[903] google   Others

[905] google   Others

[907] google   google

[909] google   google

[911] google   (direct)

[913] public.tableausoftware.com google

[915] google   (direct)

[917] (direct)   (direct)

[919] (direct)   (direct)

[921] (direct)   Others

[923] Others   Others

[925] google   google

[927] google   google

[929] google   google

[931] google   google

[933] Others   Others

[935] Others   (direct)

[937] google   Others

[939] Others   Others

[941] Others   public.tableausoftware.com

[943] Others   visualisingdata.com

[945] google   public.tableausoftware.com

[947] Others   google

[949] Others   visualisingdata.com

[951] google   public.tableausoftware.com

[953] Others   google

[955] google   (direct)

[957] google   google

[959] google   google

[961] google   google

[963] google   google

[965] google   google

[967] google   google

[969] Others   (direct)

[971] google   Others

[973] google   google

[975] google   google

[977] google   google

[979] Others   Others

[981] t.co   Others

[983] google   tableausoftware.com

[985] tableausoftware.com   public.tableausoftware.com

```
[987] tableausoftware.com       public.tableausoftware.com

[989] tableausoftware.com       tableausoftware.com

[991] tableausoftware.com       public.tableausoftware.com

[993] public.tableausoftware.com tableausoftware.com

[995] tableausoftware.com       public.tableausoftware.com

[997] tableausoftware.com       tableausoftware.com

[999] Others                (direct)

[ reached getOption("max.print") -- omitted 31109 entries ]

9 Levels: (direct) facebook google Others ... visualisingdata.com

> levels(data$Sourcegroup)

NULL

> table(data$Sourcegroup)


        (direct)              facebook              google

          7532                  92                 11542

          Others public.tableausoftware.com           reddit.com

          5360                 1354                  616

          t.co      tableausoftware.com      visualisingdata.com

          2249                 2388                  976

> data$Sourcegroup=factor(data$Sourcegroup,

+           levels =
c("(direct)","google","public.tableausoftware.com","t.co","visualisingdata.com","facebook","Ot
hers","reddit.com","tableausoftware.com"),

+           labels = c(0,1,2,3,4,5,6,7,8))

> data
```

A tibble: 32,109 x 8

Bounces Exits Continent Sourcegroup Timeinpage Uniquepageviews Visits BouncesNew

<dbl> <dbl> <fct> <fct> <dbl> <dbl> <dbl> <dbl>

	Bounces	Exits	Continent	Sourcegroup	Timeinpage	Uniquepageviews	Visits	BouncesNew
1	0	0	4	0	18	1	0	0
2	0	0	3	0	4	1	0	0
3	0	0	3	6	35	1	0	0
4	0	0	3	2	70	1	0	0
5	0	0	3	2	81	1	0	0
6	0	0	3	2	75	1	0	0
7	0	0	3	2	186	1	0	0
8	0	0	3	0	710	1	0	0
9	0	0	4	0	712	1	1	0
10	0	0	1	6	344	1	1	0

... with 32,099 more rows

> data$Timeinpage=scale(data$Timeinpage)

> View(data)

> chisq.test(data$Uniquepageviews,data$Visits)

        Pearson's Chi-squared test

data: data$Uniquepageviews and data$Visits

X-squared = 437229, df = 272, p-value < 2.2e-16

Warning message:

In chisq.test(data$Uniquepageviews, data$Visits) :

  Chi-squared approximation may be incorrect

> factor(data$Exits)

  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1

 [40] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

 [79] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[118] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 0 0 0 0

[157] 0 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2

[196] 2 2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[235] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 0 0 0

[274] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1

[313] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[352] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2

[391] 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[430] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 3 0 0 0 0 0 0 1 1 1 1 1 1 1

[469] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 0 0 0 0 0 0 0 0 1 1 1

[508] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 0 0 0 0

[547] 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[586] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 0 0 0 0 0 0 0 0 0 0 0

[625] 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[664] 1 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1

[703] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[742] 1 1 1 1 1 1 1 2 2 2 2 2 3 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[781] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 0

