Kira Plastinina Customer Analysis

STEPHEN ODHIAMBO OGAJA

2022-06-05

1. Problem Definition

a) Specifying the Question

What are the customer behaviours of different customer groups?

b) Defining the metric of success

The project will be considered a success when we can accurately come up with an understanding of the different customer groups and their characteristics.

c) Understanding the context

Kira Plastinina (https://kiraplastinina.ru/) is a Russian brand that is sold through a defunct chain of retail stores in Russia, Ukraine, Kazakhstan, Belarus, China, Philippines, and Armenia. The brand's Sales and Marketing team would like to understand their customer's behavior from data that they have collected over the past year. More specifically, they would like to learn the characteristics of customer groups.

d) Recording the experimental design

```
1. Data sourcing/loading
2. Data Understanding
3. Data Relevance
4. External Dataset Validation
5. Data Preparation
6. Univariate Analysis
7. Bivariate Analysis
8. Multivariate Analysis
9. Implementing the solution
10. Challenging the solution
11. Conclusion
12. Follow up questions
```

e) Data Relevance

For the data to be relevant it should be able to provide meaningful insights that can be used to learn the customer behavours of the different customer groups and their characteristics.

2. Data Sourcing

Libraries

```
#install.packages('VIM')
#install.packages('CatEncoders')
library(CatEncoders)
```

```
##
## Attaching package: 'CatEncoders'

## The following object is masked from 'package:base':
##
##     transform
```

```
#install.packages("readxl")
library(readxl)
```

```
# let's load the required libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
#install.packages("e1071")
library(e1071)
library(magrittr)
library(knitr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v stringr 1.4.0
## v tidyr   1.2.0     v forcats 0.5.1

## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x data.table::between() masks dplyr::between()
## x tidyr::extract()      masks magrittr::extract()
## x dplyr::filter()       masks stats::filter()
## x data.table::first()   masks dplyr::first()
## x dplyr::lag()          masks stats::lag()
## x data.table::last()    masks dplyr::last()
## x purrr::set_names()    masks magrittr::set_names()
## x purrr::transpose()    masks data.table::transpose()
```

```
#install.packages("factoextra")
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
#install.packages("devtools")
library(devtools)
```

```
## Loading required package: usethis
```

```
#install.packages("Hmisc")
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following object is masked from 'package:e1071':
##
##     impute

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(CatEncoders)
```

```
library(data.table)
shopping <- fread("http://bit.ly/EcommerceCustomersDataset")
head(shopping)
```

a) Reading the Data

```
##    Administrative Administrative_Duration Informational Informational_Duration
## 1:              0                       0             0                      0
## 2:              0                       0             0                      0
## 3:              0                      -1             0                     -1
## 4:              0                       0             0                      0
## 5:              0                       0             0                      0
## 6:              0                       0             0                      0
##    ProductRelated ProductRelated_Duration BounceRates ExitRates PageValues
## 1:              1                0.000000  0.20000000 0.2000000          0
## 2:              2               64.000000  0.00000000 0.1000000          0
## 3:              1               -1.000000  0.20000000 0.2000000          0
## 4:              2                2.666667  0.05000000 0.1400000          0
## 5:             10              627.500000  0.02000000 0.0500000          0
## 6:             19              154.216667  0.01578947 0.0245614          0
##    SpecialDay Month OperatingSystems Browser Region TrafficType
## 1:          0   Feb                1       1      1           1
## 2:          0   Feb                2       2      1           2
## 3:          0   Feb                4       1      9           3
## 4:          0   Feb                3       2      2           4
## 5:          0   Feb                3       3      1           4
## 6:          0   Feb                2       2      1           3
##          VisitorType Weekend Revenue
## 1: Returning_Visitor   FALSE   FALSE
## 2: Returning_Visitor   FALSE   FALSE
## 3: Returning_Visitor   FALSE   FALSE
## 4: Returning_Visitor   FALSE   FALSE
## 5: Returning_Visitor    TRUE   FALSE
## 6: Returning_Visitor   FALSE   FALSE
```

```
# Number of records
cat('Number of rows = ', nrow(shopping), 'and the number of columns = ', ncol(shopping),'.')
```

b) Checking the Data

```
## Number of rows =  12330 and the number of columns =  18 .
```

```
# top dataset preview
head(shopping, 5)
```

```
##    Administrative Administrative_Duration Informational Informational_Duration
## 1:              0                       0             0                      0
## 2:              0                       0             0                      0
## 3:              0                      -1             0                     -1
## 4:              0                       0             0                      0
## 5:              0                       0             0                      0
##    ProductRelated ProductRelated_Duration BounceRates ExitRates PageValues
## 1:              1                0.000000        0.20      0.20          0
## 2:              2               64.000000        0.00      0.10          0
## 3:              1               -1.000000        0.20      0.20          0
## 4:              2                2.666667        0.05      0.14          0
## 5:             10              627.500000        0.02      0.05          0
##    SpecialDay Month OperatingSystems Browser Region TrafficType
## 1:          0   Feb                1       1      1           1
## 2:          0   Feb                2       2      1           2
## 3:          0   Feb                4       1      9           3
## 4:          0   Feb                3       2      2           4
## 5:          0   Feb                3       3      1           4
##          VisitorType Weekend Revenue
## 1: Returning_Visitor   FALSE   FALSE
## 2: Returning_Visitor   FALSE   FALSE
## 3: Returning_Visitor   FALSE   FALSE
## 4: Returning_Visitor   FALSE   FALSE
## 5: Returning_Visitor    TRUE   FALSE
```

```
# bottom dataset preview
tail(shopping, 5)
```

```
##    Administrative Administrative_Duration Informational Informational_Duration
## 1:              3                     145             0                      0
## 2:              0                       0             0                      0
## 3:              0                       0             0                      0
## 4:              4                      75             0                      0
## 5:              0                       0             0                      0
##    ProductRelated ProductRelated_Duration BounceRates  ExitRates PageValues
## 1:             53                1783.792 0.007142857 0.02903061   12.24172
## 2:              5                 465.750 0.000000000 0.02133333    0.00000
## 3:              6                 184.250 0.083333333 0.08666667    0.00000
## 4:             15                 346.000 0.000000000 0.02105263    0.00000
## 5:              3                  21.250 0.000000000 0.06666667    0.00000
##    SpecialDay Month OperatingSystems Browser Region TrafficType
## 1:          0   Dec                4       6      1           1
## 2:          0   Nov                3       2      1           8
## 3:          0   Nov                3       2      1          13
## 4:          0   Nov                2       2      3          11
## 5:          0   Nov                3       2      1           2
##          VisitorType Weekend Revenue
## 1: Returning_Visitor    TRUE   FALSE
```

```
## 2: Returning_Visitor    TRUE   FALSE
## 3: Returning_Visitor    TRUE   FALSE
## 4: Returning_Visitor   FALSE   FALSE
## 5:       New_Visitor    TRUE   FALSE
```

```
# let's see the structure of the dataset
str(shopping)
```

c) Checking Datatypes

```
## Classes 'data.table' and 'data.frame':   12330 obs. of  18 variables:
##  $ Administrative         : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Administrative_Duration: num  0 0 -1 0 0 0 -1 -1 0 0 ...
##  $ Informational          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Informational_Duration : num  0 0 -1 0 0 0 -1 -1 0 0 ...
##  $ ProductRelated         : int  1 2 1 2 10 19 1 1 2 3 ...
##  $ ProductRelated_Duration: num  0 64 -1 2.67 627.5 ...
##  $ BounceRates            : num  0.2 0 0.2 0.05 0.02 ...
##  $ ExitRates              : num  0.2 0.1 0.2 0.14 0.05 ...
##  $ PageValues             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ SpecialDay             : num  0 0 0 0 0 0 0.4 0 0.8 0.4 ...
##  $ Month                  : chr  "Feb" "Feb" "Feb" "Feb" ...
##  $ OperatingSystems       : int  1 2 4 3 3 2 2 1 2 2 ...
##  $ Browser                : int  1 2 1 2 3 2 4 2 2 4 ...
##  $ Region                 : int  1 1 9 2 1 1 3 1 2 1 ...
##  $ TrafficType            : int  1 2 3 4 4 3 3 5 3 2 ...
##  $ VisitorType            : chr  "Returning_Visitor" "Returning_Visitor" "Returning_Visitor" "Return
##  $ Weekend                : logi  FALSE FALSE FALSE FALSE TRUE FALSE ...
##  $ Revenue                : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

The columns have the right datatypes

3. External Data Validation

The data was provided by the company about the brand and was based on a previous related customer reaction data, there is no reason for external validation

4. Data Preparation

a) Validation

```
colnames(shopping)
```

Checking to ensure the columns are valid for this analysis

```
##  [1] "Administrative"          "Administrative_Duration"
##  [3] "Informational"           "Informational_Duration"
##  [5] "ProductRelated"          "ProductRelated_Duration"
##  [7] "BounceRates"             "ExitRates"
##  [9] "PageValues"              "SpecialDay"
## [11] "Month"                   "OperatingSystems"
## [13] "Browser"                 "Region"
## [15] "TrafficType"             "VisitorType"
## [17] "Weekend"                 "Revenue"
```

The columns are valid

```
# let's see the dataset summary
summary(shopping)
```

let's check for invalid values/ Anomalies

```
##  Administrative   Administrative_Duration Informational
##  Min.   : 0.000   Min.   :  -1.00         Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.:   0.00         1st Qu.: 0.000
##  Median : 1.000   Median :   8.00         Median : 0.000
##  Mean   : 2.318   Mean   :  80.91         Mean   : 0.504
##  3rd Qu.: 4.000   3rd Qu.:  93.50         3rd Qu.: 0.000
##  Max.   :27.000   Max.   :3398.75         Max.   :24.000
##  NA's   :14       NA's   :14              NA's   :14
##  Informational_Duration ProductRelated   ProductRelated_Duration
##  Min.   :  -1.00        Min.   :  0.00   Min.   :   -1.0
##  1st Qu.:   0.00        1st Qu.:  7.00   1st Qu.:  185.0
##  Median :   0.00        Median : 18.00   Median :  599.8
##  Mean   :  34.51        Mean   : 31.76   Mean   : 1196.0
##  3rd Qu.:   0.00        3rd Qu.: 38.00   3rd Qu.: 1466.5
##  Max.   :2549.38        Max.   :705.00   Max.   :63973.5
##  NA's   :14             NA's   :14       NA's   :14
##   BounceRates         ExitRates         PageValues        SpecialDay
##  Min.   :0.000000   Min.   :0.00000   Min.   :  0.000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.01429   1st Qu.:  0.000   1st Qu.:0.00000
##  Median :0.003119   Median :0.02512   Median :  0.000   Median :0.00000
##  Mean   :0.022152   Mean   :0.04300   Mean   :  5.889   Mean   :0.06143
##  3rd Qu.:0.016684   3rd Qu.:0.05000   3rd Qu.:  0.000   3rd Qu.:0.00000
##  Max.   :0.200000   Max.   :0.20000   Max.   :361.764   Max.   :1.00000
##  NA's   :14         NA's   :14
##     Month           OperatingSystems    Browser           Region
##  Length:12330       Min.   :1.000    Min.   : 1.000   Min.   :1.000
##  Class :character   1st Qu.:2.000    1st Qu.: 2.000   1st Qu.:1.000
##  Mode  :character   Median :2.000    Median : 2.000   Median :3.000
##                     Mean   :2.124    Mean   : 2.357   Mean   :3.147
##                     3rd Qu.:3.000    3rd Qu.: 2.000   3rd Qu.:4.000
##                     Max.   :8.000    Max.   :13.000   Max.   :9.000
##
##   TrafficType    VisitorType         Weekend         Revenue
##  Min.   : 1.00   Length:12330       Mode :logical   Mode :logical
```

```
##  1st Qu.: 2.00   Class :character   FALSE:9462      FALSE:10422
##  Median : 2.00   Mode  :character   TRUE :2868      TRUE :1908
##  Mean   : 4.07
##  3rd Qu.: 4.00
##  Max.   :20.00
##
```

There are no anomalies, but we have some missing values

b) Consistency

```
# checking for missing values
colSums(is.na(shopping))
```

```
##          Administrative Administrative_Duration           Informational
##                      14                      14                      14
##  Informational_Duration          ProductRelated ProductRelated_Duration
##                      14                      14                      14
##             BounceRates               ExitRates              PageValues
##                      14                      14                       0
##              SpecialDay                   Month        OperatingSystems
##                       0                       0                       0
##                 Browser                  Region             TrafficType
##                       0                       0                       0
##             VisitorType                 Weekend                 Revenue
##                       0                       0                       0
```

We have some missing values which we will now impute

```
shopping <- na.omit(shopping)
```

We removed all the missing values

d) Completeness

```
# lets check for duplicates
sum(duplicated(shopping))
```

```
## [1] 117
```

```
# removing duplicates
shop <- distinct(shopping)
```

```
# checking for duplicates
sum(duplicated(shop))
```

```
## [1] 0
```

The dataset had some duplicated values which we removed

d) Uniformity

```
# let's check the uniformity of the column names
colnames(shop)
```

```
##  [1] "Administrative"          "Administrative_Duration"
##  [3] "Informational"           "Informational_Duration"
##  [5] "ProductRelated"          "ProductRelated_Duration"
##  [7] "BounceRates"             "ExitRates"
##  [9] "PageValues"              "SpecialDay"
## [11] "Month"                   "OperatingSystems"
## [13] "Browser"                 "Region"
## [15] "TrafficType"             "VisitorType"
## [17] "Weekend"                 "Revenue"
```

The column names are uniform and have no white spaces

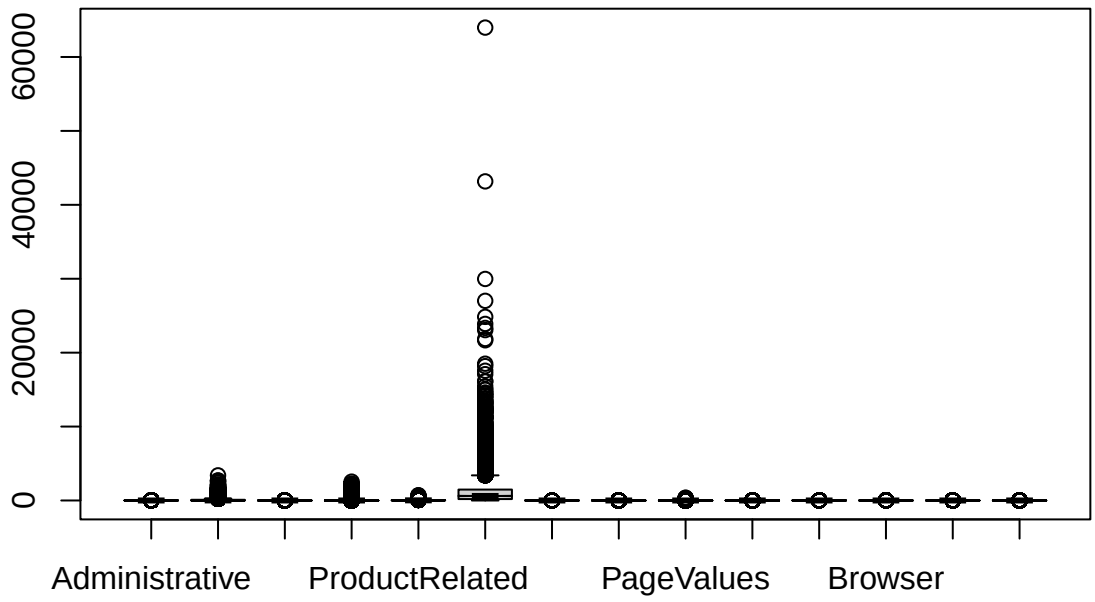
```
num_col <- Filter(is.numeric, shop)
desc_stats <- data.frame(
    min = apply(num_col, 2, min),
    median = apply(num_col, 2, median),
    mean_df = apply(num_col, 2, mean),
    SD = apply(num_col, 2, sd),
    max = apply(num_col, 2, max),
    skew = apply(num_col,2, skewness),
    Kurt = apply(num_col,2, kurtosis)
)

statistics <- round(desc_stats, 1)
statistics
```

boxplots for numerical columns to remove outliers