[820] 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 0 0 0 0 0 0 0 0 0 0 0 1

[859] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 0

[898] 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2

[937] 2 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[976] 1 1 1 1 1 1 1 2 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1

[reached getOption("max.print") -- omitted 31109 entries]

Levels: 0 1 2 3 4 5 6 7 8 9 10 12 15 27 33 36

> chisq.test(data$Exits,data$Continent)

        Pearson's Chi-squared test

data: data$Exits and data$Continent

X-squared = 189.66, df = 75, p-value = 6.733e-12

Warning message:

In chisq.test(data$Exits, data$Continent) :

  Chi-squared approximation may be incorrect

> chisq.test(data$Exits,data$Sourcegroup)

        Pearson's Chi-squared test

data: data$Exits and data$Sourcegroup

X-squared = 1748.6, df = 120, p-value < 2.2e-16

Warning message:

In chisq.test(data$Exits, data$Sourcegroup) :

  Chi-squared approximation may be incorrect

> chisq.test(data$Exits,data$Timeinpage)

        Pearson's Chi-squared test

data: data$Exits and data$Timeinpage

X-squared = 180718, df = 20160, p-value < 2.2e-16

Warning message:

In chisq.test(data$Exits, data$Timeinpage) :

  Chi-squared approximation may be incorrect

> chisq.test(data$Exits,data$Uniquepageviews)

        Pearson's Chi-squared test

data: data$Exits and data$Uniquepageviews

X-squared = 386724, df = 255, p-value < 2.2e-16

Warning message:

In chisq.test(data$Exits, data$Uniquepageviews) :

  Chi-squared approximation may be incorrect

> chisq.test(data$Exits,data$Visits)

```
        Pearson's Chi-squared test


data:  data$Exits and data$Visits

X-squared = 368000, df = 240, p-value < 2.2e-16


Warning message:

In chisq.test(data$Exits, data$Visits) :

  Chi-squared approximation may be incorrect

> chisq.test(data$Exits,data$Bounces)


        Pearson's Chi-squared test


data:  data$Exits and data$Bounces

X-squared = 280325, df = 195, p-value < 2.2e-16


Warning message:

In chisq.test(data$Exits, data$Bounces) :

  Chi-squared approximation may be incorrect

> library(ggplot2)

> ggplot(data,aes(x=Bounces,y=Visits))+geom_point(color="red",shape=3)

> library(caTools)

> set.seed(123)

> train=data[1:80,]
```

```
> test=data[1:100,]

> model=lm(Bounces~.,train)

> summary(model)


Call:

lm(formula = Bounces ~ ., data = train)


Residuals:

     Min        1Q     Median       3Q        Max

-3.290e-15 -1.049e-16  3.300e-17  1.508e-16  6.208e-16


Coefficients:

             Estimate Std. Error    t value Pr(>|t|)

(Intercept)  -6.609e-16  3.014e-16 -2.193e+00   0.0320 *

Exits         7.581e-16  1.646e-16  4.605e+00 2.02e-05 ***

Continent2    1.311e-16  2.608e-16  5.020e-01   0.6171

Continent3   -5.814e-17  2.316e-16 -2.510e-01   0.8026

Continent4    1.233e-16  3.338e-16  3.690e-01   0.7131

Continent5    1.246e-16  2.847e-16  4.370e-01   0.6633

Sourcegroup1 -2.345e-17  1.870e-16 -1.250e-01   0.9006

Sourcegroup2  4.671e-16  2.365e-16  1.975e+00   0.0526 .

Sourcegroup3  6.338e-18  2.726e-16  2.300e-02   0.9815

Sourcegroup4 -4.263e-16  2.794e-16 -1.526e+00   0.1320

Sourcegroup6  1.514e-16  1.654e-16  9.150e-01   0.3635
```

Sourcegroup8   -6.566e-16 5.136e-16 -1.278e+00   0.2057

Timeinpage   -4.150e-17 1.105e-16 -3.760e-01   0.7085

Uniquepageviews 4.629e-17 2.130e-16 2.170e-01   0.8287

Visits   3.255e-16 1.723e-16 1.889e+00   0.0634 .