```
##                           min median mean_df     SD     max skew  Kurt
## Administrative              0    1.0     2.3    3.3    27.0  1.9   4.6
## Administrative_Duration    -1    9.0    81.7  177.5  3398.8  5.6  50.1
## Informational               0    0.0     0.5    1.3    24.0  4.0  26.6
## Informational_Duration     -1    0.0    34.8  141.5  2549.4  7.5  75.5
## ProductRelated              0   18.0    32.1   44.6   705.0  4.3  31.0
## ProductRelated_Duration    -1  609.5  1207.5 1919.9 63973.5  7.3 136.6
## BounceRates                 0    0.0     0.0    0.0     0.2  3.2   9.3
## ExitRates                   0    0.0     0.0    0.0     0.2  2.2   4.6
## PageValues                  0    0.0     6.0   18.7   361.8  6.3  64.9
## SpecialDay                  0    0.0     0.1    0.2     1.0  3.3   9.8
## OperatingSystems            1    2.0     2.1    0.9     8.0  2.0  10.3
## Browser                     1    2.0     2.4    1.7    13.0  3.2  12.5
## Region                      1    3.0     3.2    2.4     9.0  1.0  -0.2
## TrafficType                 1    2.0     4.1    4.0    20.0  2.0   3.5
```

```
boxplot(num_col, notch = TRUE)
```

```
## Warning in (function (z, notch = FALSE, width = NULL, varwidth = FALSE, : some
## notches went outside hinges ('box'): maybe set notch=FALSE
```



The Administrative, BounceRates and Browser columns have outliers but they will be retained for further analysis

5. Descriptive Analysis

```
# descriptive statistics
describe(shop)
```

```
## shop
##
##  18  Variables      12199  Observations
## ---------------------------------------------------------------------------
## Administrative
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0       27    0.898     2.34    3.193        0        0
##      .25      .50      .75      .90      .95
##        0        1        4        7        9
##
```

```
## lowest :  0  1  2  3  4, highest: 22 23 24 26 27
## ---------------------------------------------------------------------------
## Administrative_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0     3336    0.896    81.68    125.8     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     9.00    94.75   227.05   352.23
##
## lowest :   -1.000000    0.000000    1.333333    2.000000    3.000000
## highest: 2407.423810 2629.253968 2657.318056 2720.500000 3398.750000
## ---------------------------------------------------------------------------
## Informational
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0       17    0.516   0.5088   0.8763        0        0
##      .25      .50      .75      .90      .95
##        0        0        0        2        3
##
## lowest :  0  1  2  3  4, highest: 12 13 14 16 24
##
## Value          0     1     2     3     4     5     6     7     8     9    10
## Frequency   9569  1041   727   380   222    99    78    36    14    15     7
## Proportion 0.784 0.085 0.060 0.031 0.018 0.008 0.006 0.003 0.001 0.001 0.001
##
## Value         11    12    13    14    16    24
## Frequency      1     5     1     2     1     1
## Proportion 0.000 0.000 0.000 0.000 0.000 0.000
## ---------------------------------------------------------------------------
## Informational_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0     1259    0.488    34.84    64.71        0        0
##      .25      .50      .75      .90      .95
##        0        0        0       74      199
##
## lowest :   -1.000    0.000    1.000    1.500    2.000
## highest: 2166.500 2195.300 2252.033 2256.917 2549.375
## ---------------------------------------------------------------------------
## ProductRelated
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0      311    0.999    32.06    36.51        2        3
##      .25      .50      .75      .90      .95
##        8       18       38       74      110
##
## lowest :   0   1   2   3   4, highest: 518 534 584 686 705
## ---------------------------------------------------------------------------
## ProductRelated_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0     9552        1     1208     1491      0.0     44.4
##      .25      .50      .75      .90      .95
##    193.6    609.5   1477.6   2905.0   4313.5
##
## lowest :    -1.000000     0.000000     0.500000     1.000000     2.333333
## highest: 24844.156200 27009.859430 29970.465970 43171.233380 63973.522230
## ---------------------------------------------------------------------------
## BounceRates
```

```
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0     1872    0.908  0.02045  0.03289  0.00000  0.00000
##      .25      .50      .75      .90      .95
##  0.00000  0.00293  0.01667  0.05000  0.15000
##
## lowest : 0.0000000 0.0000273 0.0000335 0.0000383 0.0000394
## highest: 0.1750000 0.1769231 0.1800000 0.1833333 0.2000000
## ------------------------------------------------------------------------------
## ExitRates
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0     4777        1   0.0415  0.04232 0.004545 0.007407
##      .25      .50      .75      .90      .95
## 0.014223 0.025000 0.048485 0.100000 0.175000
##
## lowest : 0.000000000 0.000175593 0.000250438 0.000262123 0.000263158
## highest: 0.183333333 0.186666667 0.188888889 0.192307692 0.200000000
## ------------------------------------------------------------------------------
## PageValues
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0     2704    0.532    5.953    10.65     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00     0.00    19.12    38.31
##
## lowest :   0.00000000   0.03803454   0.06704955   0.09354695   0.09862140
## highest: 261.49128570 270.78469310 287.95379280 360.95338390 361.76374190
## ------------------------------------------------------------------------------
## SpecialDay
##        n  missing distinct     Info     Mean      Gmd
##    12199        0        6    0.277  0.06197   0.1142
##
## lowest : 0.0 0.2 0.4 0.6 0.8, highest: 0.2 0.4 0.6 0.8 1.0
##
## Value        0.0   0.2   0.4   0.6   0.8   1.0
## Frequency  10950   178   243   350   324   154
## Proportion 0.898 0.015 0.020 0.029 0.027 0.013
## ------------------------------------------------------------------------------
## Month
##        n  missing distinct
##    12199        0       10
##
## lowest : Aug  Dec  Feb  Jul  June, highest: Mar  May  Nov  Oct  Sep
##
## Value        Aug   Dec   Feb   Jul  June   Mar   May   Nov   Oct   Sep
## Frequency    433  1706   182   432   285  1853  3328  2983   549   448
## Proportion 0.035 0.140 0.015 0.035 0.023 0.152 0.273 0.245 0.045 0.037
## ------------------------------------------------------------------------------
## OperatingSystems
##        n  missing distinct     Info     Mean      Gmd
##    12199        0        8    0.828    2.124   0.8615
##
## lowest : 1 2 3 4 5, highest: 4 5 6 7 8
##
## Value          1     2     3     4     5     6     7     8
## Frequency   2548  6536  2530   478     6    19     7    75
```

```
## Proportion 0.209 0.536 0.207 0.039 0.000 0.002 0.001 0.006
## ---------------------------------------------------------------------------
## Browser
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0       13    0.723    2.358    1.316        1        1
##      .25      .50      .75      .90      .95
##        2        2        2        4        5
##
## lowest :  1  2  3  4  5, highest:  9 10 11 12 13
##
## Value          1     2     3     4     5     6     7     8     9    10    11
## Frequency   2426  7878   105   730   466   174    49   135     1   163     6
## Proportion 0.199 0.646 0.009 0.060 0.038 0.014 0.004 0.011 0.000 0.013 0.000
##
## Value         12    13
## Frequency     10    56
## Proportion 0.001 0.005
## ---------------------------------------------------------------------------
## Region
##        n  missing distinct     Info     Mean      Gmd
##    12199        0        9    0.933    3.153    2.553
##
## lowest : 1 2 3 4 5, highest: 5 6 7 8 9
##
## Value          1     2     3     4     5     6     7     8     9
## Frequency   4711  1127  2382  1168   317   800   758   431   505
## Proportion 0.386 0.092 0.195 0.096 0.026 0.066 0.062 0.035 0.041
## ---------------------------------------------------------------------------
## TrafficType
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    12199        0       20    0.954    4.075    3.735        1        1
##      .25      .50      .75      .90      .95
##        2        2        4       11       13
##
## lowest :  1  2  3  4  5, highest: 16 17 18 19 20
##
## Value          1     2     3     4     5     6     7     8     9    10    11
## Frequency   2383  3907  2017  1066   260   443    40   343    41   450   247
## Proportion 0.195 0.320 0.165 0.087 0.021 0.036 0.003 0.028 0.003 0.037 0.020
##
## Value         12    13    14    15    16    17    18    19    20
## Frequency      1   728    13    36     3     1    10    17   193
## Proportion 0.000 0.060 0.001 0.003 0.000 0.000 0.001 0.001 0.016
## ---------------------------------------------------------------------------
## VisitorType
##        n  missing distinct
##    12199        0        3
##
## Value            New_Visitor             Other Returning_Visitor
## Frequency               1693                81             10425
## Proportion             0.139             0.007             0.855
## ---------------------------------------------------------------------------
## Weekend
##        n  missing distinct
```

```
##    12199        0        2
##
## Value      FALSE  TRUE
## Frequency   9343  2856
## Proportion 0.766 0.234
## ---------------------------------------------------------------------------
## Revenue
##        n  missing distinct
##    12199        0        2
##
## Value      FALSE  TRUE
## Frequency  10291  1908
## Proportion 0.844 0.156
## ---------------------------------------------------------------------------
```

Month The website had the most visitors during the months of May, with 27.3% of total activity for the entire period and november which had 24.5% of all viisits.

Browser 53.6% of all visitors were using operating system type 2 while 20.9% were using type 1

Region 38.6% of all visitors were from region 1 followed by region 3 which had 19.5% of all visitors

Visitor Type 193.9% of the visitors were new visitors while 85.5% were return visitors.

Weekend 23.4% of site visits were during the weekend while the rest 76.6% were during the weekdays.

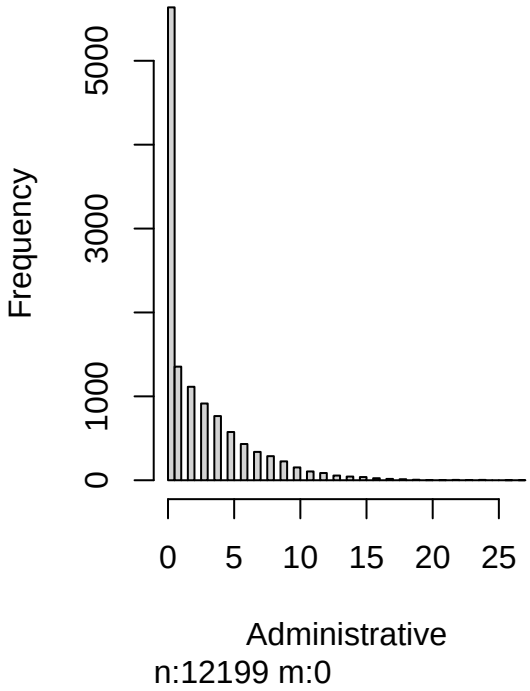
Traffic Type Most of the visitors were directed from traffic type 2 which brought in 32% of all visitors followed by type 1 which brought in 19.5% of all visitors and type 2 wuth 16.5%. Trafic type 12,17 and 16 did not bring in any visitors for the enter duration under analysis.

```
num_col <- Filter(is.numeric, shop)
desc_stats <- data.frame(
    min = apply(num_col, 2, min),
    median = apply(num_col, 2, median),
    mean_df = apply(num_col, 2, mean),
    SD = apply(num_col, 2, sd),
    max = apply(num_col, 2, max),
    skew = apply(num_col,2, skewness),
    Kurt = apply(num_col,2, kurtosis)
)

statistics <- round(desc_stats, 1)
statistics
```

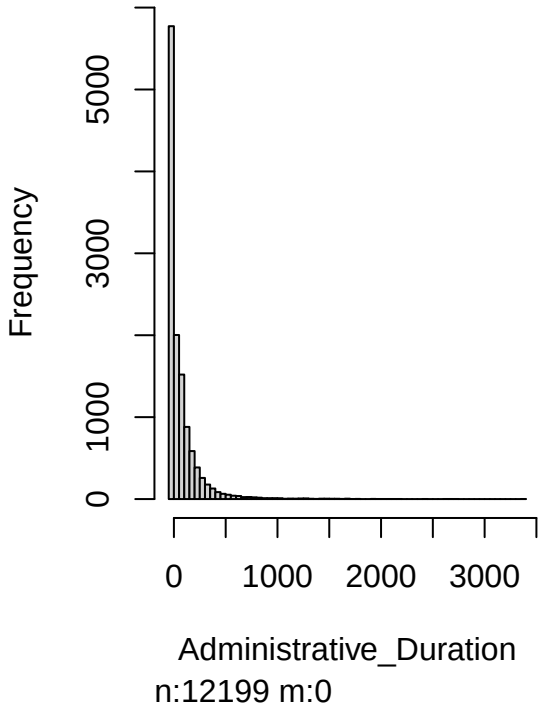
let's make a dataframe of numeric variables for descriptive statistics

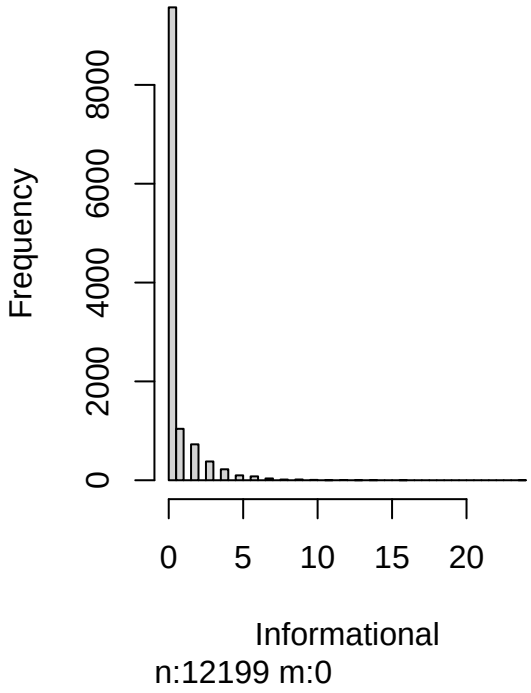
```
##                           min median mean_df     SD     max skew  Kurt
## Administrative              0    1.0     2.3    3.3    27.0  1.9   4.6
## Administrative_Duration    -1    9.0    81.7  177.5  3398.8  5.6  50.1
## Informational               0    0.0     0.5    1.3    24.0  4.0  26.6
## Informational_Duration     -1    0.0    34.8  141.5  2549.4  7.5  75.5
## ProductRelated              0   18.0    32.1   44.6   705.0  4.3  31.0
## ProductRelated_Duration    -1  609.5  1207.5 1919.9 63973.5  7.3 136.6
## BounceRates                 0    0.0     0.0    0.0     0.2  3.2   9.3
## ExitRates                   0    0.0     0.0    0.0     0.2  2.2   4.6
## PageValues                  0    0.0     6.0   18.7   361.8  6.3  64.9
## SpecialDay                  0    0.0     0.1    0.2     1.0  3.3   9.8
## OperatingSystems            1    2.0     2.1    0.9     8.0  2.0  10.3
## Browser                     1    2.0     2.4    1.7    13.0  3.2  12.5
## Region                      1    3.0     3.2    2.4     9.0  1.0  -0.2
## TrafficType                 1    2.0     4.1    4.0    20.0  2.0   3.5
```

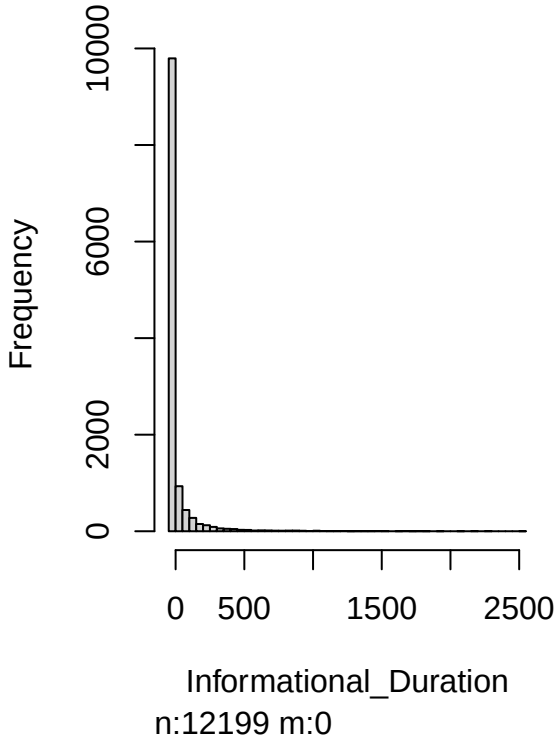
```r
par(mfrow = c(1,2))
for (i in 1:13) {
    hist(num_col[, ..i ], main = names(num_col)[i], xlab = NULL)
}
```

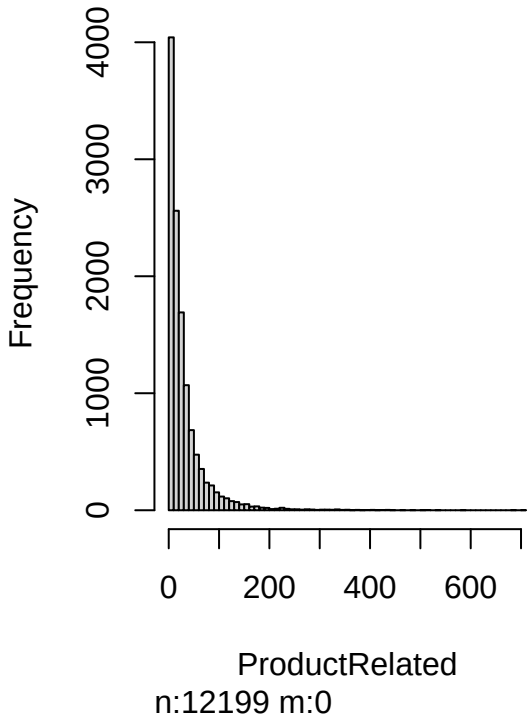


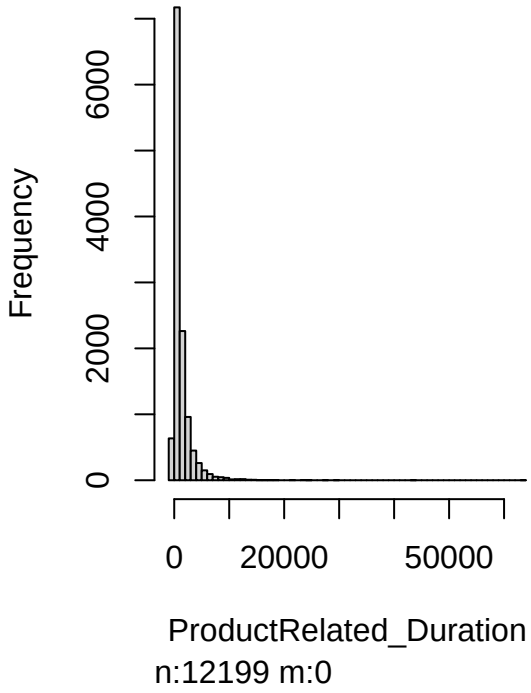
let's plot data distribution histograms for the numerical columns

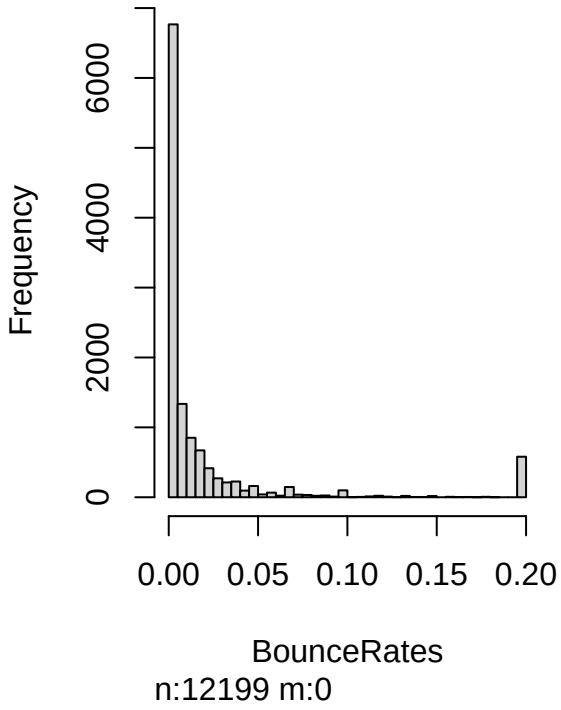


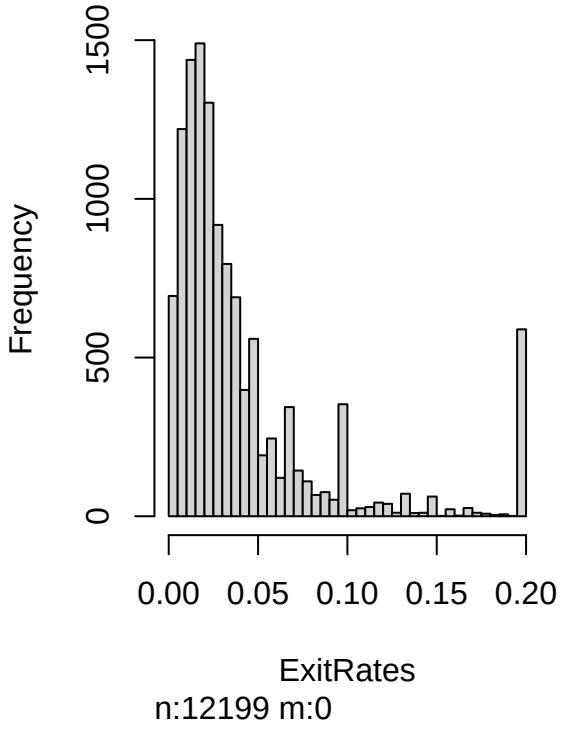

Frequency
0
2000
4000
6000
8000
0
5
10
15
20
Informational
n:12199 m:0

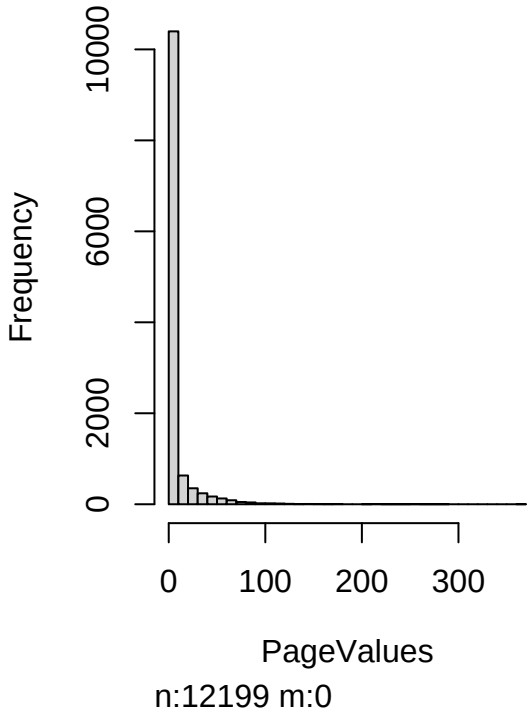

Frequency
0
2000
6000
10000
0
500
1500
2500
Informational_Duration
n:12199 m:0

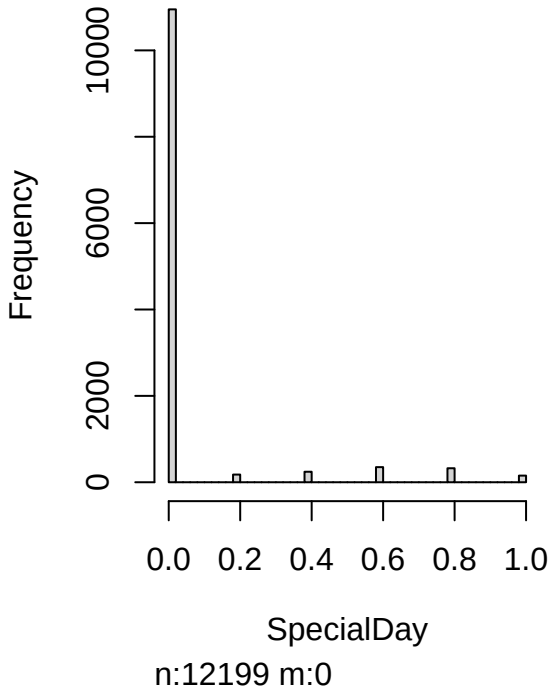

Frequency
0
1000
2000
3000
4000
0
200
400
600
ProductRelated
n:12199 m:0

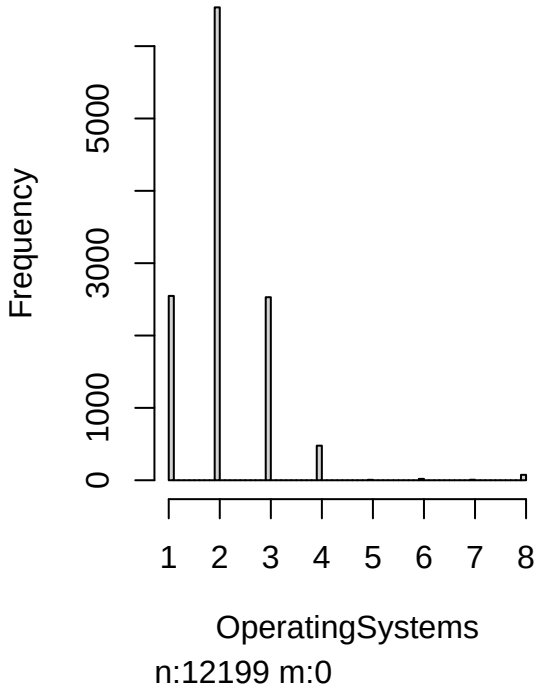

Frequency
0
2000
4000
6000
0
20000
50000
ProductRelated_Duration
n:12199 m:0

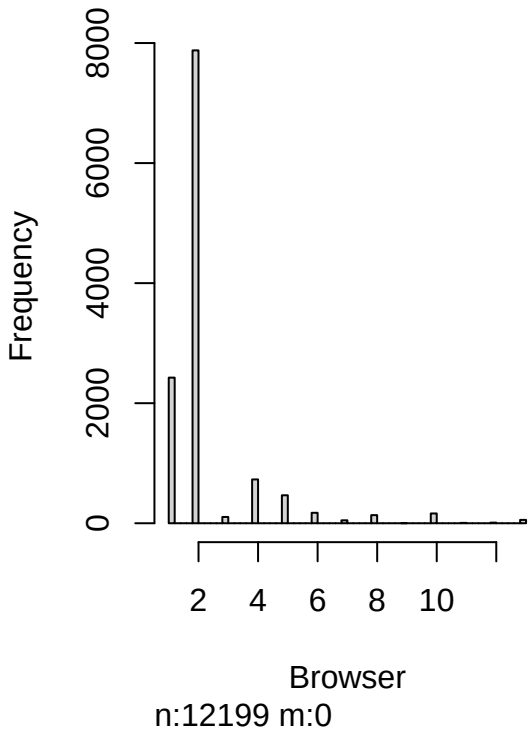

Frequency
0
2000
4000
6000
0.00
0.05
0.10
0.15
0.20
BounceRates
n:12199 m:0

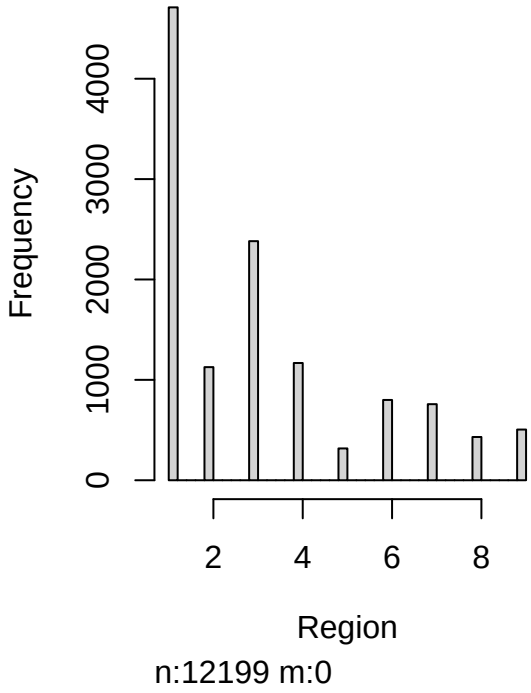

Frequency
0
500
1000
1500
0.00
0.05
0.10
0.15
0.20
ExitRates
n:12199 m:0


Frequency
0
2000
6000
10000
0
100
200
300
PageValues
n:12199 m:0


Frequency
0
2000
6000
10000
0.0
0.2
0.4
0.6
0.8
1.0
SpecialDay
n:12199 m:0

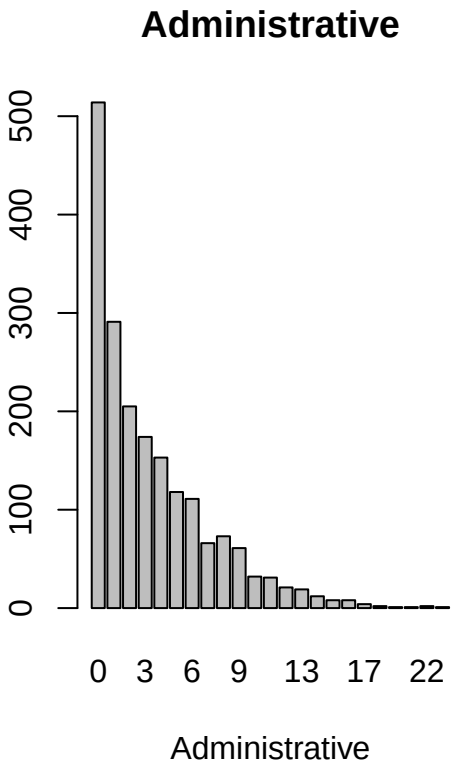

Frequency
0
1000
3000
5000
1
2
3
4
5
6
7
8
OperatingSystems
n:12199 m:0

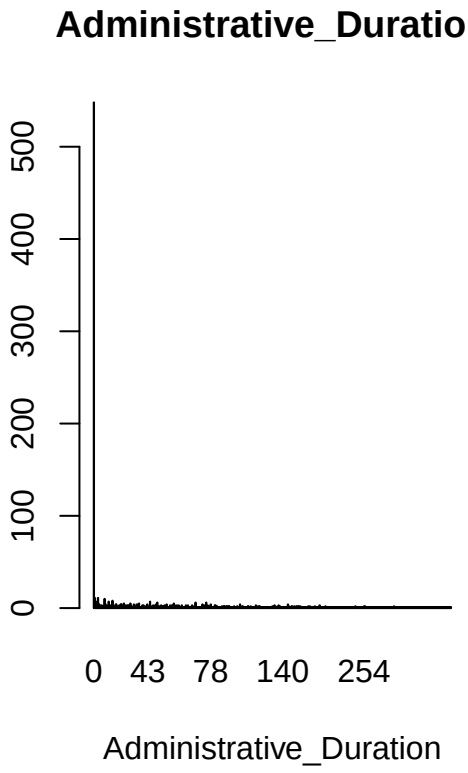

Frequency
0
2000
4000
6000
8000
2
4
6
8
10
Browser
n:12199 m:0

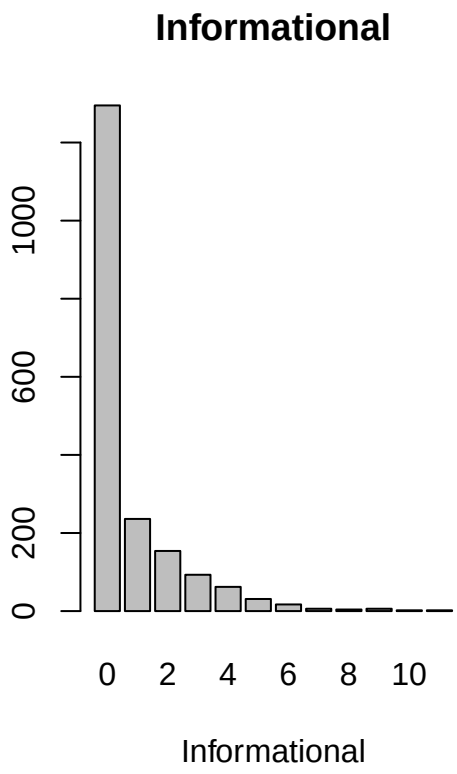


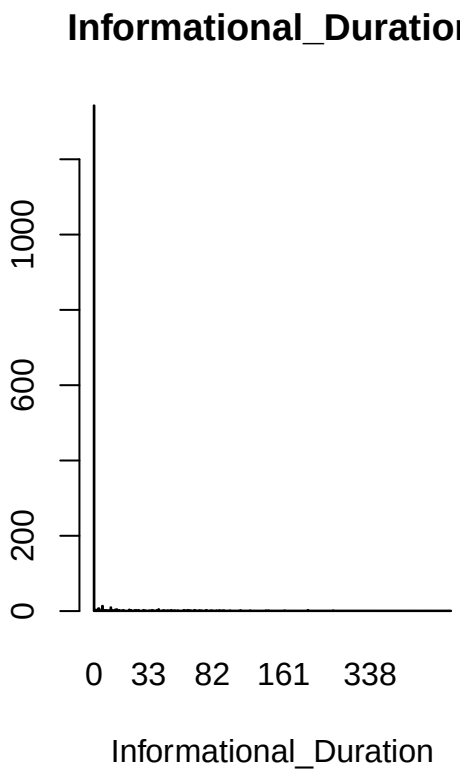
Analysis on The Revenue Target Column

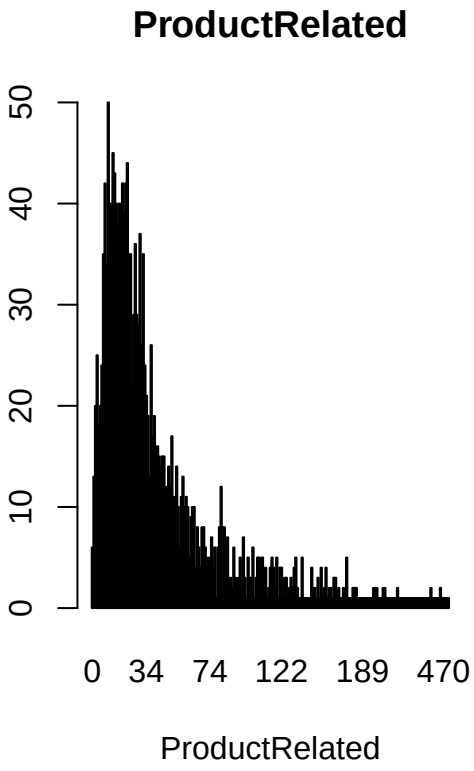
```
# let's select the revenue column
Revenue <- shop[shop$Revenue == TRUE]