BouncesNew   1.000e+02 2.212e-14 4.521e+15 < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.859e-16 on 64 degrees of freedom

Multiple R-squared: 1,   Adjusted R-squared: 1

F-statistic: 1.157e+31 on 15 and 64 DF, p-value: < 2.2e-16

Warning message:

In summary.lm(model) : essentially perfect fit: summary may be unreliable

> factor(data$Continent)

  [1] 4 3 3 3 3 3 3 3 3 4 1 2 2 3 3 5 3 3 3 3 4 5 1 2 2 3 3 3 3 1 2 3 4 3 3 3 3 2 2 2

 [40] 3 3 3 3 3 3 3 3 3 3 3 3 1 2 2 3 5 5 5 5 3 3 3 3 3 2 2 2 5 5 1 2 3 2 3 3 2 5

 [79] 1 3 2 2 2 3 3 1 4 1 2 3 3 1 2 5 4 5 3 3 3 3 2 2 1 4 4 2 2 2 1 1 2 2 1 1 2 2 2

[118] 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 5 2 2 2 2 2 4 2 3 3 3 2 2 2 3 2 2 5 2 2 3 3

[157] 2 2 2 3 3 4 2 2 3 4 4 5 3 2 2 1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 5 3 3 3 5 3 2

[196] 2 3 3 2 2 2 2 2 2 3 4 2 3 3 3 3 2 2 2 3 3 3 3 2 3 3 3 3 3 2 2 2 3 4 2 5 4 5 5

[235] 3 3 2 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 5 5 3 3 2 3 2 3 2 2 2

[274] 2 2 2 2 2 2 3 3 2 2 2 1 1 1 3 3 3 3 3 3 2 2 2 3 3 2 1 1 3 2 2 4 3 2 2 2 2 3 3

[313] 2 1 1 3 2 4 3 2 2 2 3 2 2 3 3 3 3 4 3 3 3 2 2 2 1 2 4 2 2 1 2 2 2 1 2 2 3 3 3

[352] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 5 3 2 4 4 4 2 2 2 3 3 3 3 3 5 3

[391] 3 3 2 2 2 2 2 2 2 2 2 2 3 2 2 2 3 3 2 2 3 3 2 2 1 1 3 3 4 3 2 2 2 2 1 1 2 2 2

[430] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 2 2 3 1 2 4 3 3 4 3 3 2 3 3 2 2 3 3 2 2 3 2 1 3

[469] 3 2 2 3 2 2 1 4 2 3 3 3 3 3 3 3 3 3 3 3 5 5 3 2 3 3 1 2 2 1 1 1 2 3 3 2 3 2

[508] 3 2 5 1 1 1 1 3 3 4 4 2 2 2 1 2 1 4 1 2 2 3 3 3 3 3 3 3 3 3 2 2 2 2 2 3 3 3

[547] 3 3 3 3 3 3 3 2 2 3 3 3 3 3 2 2 3 3 3 3 2 1 1 1 2 1 2 3 3 3 3 1 4 4 3 3 0 2 2

[586] 2 3 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 2 3 2 3 3 3 1 3 2 3 3 3 3 1

[625] 2 1 2 2 3 3 3 2 3 1 3 3 3 1 2 2 1 4 1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

[664] 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3

[703] 3 2 2 2 2 3 3 3 3 3 3 4 2 3 3 3 3 3 4 4 2 3 2 1 2 0 2 3 3 3 3 3 3 3 3 3 3 3

[742] 3 3 3 3 3 3 3 3 3 3 2 3 4 3 3 3 4 3 3 3 3 4 3 3 3 3 3 2 3 3 3 3 2 2 1 1 4 4 1

[781] 0 1 1 1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 2 3 2 3 2 2 3 3 3 3 3 3 3 3 2 3

[820] 3 3 3 3 3 3 3 3 3 4 2 2 1 1 1 3 3 3 3 3 3 3 2 3 4 2 1 1 3 2 2 1 2 2 3 3 2 3 1