# let's have plots that show customer behaviour
par(mfrow = c(1,2))
for (i in 1:18) {
    barplot(table(Revenue[, ..i]),main = names(Revenue)[i], xlab = names(Revenue)[i])
}
```

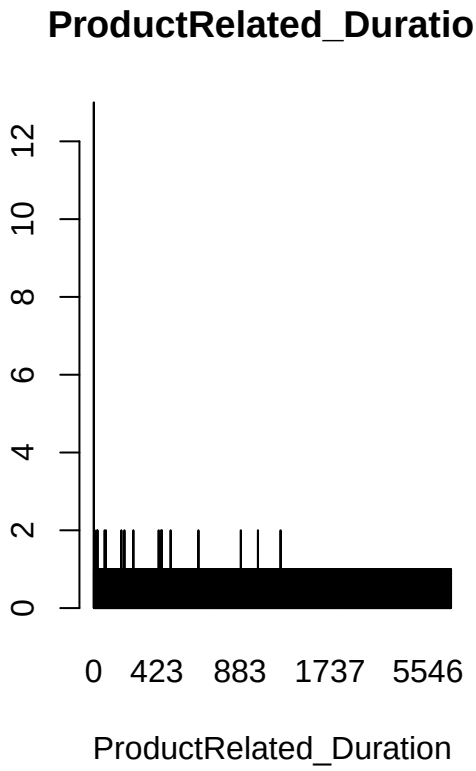

Administrative
0
100
200
300
400
500
0 3 6 9 13 17 22
Administrative

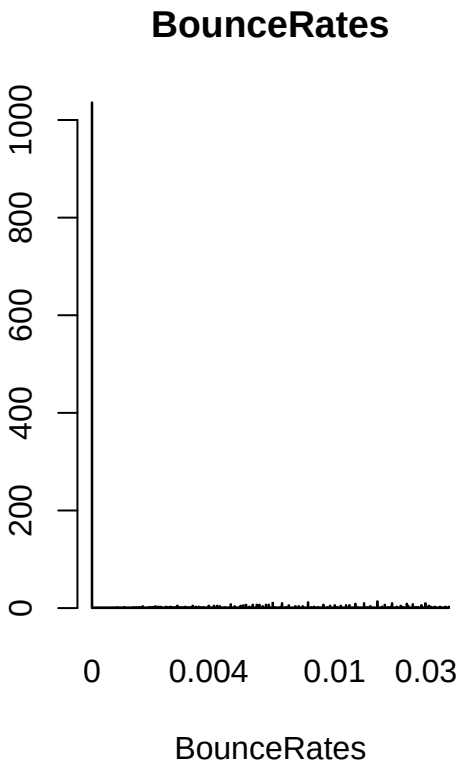

Administrative_Duration
0
100
200
300
400
500
0 43 78 140 254
Administrative_Duration

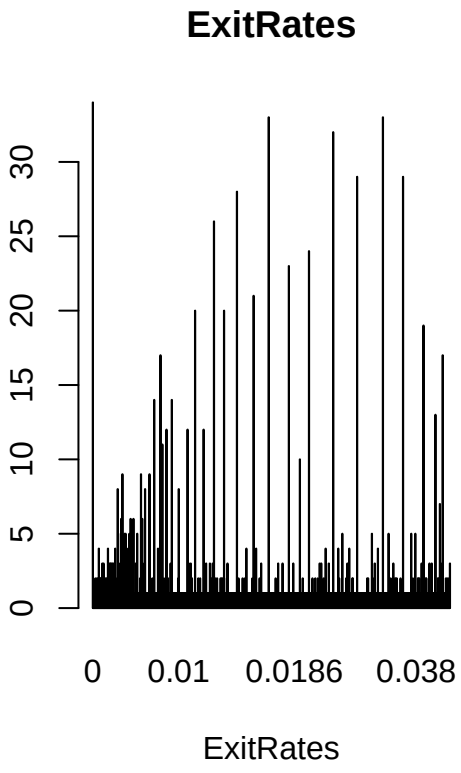

Informational
0
200
600
1000
0
2
4
6
8
10
Informational

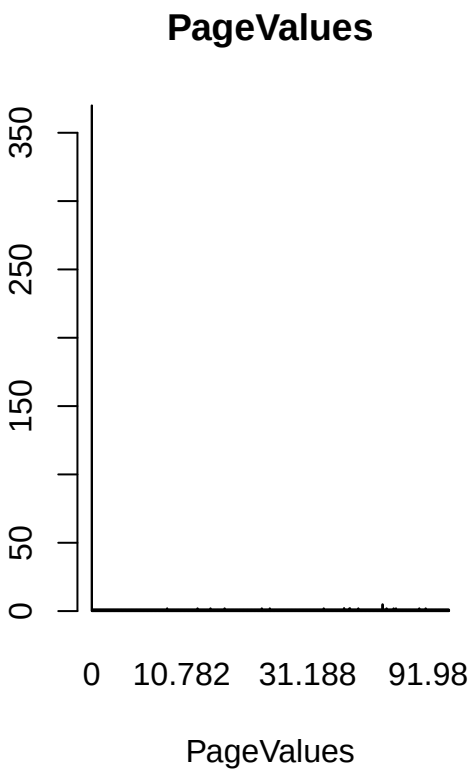

Informational_Duration
0
200
600
1000
0
33
82
161
338
Informational_Duration

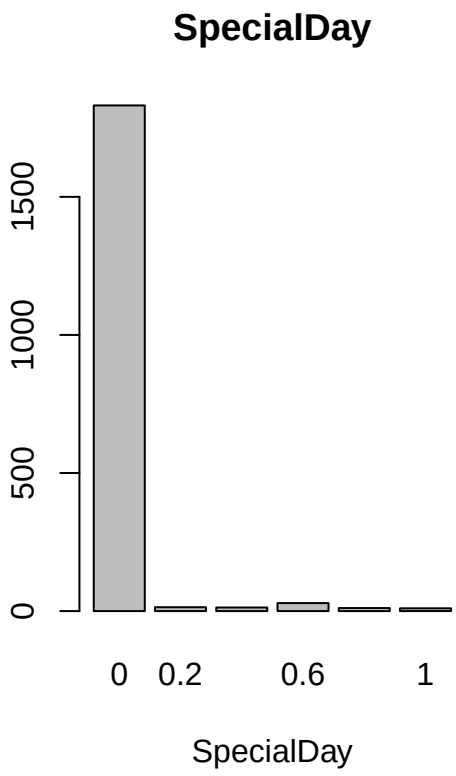

ProductRelated
0
10
20
30
40
50
0
34
74
122
189
470
ProductRelated

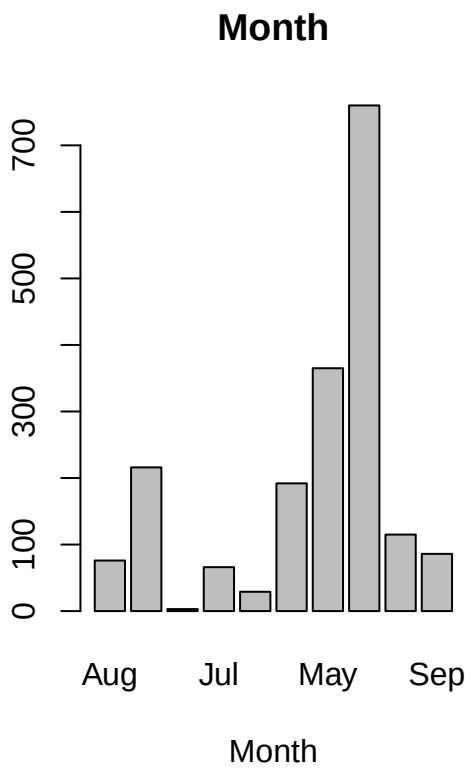

ProductRelated_Duration
0
2
4
6
8
10
12
0
423
883
1737
5546
ProductRelated_Duration

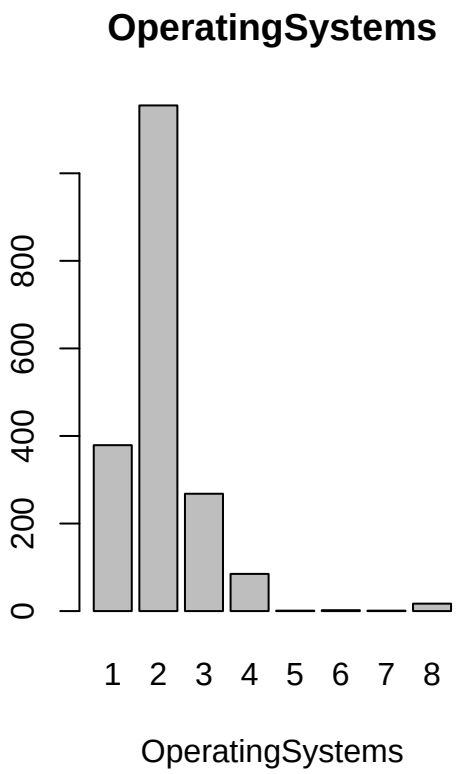

BounceRates
0
200
400
600
800
1000
0
0.004
0.01
0.03
BounceRates

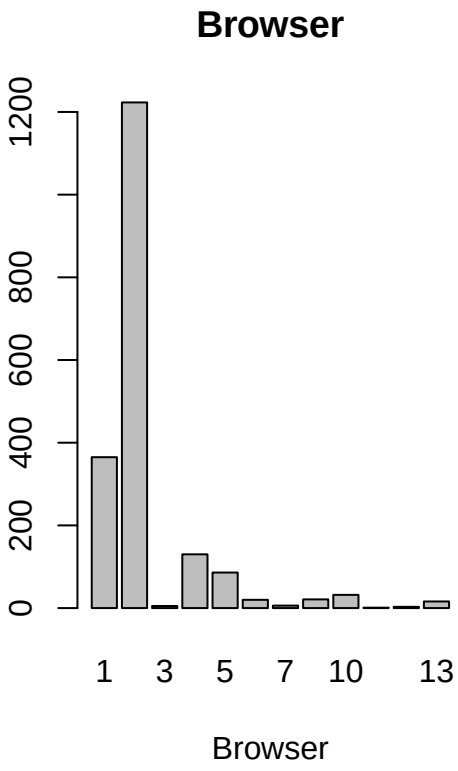

ExitRates
0
5
10
15
20
25
30
0
0.01
0.0186
0.038
ExitRates

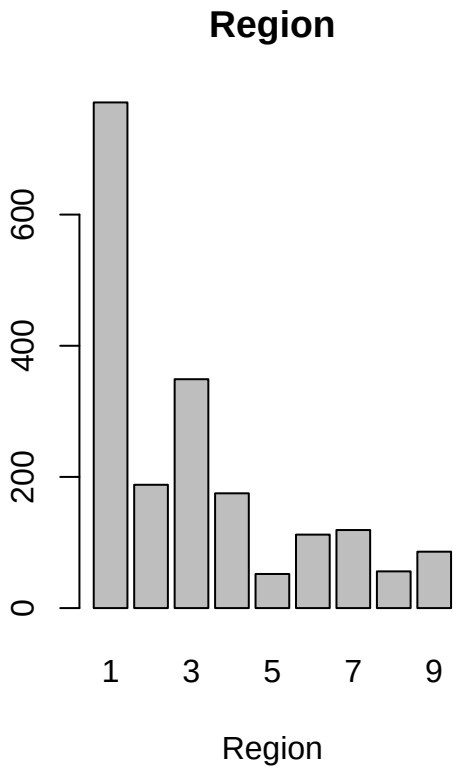

PageValues
0
50
150
250
350
0
10.782
31.188
91.98
PageValues

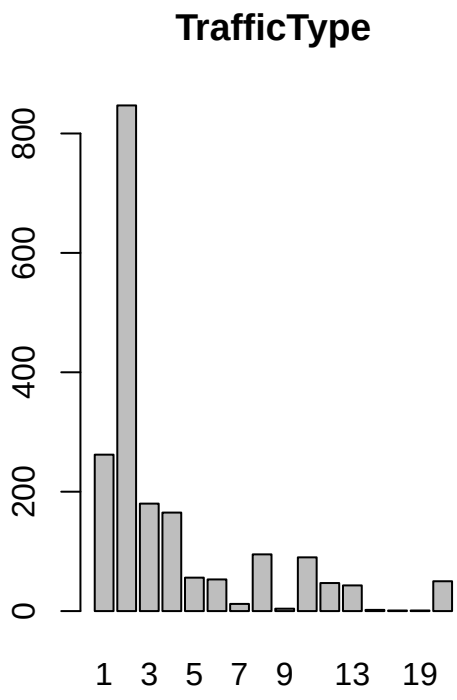

SpecialDay
0
500
1000
1500
0
0.2
0.6
1
SpecialDay

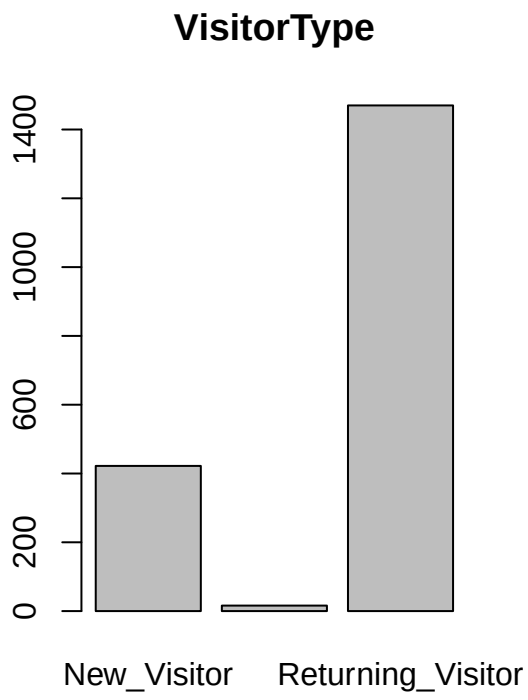

Month
0
100
300
500
700
Aug
Jul
May
Sep
Month

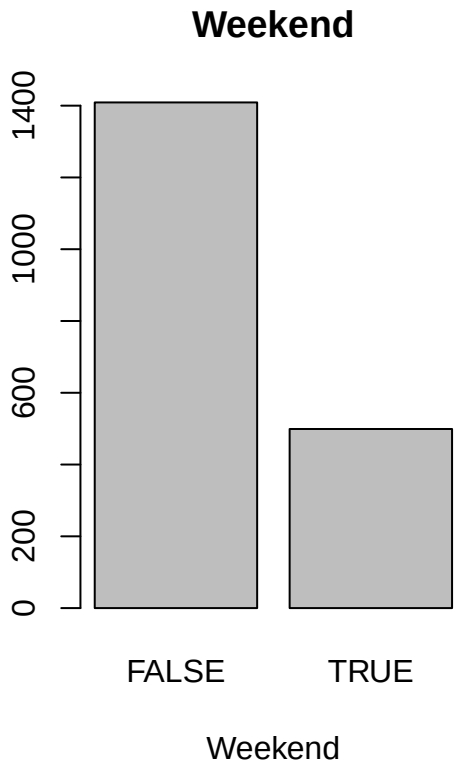

OperatingSystems
0
200
400
600
800
1
2
3
4
5
6
7
8
OperatingSystems

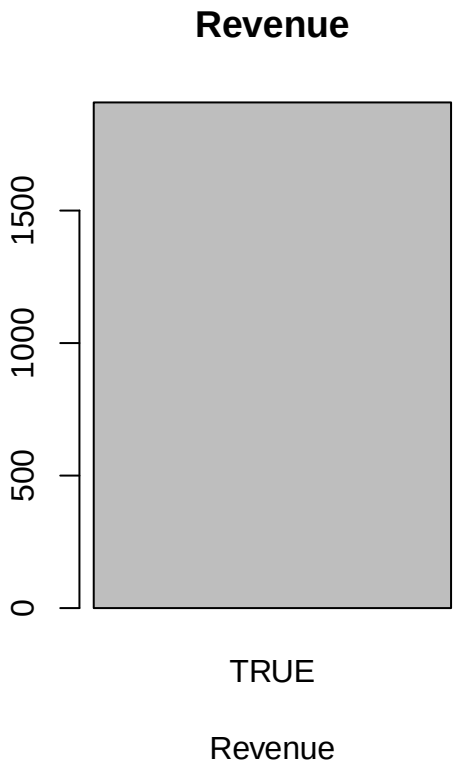

Browser
0
200
400
600
800
1200
1
3
5
7
10
13
Browser


Region
0
200
400
600
1
3
5
7
9
Region


TrafficType
0
200
400
600
800
1
3
5
7
9
13
19
TrafficType


VisitorType
0
200
600
1000
1400
New_Visitor
Returning_Visitor
VisitorType






```
describe(Revenue)
```

Descriptive Statistics for the Revenue column

```
## Revenue
##
##  18  Variables      1908  Observations
## ---------------------------------------------------------------------------
## Administrative
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0       23    0.974    3.394    3.877        0        0
##      .25      .50      .75      .90      .95
##        0        2        5        9       11
##
## lowest :  0  1  2  3  4, highest: 18 19 20 22 26
## ---------------------------------------------------------------------------
## Administrative_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0     1037    0.976    119.5    163.2     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00    52.37   151.08   311.40   478.02
##
## lowest :    0.000000    2.000000    3.000000    4.000000    4.333333
```

```
## highest: 1592.916667 1660.300000 1668.500000 2047.234848 2086.750000
## ---------------------------------------------------------------------------
## Informational
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0       12    0.685   0.7862    1.248        0        0
##      .25      .50      .75      .90      .95
##        0        0        1        3        4
##
## lowest :  0  1  2  3  4, highest:  7  8  9 10 12
##
## Value          0     1     2     3     4     5     6     7     8     9    10
## Frequency   1295   236   154    93    62    31    17     6     4     6     2
## Proportion 0.679 0.124 0.081 0.049 0.032 0.016 0.009 0.003 0.002 0.003 0.001
##
## Value         12
## Frequency      2
## Proportion 0.001
## ---------------------------------------------------------------------------
## Informational_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0      426    0.651    57.61    102.1      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0     19.0    173.9    339.0
##
## lowest :    0.000    2.000    3.000    4.000    5.000
## highest: 1488.000 1636.000 1652.000 1665.067 1767.667
## ---------------------------------------------------------------------------
## ProductRelated
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0      225        1    48.21    50.45      5.0      8.0
##      .25      .50      .75      .90      .95
##     15.0     29.0     57.0    113.3    159.0
##
## lowest :   0   1   2   3   4, highest: 439 470 501 517 534
## ---------------------------------------------------------------------------
## ProductRelated_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0     1879        1     1876     2022    121.0    258.4
##      .25      .50      .75      .90      .95
##    541.9   1109.9   2266.0   4532.6   6225.2
##
## lowest :     0.00     5.00     7.00     8.00    10.00
## highest: 14568.16 17550.58 18504.13 21672.24 27009.86
## ---------------------------------------------------------------------------
## BounceRates
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0      536     0.84 0.005117 0.007957 0.000000 0.000000
##      .25      .50      .75      .90      .95
## 0.000000 0.000000 0.006452 0.014286 0.022050
##
## lowest : 0.00000000 0.00003940 0.00007270 0.00007500 0.00008010
## highest: 0.06842105 0.08333333 0.10000000 0.11071429 0.20000000
## ---------------------------------------------------------------------------
## ExitRates
```

```
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0     1136        1  0.01956  0.01552 0.002895 0.004762
##      .25      .50      .75      .90      .95
## 0.009521 0.016000 0.025000 0.037423 0.048140
##
## lowest : 0.000000000 0.000480769 0.000586510 0.000589971 0.000732601
## highest: 0.100000000 0.108333333 0.116666667 0.130952381 0.200000000
## ---------------------------------------------------------------------------
## PageValues
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0     1518    0.993    27.26    32.17    0.000    0.000
##      .25      .50      .75      .90      .95
##    3.641   16.758   38.898   63.606   88.798
##
## lowest :   0.00000000   0.06704955   0.09354695   0.09862140   0.12069991
## highest: 261.49128570 270.78469310 287.95379280 360.95338390 361.76374190
## ---------------------------------------------------------------------------
## SpecialDay
##        n  missing distinct     Info     Mean      Gmd
##     1908        0        6    0.116  0.02317  0.04493
##
## lowest : 0.0 0.2 0.4 0.6 0.8, highest: 0.2 0.4 0.6 0.8 1.0
##
## Value        0.0   0.2   0.4   0.6   0.8   1.0
## Frequency   1831    14    13    29    11    10
## Proportion 0.960 0.007 0.007 0.015 0.006 0.005
## ---------------------------------------------------------------------------
## Month
##        n  missing distinct
##     1908        0       10
##
## lowest : Aug  Dec  Feb  Jul  June, highest: Mar  May  Nov  Oct  Sep
##
## Value        Aug   Dec   Feb   Jul  June   Mar   May   Nov   Oct   Sep
## Frequency     76   216     3    66    29   192   365   760   115    86
## Proportion 0.040 0.113 0.002 0.035 0.015 0.101 0.191 0.398 0.060 0.045
## ---------------------------------------------------------------------------
## OperatingSystems
##        n  missing distinct     Info     Mean      Gmd
##     1908        0        8    0.767    2.093   0.8178
##
## lowest : 1 2 3 4 5, highest: 4 5 6 7 8
##
## Value          1     2     3     4     5     6     7     8
## Frequency    379  1155   268    85     1     2     1    17
## Proportion 0.199 0.605 0.140 0.045 0.001 0.001 0.001 0.009
## ---------------------------------------------------------------------------
## Browser
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0       12    0.729    2.453    1.455        1        1
##      .25      .50      .75      .90      .95
##        2        2        2        4        6
##
## lowest :  1  2  3  4  5, highest:  8 10 11 12 13
```

```
##
## Value          1     2     3     4     5     6     7     8    10    11    12
## Frequency    365  1223     5   130    86    20     6    21    32     1     3
## Proportion 0.191 0.641 0.003 0.068 0.045 0.010 0.003 0.011 0.017 0.001 0.002
##
## Value         13
## Frequency     16
## Proportion 0.008
## ---------------------------------------------------------------------------
## Region
##        n  missing distinct     Info     Mean      Gmd
##     1908        0        9    0.926    3.082    2.535
##
## lowest : 1 2 3 4 5, highest: 5 6 7 8 9
##
## Value          1     2     3     4     5     6     7     8     9
## Frequency    771   188   349   175    52   112   119    56    86
## Proportion 0.404 0.099 0.183 0.092 0.027 0.059 0.062 0.029 0.045
## ---------------------------------------------------------------------------
## TrafficType
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1908        0       16    0.908    4.021    3.575        1        1
##      .25      .50      .75      .90      .95
##        2        2        4       10       13
##
## lowest :  1  2  3  4  5, highest: 13 14 16 19 20
##
## Value          1     2     3     4     5     6     7     8     9    10    11
## Frequency    262   847   180   165    56    53    12    95     4    90    47
## Proportion 0.137 0.444 0.094 0.086 0.029 0.028 0.006 0.050 0.002 0.047 0.025
##
## Value         13    14    16    19    20
## Frequency     43     2     1     1    50
## Proportion 0.023 0.001 0.001 0.001 0.026
## ---------------------------------------------------------------------------
## VisitorType
##        n  missing distinct
##     1908        0        3
##
## Value            New_Visitor             Other Returning_Visitor
## Frequency                422                16              1470
## Proportion             0.221             0.008             0.770
## ---------------------------------------------------------------------------
## Weekend
##        n  missing distinct
##     1908        0        2
##
## Value      FALSE  TRUE
## Frequency   1409   499
## Proportion 0.738 0.262
## ---------------------------------------------------------------------------
## Revenue
##        n  missing distinct    value
##     1908        0        1     TRUE
```