[859] 4 2 2 3 3 2 3 2 2 3 3 2 4 4 2 2 4 2 0 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 3

[898] 3 3 3 3 3 3 1 3 1 3 4 2 3 3 2 1 0 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 1 2

[937] 3 3 3 3 3 3 3 2 3 3 3 2 3 2 3 3 3 2 3 3 3 3 2 4 3 3 3 2 1 3 4 2 1 1 2 3 3 3

[976] 3 3 3 2 3 2 3 3 2 2 3 3 2 2 3 2 3 2 2 2 3 3 3 1 3

[reached getOption("max.print") -- omitted 31109 entries]

Levels: 0 1 2 3 4 5

```
> cor(data$Uniquepageviews,data$Visits)
[1] 0.8144457
> anov<-aov(Uniquepageviews~Visits,data = data)
> summary(anov)
        Df Sum Sq Mean Sq F value Pr(>F)
```

```
Visits        1  8052  8052  63257 <2e-16 ***

Residuals  32107  4087      0

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anoo<-aov(Exits~.,data = data)

> summary(anoo)

                 Df Sum Sq Mean Sq   F value   Pr(>F)

Bounces          1  10578   10578 1.043e+05  < 2e-16 ***

Continent        5      3       1 5.960e+00 1.62e-05 ***

Sourcegroup      8      7       1 8.760e+00 4.89e-12 ***

Timeinpage       1    130     130 1.279e+03  < 2e-16 ***

Uniquepageviews  1   1573    1573 1.552e+04  < 2e-16 ***

Visits           1      1       1 5.014e+00   0.0251 *

Residuals    32091   3254       0

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> newModel=lm(Exits~Bounces+Sourcegroup+Timeinpage+Visits,train)

> summary(newModel)


Call:

lm(formula = Exits ~ Bounces + Sourcegroup + Timeinpage + Visits,

   data = train)


Residuals:
```

```
     Min     1Q  Median     3Q     Max
-0.38226 -0.25419 -0.14231  0.09956  1.54803


Coefficients:

        Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.32836   0.09800   3.350  0.0013 **

Bounces      1.00329   0.12999   7.718 6.04e-11 ***

Sourcegroup1 0.15684   0.14386   1.090  0.2794

Sourcegroup2 -0.11913   0.19366  -0.615  0.5404

Sourcegroup3 -0.19379   0.19460  -0.996  0.3228

Sourcegroup4 0.21070   0.21757   0.968  0.3362

Sourcegroup6 0.04503   0.12371   0.364  0.7170

Sourcegroup8 -0.09735   0.42706  -0.228  0.8203

Timeinpage   -0.05289   0.08965  -0.590  0.5571

Visits       -0.24411   0.12608  -1.936  0.0569 .

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.4143 on 70 degrees of freedom

Multiple R-squared:  0.7231,  Adjusted R-squared:  0.6875

F-statistic: 20.31 on 9 and 70 DF,  p-value: 2.739e-16


> predExits=predict(newModel,test)

> View(predExits)
```

```
> predExits=round(predExits)

> View(predExits)

> factor(predExits)
```

 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42

 0 0 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1

 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63

 1 1 1 1 1 1 1 1 1 2 2 3 3 0 0 0 0 0 0 0 0

 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84

 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0

 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1

Levels: 0 1 2 3

```
> final_data=cbind(test,predExits)

> View(final_data)

> ggplot(data,aes(x=Bounces+Timeinpage,y=Exits))+geom_point(color="red",shape=3)

> anooo<-aov(Timeinpage~.,data = data)

> summary(anooo)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Bounces	1	382	382.2	422.868	< 2e-16	***
Exits	1	838	838.1	927.283	< 2e-16	***
Continent	5	31	6.1	6.780	2.51e-06	***
Sourcegroup	8	10	1.2	1.374	0.202	

```
Uniquepageviews   1   1151  1151.4 1273.826  < 2e-16 ***

Visits        1   690   689.8  763.163  < 2e-16 ***

Residuals     32091  29006   0.9

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> library(psych)