```
##
## Value      TRUE
## Frequency  1908
## Proportion    1
## --------------------------------------------------------------------------------
```

The website had the most revenue during the months of Nov, with 39.8% of total activity for the entire period and november which had 19.1% of all viisits.

60.5% of all revenue were visitors using operating system type 2 while 19.9% were using type 1

Browser 2 had brought in the most revenue with 64.4% while browser 1 brough in 19.1% of the revenue.

40.4% of all revenue were from region 1 followed by region 3 which had 18.3% of all revenue

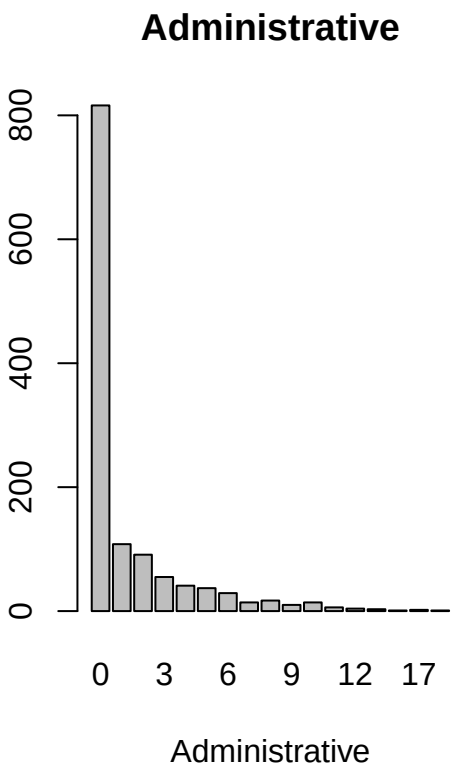
Most of the revenue came from traffic type 2 which brought in 44.4% of all revenue followed by type 1 which brought in 13.7% of all revenue.

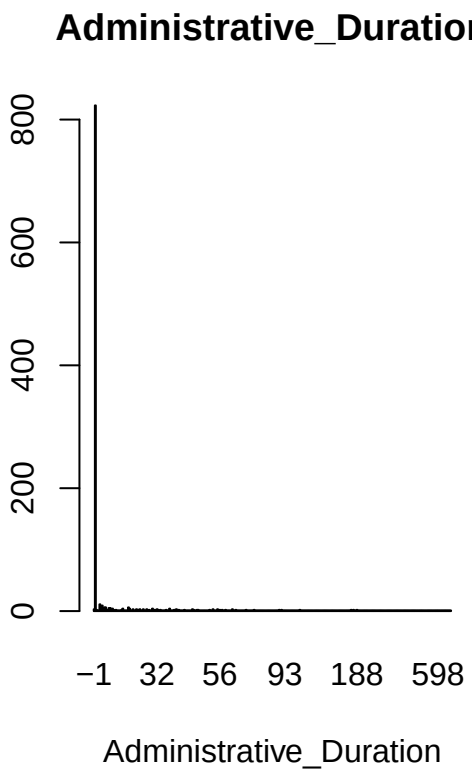
22.1% of the revenue was from new visitors while 77% were return visitors.

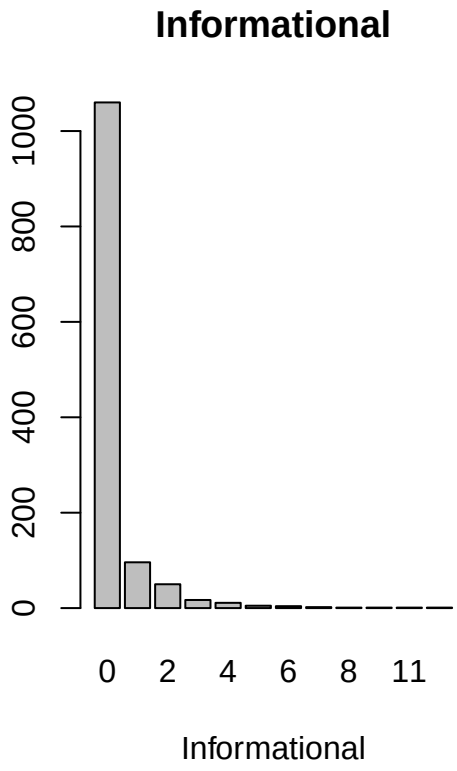
26.2% of site revenue was gained during the weekend while the rest 73.8% was gained during the weekdays.

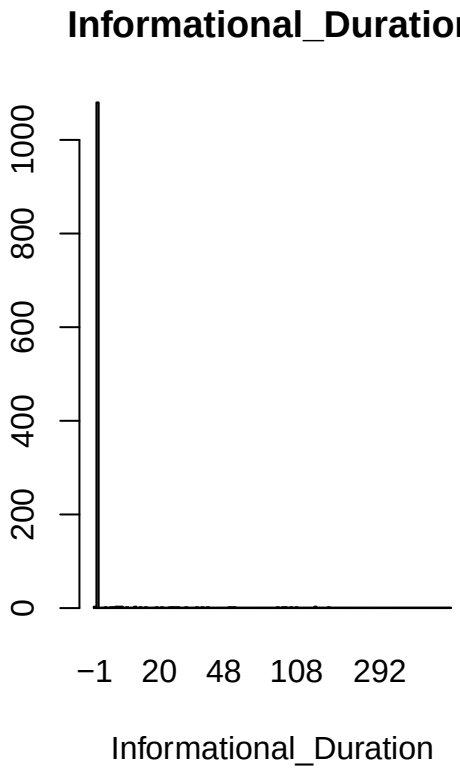
Analysis on the near holiday revenues

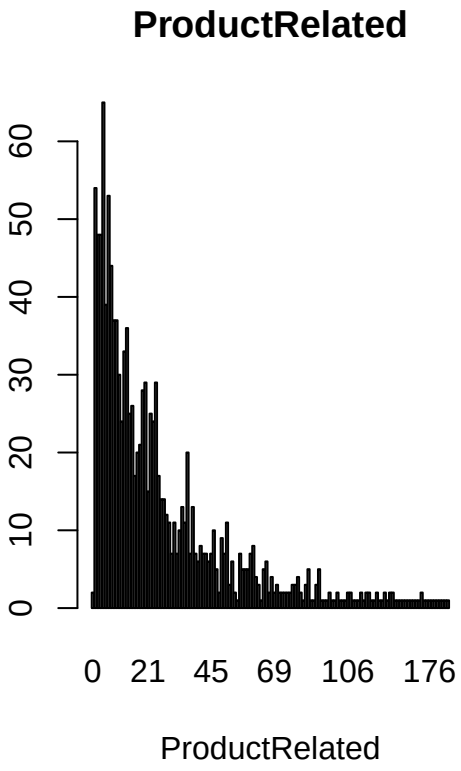
```
holiday <- shop[shop$SpecialDay > 0,]
par(mfrow = c(1,2))
for (i in 1:18) {
    barplot(table(holiday[, ..i]),main = names(holiday)[i], xlab = names(holiday)[i])
}
```

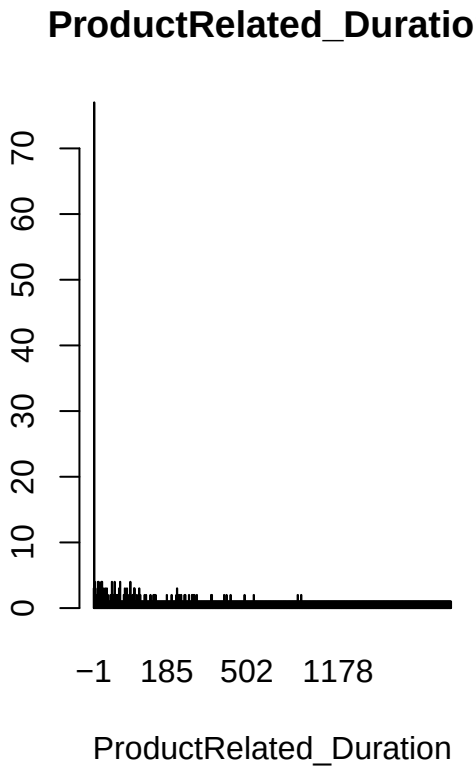

Administrative
0
200
400
600
800
0
3
6
9
12
17
Administrative

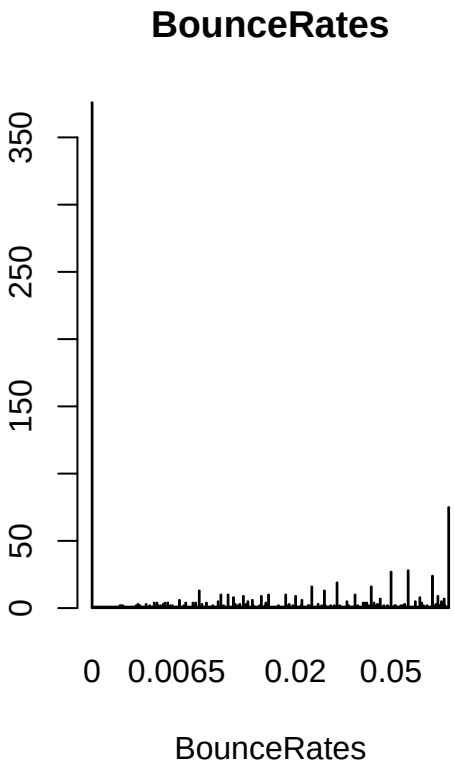

Administrative_Duration
0
200
400
600
800
−1
32
56
93
188
598
Administrative_Duration

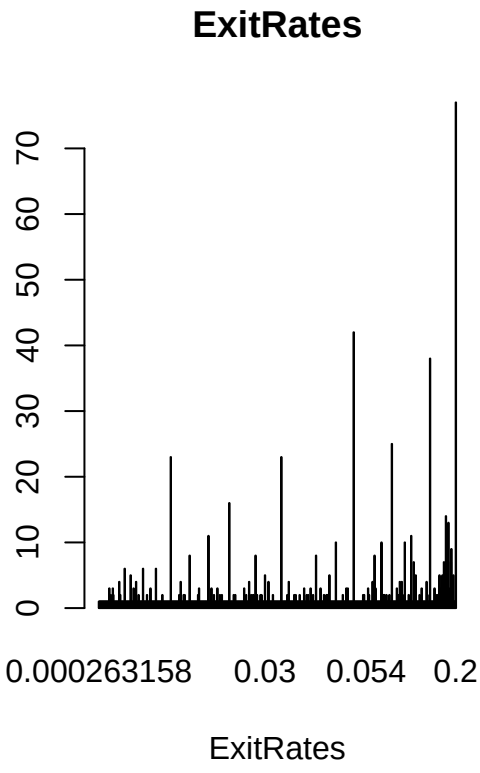

Informational
0
200
400
600
800
1000
0
2
4
6
8
11
Informational

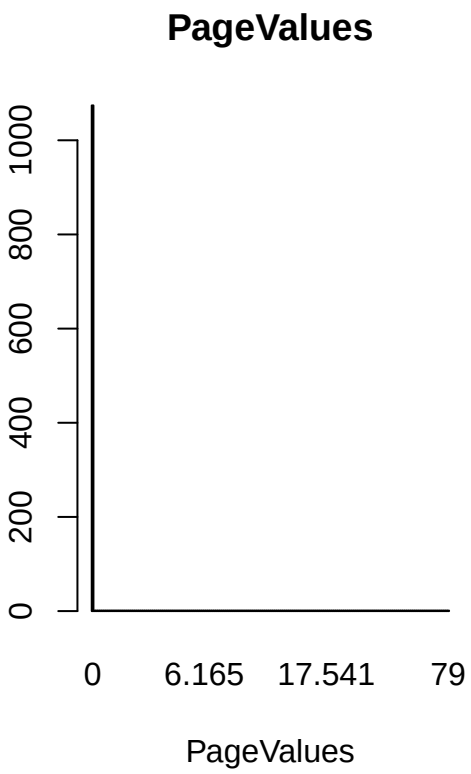

Informational_Duration
0
200
400
600
800
1000
−1
20
48
108
292
Informational_Duration

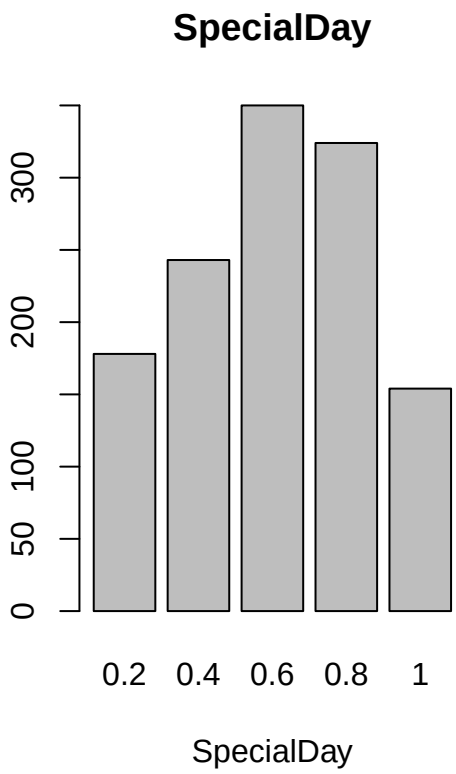

ProductRelated
0
10
20
30
40
50
60
0
21
45
69
106
176
ProductRelated

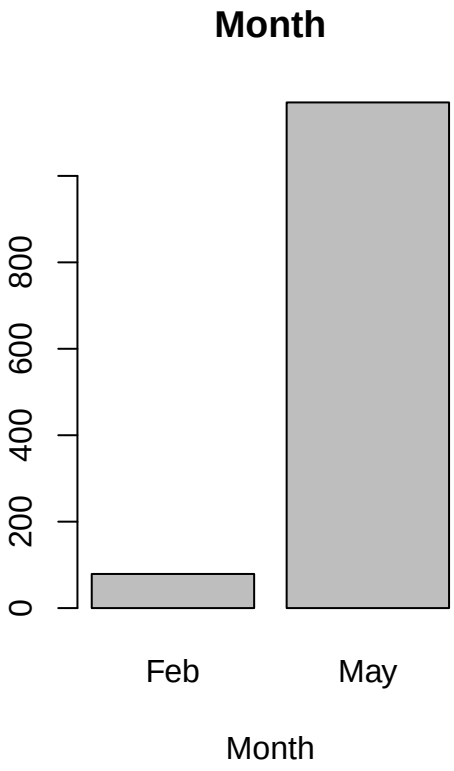

ProductRelated_Duration
0
10
20
30
40
50
60
70
−1
185
502
1178
ProductRelated_Duration

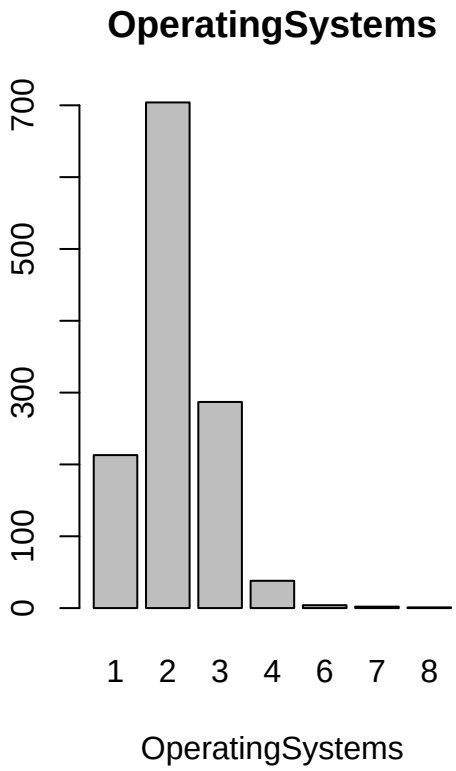

BounceRates
0
50
150
250
350
0
0.0065
0.02
0.05
BounceRates

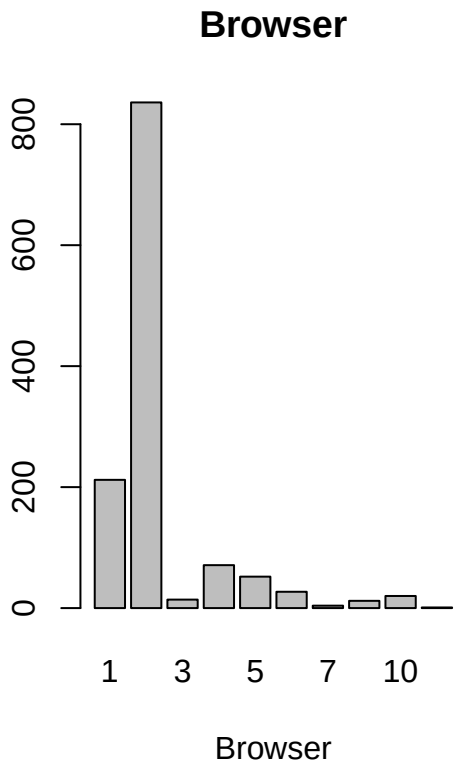

ExitRates
0
10
20
30
40
50
60
70
0.000263158
0.03
0.054
0.2
ExitRates

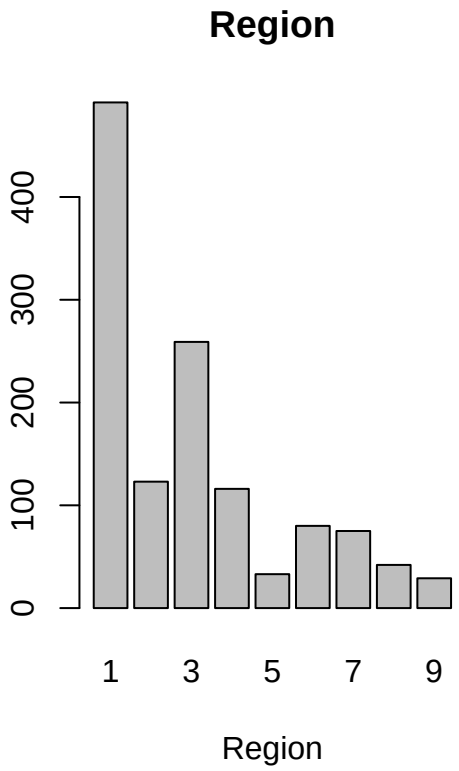

PageValues
0
200
400
600
800
1000
0
6.165
17.541
79
PageValues

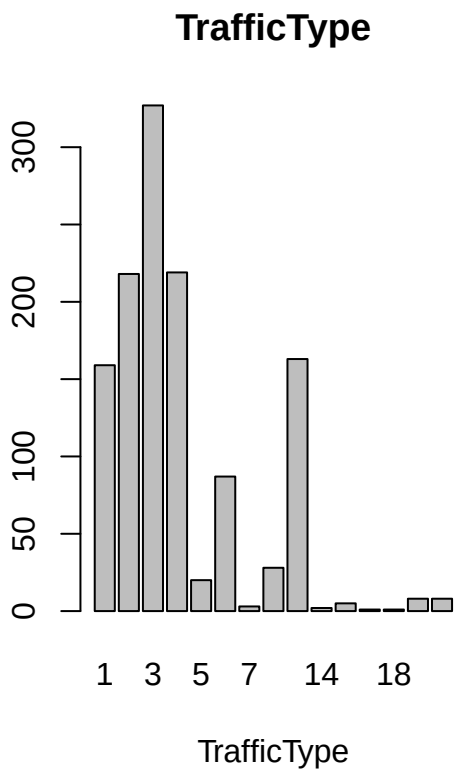

SpecialDay
0
50
100
200
300
0.2
0.4
0.6
0.8
1
SpecialDay

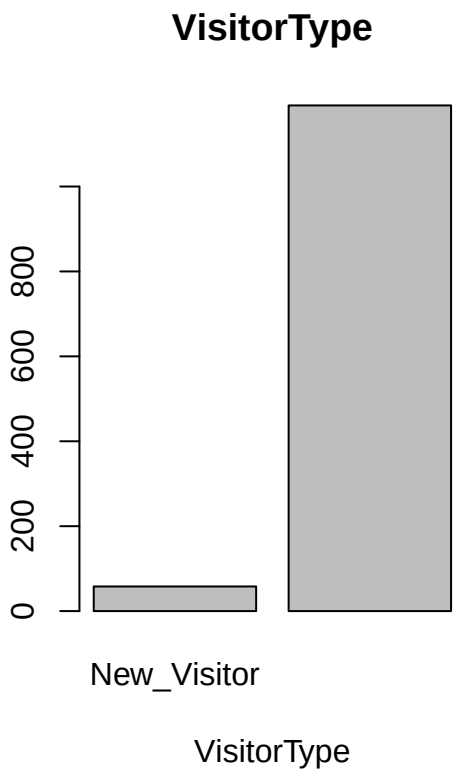

Month
0
200
400
600
800
Feb
May
Month

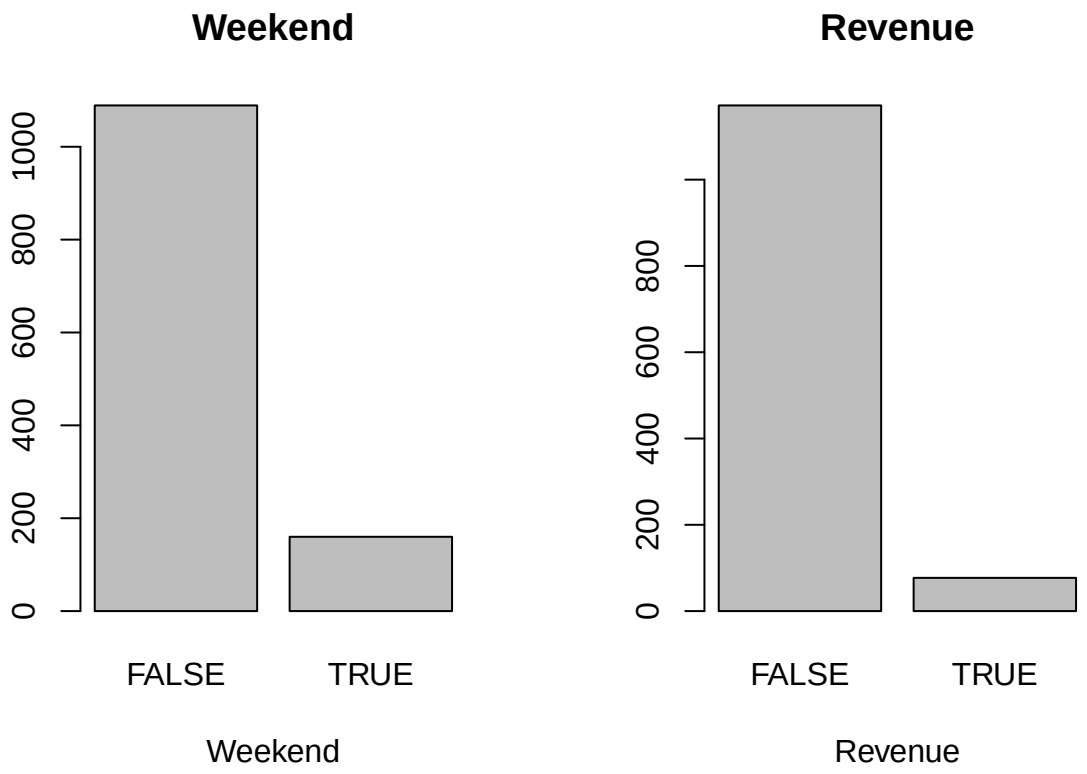

OperatingSystems
0
100
300
500
700
1
2
3
4
6
7
8
OperatingSystems


Browser
0
200
400
600
800
1
3
5
7
10
Browser


Region
0
100
200
300
400
1
3
5
7
9
Region


TrafficType
0
50
100
200
300
1
3
5
7
14
18
TrafficType


VisitorType
0
200
400
600
800
New_Visitor
VisitorType




```
describe(holiday)
```

Descriptive Statistics for the near holiday column

```
## holiday
##
##  18  Variables      1249  Observations
## --------------------------------------------------------------------------------
## Administrative
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0       17     0.72    1.339    2.139        0        0
##      .25      .50      .75      .90      .95
##        0        0        2        5        7
##
## lowest :  0  1  2  3  4, highest: 12 14 16 17 24
##
## Value          0     1     2     3     4     5     6     7     8     9    10
## Frequency    816   108    91    55    41    37    29    14    17    10    14
## Proportion 0.653 0.086 0.073 0.044 0.033 0.030 0.023 0.011 0.014 0.008 0.011
##
## Value         11    12    14    16    17    24
## Frequency      6     4     3     1     2     1
## Proportion 0.005 0.003 0.002 0.001 0.002 0.001
```

```
## ---------------------------------------------------------------------------
## Administrative_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0      313    0.714    42.27    73.42      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0     26.0    118.8    234.3
##
## lowest :   -1.000    0.000    2.000    3.000    3.500
## highest:  975.750 1023.000 1261.917 1321.250 1474.500
## ---------------------------------------------------------------------------
## Informational
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0       12    0.388   0.3155   0.5707        0        0
##      .25      .50      .75      .90      .95
##        0        0        0        1        2
##
## lowest :  0  1  2  3  4, highest:  7  8  9 11 12
##
## Value          0     1     2     3     4     5     6     7     8     9    11
## Frequency   1060    96    50    17    11     5     4     2     1     1     1
## Proportion 0.849 0.077 0.040 0.014 0.009 0.004 0.003 0.002 0.001 0.001 0.001
##
## Value         12
## Frequency      1
## Proportion 0.001
## ---------------------------------------------------------------------------
## Informational_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0      133    0.353    22.13    42.38      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0      0.0     21.6    105.8
##
## lowest :   -1.000    0.000    1.000    1.500    2.000
## highest: 1150.000 1258.833 1360.000 1636.000 1779.167
## ---------------------------------------------------------------------------
## ProductRelated
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0      136    0.999    27.42       30      2.0      3.0
##      .25      .50      .75      .90      .95
##      7.0     17.0     36.0     63.2     89.0
##
## lowest :   0   1   2   3   4, highest: 195 204 220 313 440
## ---------------------------------------------------------------------------
## ProductRelated_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0     1048        1    954.5     1186        0       23
##      .25      .50      .75      .90      .95
##      127      471     1248     2430     3630
##
## lowest :    -1.000     0.000     3.000     4.000     5.000
## highest:  7968.903  8388.316  9143.436  9951.869 11308.098
## ---------------------------------------------------------------------------
## BounceRates
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

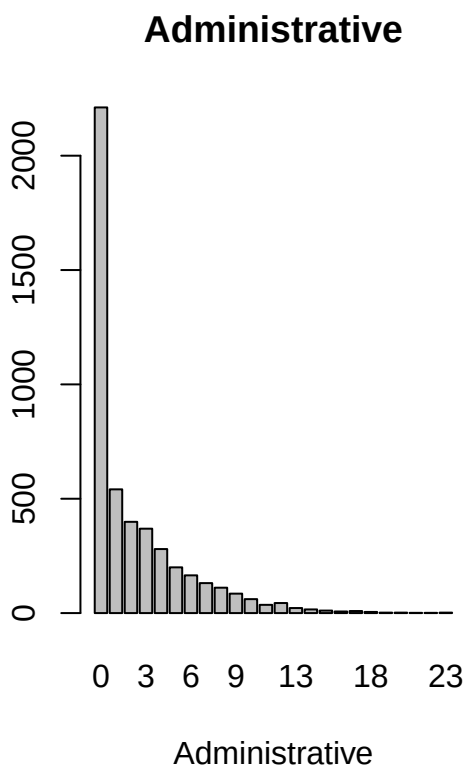
```
##     1249        0      397    0.972  0.03367  0.04713  0.00000  0.00000
##      .25      .50      .75      .90      .95
##  0.00000  0.01176  0.04000  0.10000  0.20000
##
## lowest : 0.000000000 0.000217391 0.000287356 0.000531915 0.000555556
## highest: 0.160000000 0.164285714 0.176923077 0.183333333 0.200000000
## ---------------------------------------------------------------------------
## ExitRates
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0      721        1  0.05884  0.05338 0.008333 0.012867
##      .25      .50      .75      .90      .95
## 0.022611 0.040441 0.075000 0.145556 0.200000
##
## lowest : 0.000263158 0.000409836 0.000833333 0.001449275 0.001470588
## highest: 0.180000000 0.183333333 0.188888889 0.192307692 0.200000000
## ---------------------------------------------------------------------------
## PageValues
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0      176    0.364    2.328    4.319    0.000    0.000
##      .25      .50      .75      .90      .95
##    0.000    0.000    0.000    5.815   17.132
##
## lowest :  0.0000000  0.2016637  0.5133860  0.6731276  0.9369792
## highest: 59.9677714 60.3423529 60.4338526 72.7880000 79.0000000
## ---------------------------------------------------------------------------
## SpecialDay
##        n  missing distinct     Info     Mean      Gmd
##     1249        0        5    0.948   0.6053   0.2762
##
## lowest : 0.2 0.4 0.6 0.8 1.0, highest: 0.2 0.4 0.6 0.8 1.0
##
## Value        0.2   0.4   0.6   0.8   1.0
## Frequency    178   243   350   324   154
## Proportion 0.143 0.195 0.280 0.259 0.123
## ---------------------------------------------------------------------------
## Month
##        n  missing distinct
##     1249        0        2
##
## Value        Feb   May
## Frequency     79  1170
## Proportion 0.063 0.937
## ---------------------------------------------------------------------------
## OperatingSystems
##        n  missing distinct     Info     Mean      Gmd
##     1249        0        7    0.804    2.146    0.772
##
## lowest : 1 2 3 4 6, highest: 3 4 6 7 8
##
## Value          1     2     3     4     6     7     8
## Frequency    213   704   287    38     4     2     1
## Proportion 0.171 0.564 0.230 0.030 0.003 0.002 0.001
## ---------------------------------------------------------------------------
## Browser
```

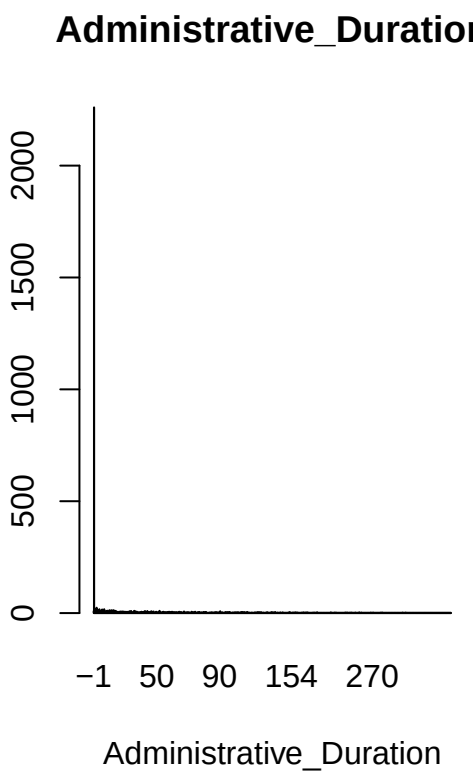
```
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0       10    0.695    2.376    1.252        1        1
##      .25      .50      .75      .90      .95
##        2        2        2        4        6
##
## lowest :  1  2  3  4  5, highest:  6  7  8 10 12
##
## Value          1     2     3     4     5     6     7     8    10    12
## Frequency    212   836    14    71    52    27     4    12    20     1
## Proportion 0.170 0.669 0.011 0.057 0.042 0.022 0.003 0.010 0.016 0.001
## ---------------------------------------------------------------------------
## Region
##        n  missing distinct     Info     Mean      Gmd
##     1249        0        9    0.928    2.999    2.384
##
## lowest : 1 2 3 4 5, highest: 5 6 7 8 9
##
## Value          1     2     3     4     5     6     7     8     9
## Frequency    492   123   259   116    33    80    75    42    29
## Proportion 0.394 0.098 0.207 0.093 0.026 0.064 0.060 0.034 0.023
## ---------------------------------------------------------------------------
## TrafficType
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1249        0       15    0.967    4.781    4.036        1        1
##      .25      .50      .75      .90      .95
##        2        3        5       13       13
##
## lowest :  1  2  3  4  5, highest: 15 17 18 19 20
##
## Value          1     2     3     4     5     6     7    11    13    14    15
## Frequency    159   218   327   219    20    87     3    28   163     2     5
## Proportion 0.127 0.175 0.262 0.175 0.016 0.070 0.002 0.022 0.131 0.002 0.004
##
## Value         17    18    19    20
## Frequency      1     1     8     8
## Proportion 0.001 0.001 0.006 0.006
## ---------------------------------------------------------------------------
## VisitorType
##        n  missing distinct
##     1249        0        2
##
## Value            New_Visitor Returning_Visitor
## Frequency                 58              1191
## Proportion             0.046             0.954
## ---------------------------------------------------------------------------
## Weekend
##        n  missing distinct
##     1249        0        2
##
## Value      FALSE  TRUE
## Frequency   1089   160
## Proportion 0.872 0.128
## ---------------------------------------------------------------------------
## Revenue
```