> describe(data)

          vars    n mean  sd median trimmed  mad   min   max  range  skew

Bounces        1 32109 0.71 0.71  1.00   0.68 0.00  0.00 30.00 30.00  7.06

Exits         2 32109 0.91 0.70  1.00   0.89 0.00  0.00 36.00 36.00 11.11

Continent*     3 32109 3.66 0.84  4.00   3.73 0.00  1.00  6.00  5.00 -0.46

Sourcegroup*    4 32109 3.52 2.66  2.00   3.19 1.48  1.00  9.00  8.00  0.86

Timeinpage     5 32109 0.00 1.00 -0.19  -0.16 0.00 -0.19 118.32 118.51 57.33

Uniquepageviews  6 32109 1.11 0.61  1.00   1.00 0.00  1.00 45.00 44.00 24.41

Visits        7 32109 0.91 0.73  1.00   0.88 0.00  0.00 45.00 45.00 13.75

BouncesNew     8 32109 0.01 0.01  0.01   0.01 0.00  0.00  0.30  0.30  7.06

        kurtosis   se

Bounces       219.35 0.00

Exits        427.57 0.00

Continent*      1.27 0.00

Sourcegroup*    -0.78 0.01

Timeinpage    6234.83 0.01

Uniquepageviews  1273.98 0.00

Visits        633.85 0.00
```

BouncesNew   219.35 0.00

```
> summary(data)
```

Bounces	Exits	Continent	Sourcegroup	Timeinpage.V1
Min. : 0.000	Min. : 0.000	0: 321	1 :11542	Min. : -0.18554
1st Qu.: 0.000	1st Qu.: 1.000	1: 3171	0 : 7532	1st Qu.: -0.18554
Median : 1.000	Median : 1.000	2: 6470	6 : 5360	Median : -0.18554
Mean : 0.713	Mean : 0.906	3:20043	8 : 2388	Mean : 0.00000
3rd Qu.: 1.000	3rd Qu.: 1.000	4: 1356	3 : 2249	3rd Qu.: -0.16019
Max. :30.000	Max. :36.000	5: 748	2 : 1354	Max. :118.32391
			(Other): 1684	

Uniquepageviews	Visits	BouncesNew
Min. : 1.000	Min. : 0.000	Min. :0.00000
1st Qu.: 1.000	1st Qu.: 1.000	1st Qu.:0.00000
Median : 1.000	Median : 1.000	Median :0.01000
Mean : 1.114	Mean : 0.906	Mean :0.00713
3rd Qu.: 1.000	3rd Qu.: 1.000	3rd Qu.:0.01000
Max. :45.000	Max. :45.000	Max. :0.30000

```
> #one way annova of bounces

> aov1<-aov(Timeinpage~Bounces,data = data)

> summary(aov1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bounces	1	382	382.2	386.8	<2e-16 ***
Residuals	32107	31726	1.0		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> print(model.tables(aov1,"means"),digits = 3)

Tables of means

Grand mean

1.227386e-15

 Bounces

Bounces

  0   1   2   3   4   5   6   7   8   9   12   24   29   30

 0.11 -0.04 -0.20 -0.35 -0.51 -0.66 -0.81 -0.97 -1.12 -1.28 -1.74 -3.59 -4.36 -4.51

Warning message:

In replications(paste("~", xx), data = mf) : non-factors ignored: Bounces

> #one way annova of Continent

> aov2<-aov(Timeinpage~Continent,data = data)

> summary(aov2)

          Df Sum Sq Mean Sq F value   Pr(>F)

Continent   5   35   6.917   6.923 1.81e-06 ***

Residuals   32103 32073   0.999

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> print(model.tables(aov2,"means"),digits = 3)

Tables of means

Grand mean

1.227386e-15

 Continent

      0   1   2   3   4   5

   -0.0455   -0.0313   -0.0325 1.11e-02   -0.0176   0.168

rep 321.0000 3171.0000 6470.0000 2.00e+04 1356.0000 748.000

> aov3<-aov(Timeinpage~Bounces*Continent,data = data)

> summary(aov3)

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Bounces        1    382   382.2 388.503  < 2e-16 ***
Continent      5    34    6.7   6.846 2.16e-06 ***
Bounces:Continent    5    115   22.9  23.296  < 2e-16 ***
Residuals      32097  31578   1.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> print(model.tables(aov3,"means"),digits = 3)

Tables of means

Grand mean

1.227386e-15

 Bounces

Error in dimnames(x) <- dn :

   length of 'dimnames' [1] not equal to array extent

In addition: Warning messages:

1: In replications(paste("~", xx), data = mf) :

   non-factors ignored: Bounces

2: In replications(paste("~", xx), data = mf) :

   non-factors ignored: Bounces, Continent

> #Corelation

> library(ggplot2)

> library(GGally)

> cor(data$Timeinpage,data$Bounces)

          [,1]

[1,] -0.1091057

> cor(data[,5:6])

          Timeinpage Uniquepageviews

Timeinpage   1.0000000   0.1145925

Uniquepageviews 0.1145925   1.0000000

> ggcorr(data,label=TRUE,label_alpha=TRUE)

Warning message:

In ggcorr(data, label = TRUE, label_alpha = TRUE) :

   data in column(s) 'Continent', 'Sourcegroup' are not numeric and were ignored

> qplot(Visits,Timeinpage,data = data,geom =
c("point","smooth"),method="lm",alpha=I(1/5),se=FALSE)

`geom_smooth()` using formula 'y ~ x'

Warning message:

Ignoring unknown parameters: method, se

> qplot(Exits,Timeinpage,data = data,geom = c("point","smooth"),alpha=I(1/5))

`geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

> ggpairs(data,columns = c("Timeinpage","Bounces","Exits"),upper = list(continuous=wrap("cor",size=10)),lower = list(continuous="smooth"))

> data$Bounces=data$Bounces*0.01

> rmm<-glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits,data = data,family = "binomial")

Warning messages:

1: In eval(family$initialize) : non-integer #successes in a binomial glm!

2: glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(rmm)

Call:

glm(formula = Bounces ~ Timeinpage + Continent + Exits + Sourcegroup +

   Uniquepageviews + Visits, family = "binomial", data = data)

Deviance Residuals:

   Min   1Q   Median   3Q   Max

-2.26149 -0.02406   0.00206   0.00895   1.81288

Coefficients:

         Estimate Std. Error z value Pr(>|z|)

(Intercept)   -5.042106   0.679901 -7.416 1.21e-13 ***

```
Timeinpage       -0.406045  0.227751  -1.783  0.0746 .

Continent1        0.002277  0.693204   0.003  0.9974

Continent2       -0.006924  0.678660  -0.010  0.9919

Continent3        0.010133  0.667419   0.015  0.9879

Continent4        0.020112  0.733367   0.027  0.9781

Continent5        0.023751  0.791425   0.030  0.9761

Exits             1.390761  0.335650   4.143 3.42e-05 ***

Sourcegroup1     -0.078363  0.172016  -0.456  0.6487

Sourcegroup2     -0.252828  0.492312  -0.514  0.6076

Sourcegroup3      0.014869  0.276016   0.054  0.9570

Sourcegroup4     -0.082252  0.461487  -0.178  0.8585

Sourcegroup5     -0.024195  1.104517  -0.022  0.9825

Sourcegroup6     -0.076792  0.218269  -0.352  0.7250

Sourcegroup7     -0.009279  0.470930  -0.020  0.9843

Sourcegroup8     -0.112930  0.319076  -0.354  0.7234

Uniquepageviews  -3.236311  0.579166  -5.588 2.30e-08 ***

Visits            2.194112  0.520222   4.218 2.47e-05 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 234.937  on 32108  degrees of freedom

Residual deviance:  96.514  on 32091  degrees of freedom
```

AIC: 506.56

Number of Fisher Scoring iterations: 11