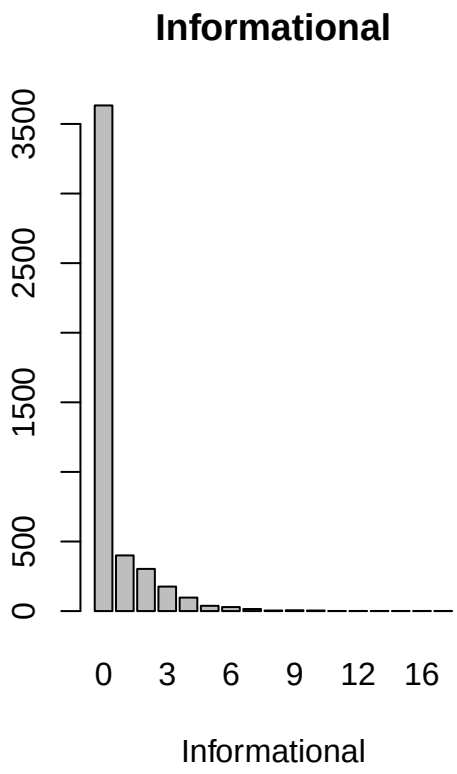
```
##        n  missing distinct
##     1249        0        2
##
## Value      FALSE  TRUE
## Frequency   1172    77
## Proportion 0.938 0.062
## ---------------------------------------------------------------------------
```

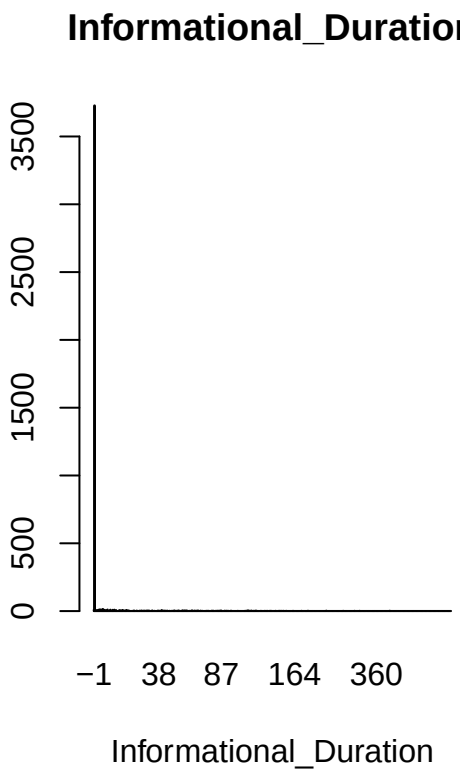
Revenue in region 1

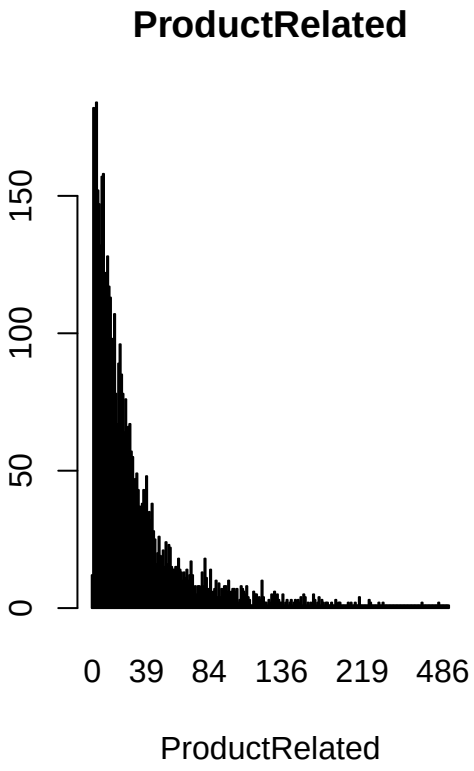
```
region <- shop[shop$Region == 1,]
par(mfrow = c(1,2))
for (i in 1:18) {
    barplot(table(region[, ..i]),main = names(region)[i], xlab = names(region)[i])

}
```

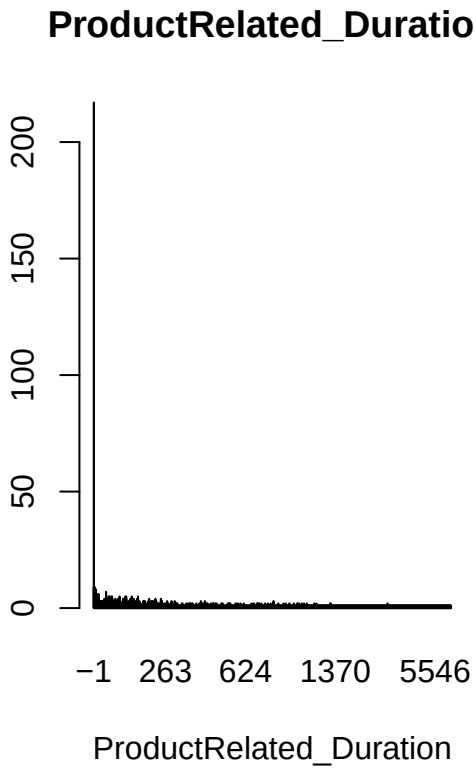


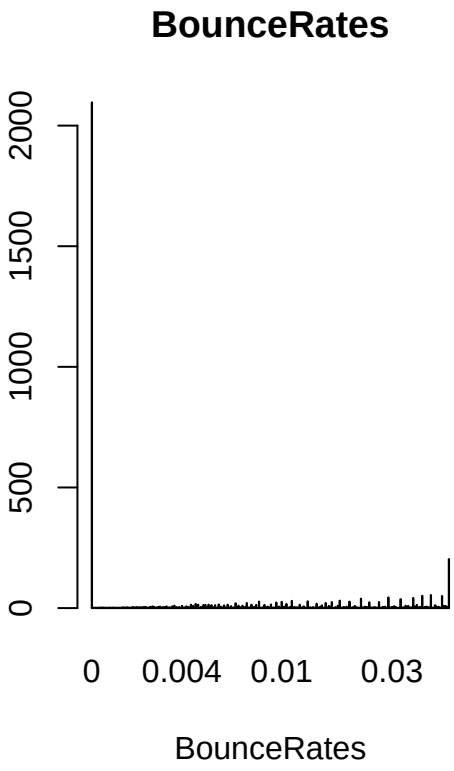


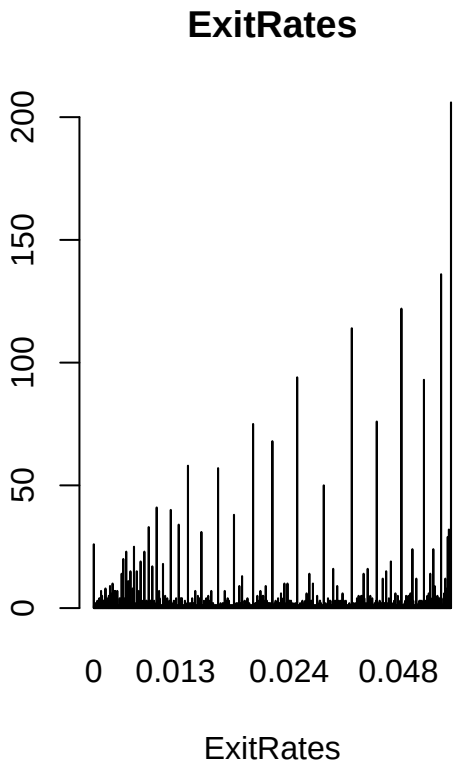

Informational
0
500
1500
2500
3500
0
3
6
9
12
16
Informational

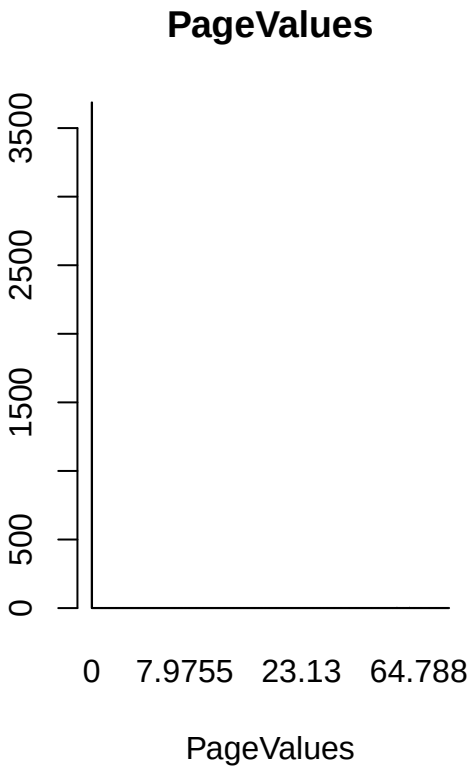

Informational_Duration
0
500
1500
2500
3500
−1
38
87
164
360
Informational_Duration

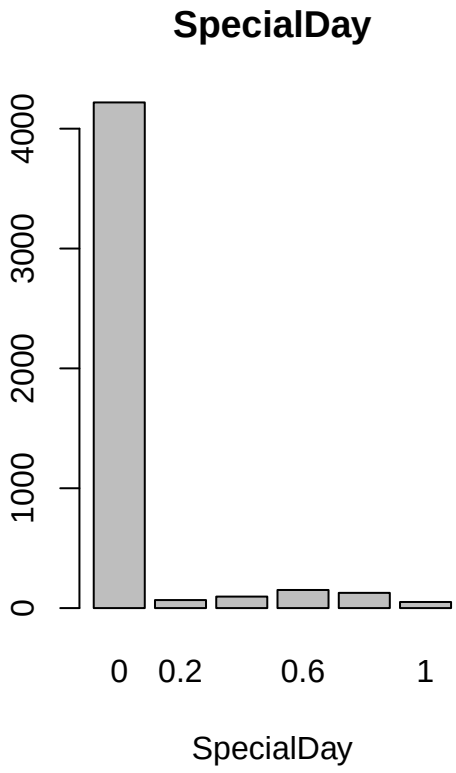

ProductRelated
0
50
100
150
0
39
84
136
219
486
ProductRelated

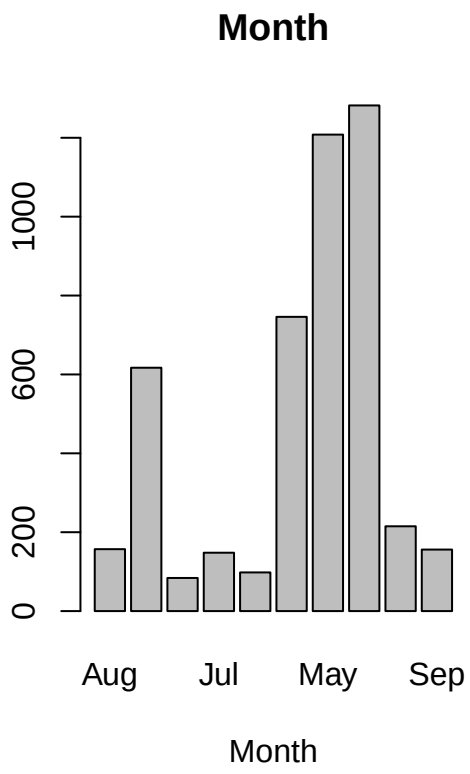

ProductRelated_Duration
0
50
100
150
200
−1
263
624
1370
5546
ProductRelated_Duration

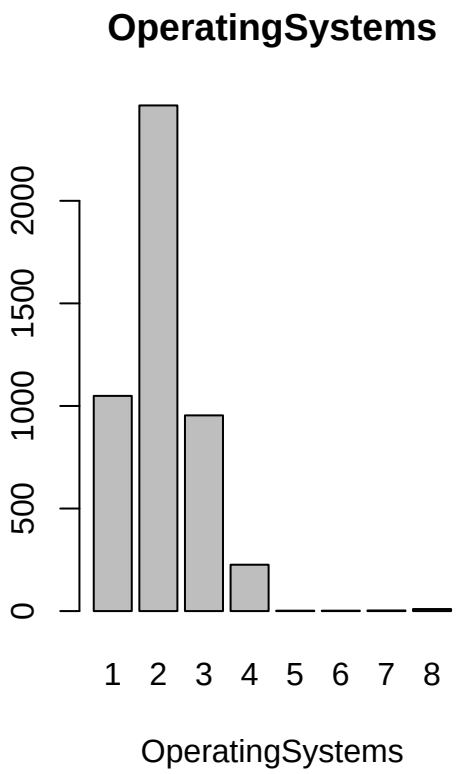

BounceRates
0
500
1000
1500
2000
0
0.004
0.01
0.03
BounceRates

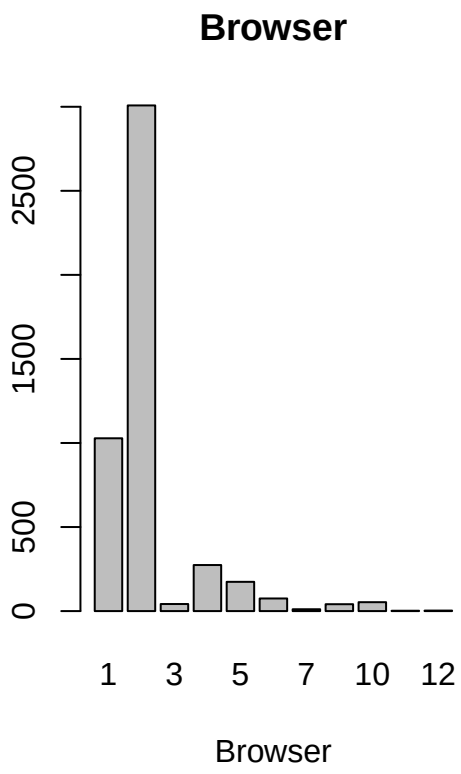

ExitRates
0
50
100
150
200
0
0.013
0.024
0.048
ExitRates

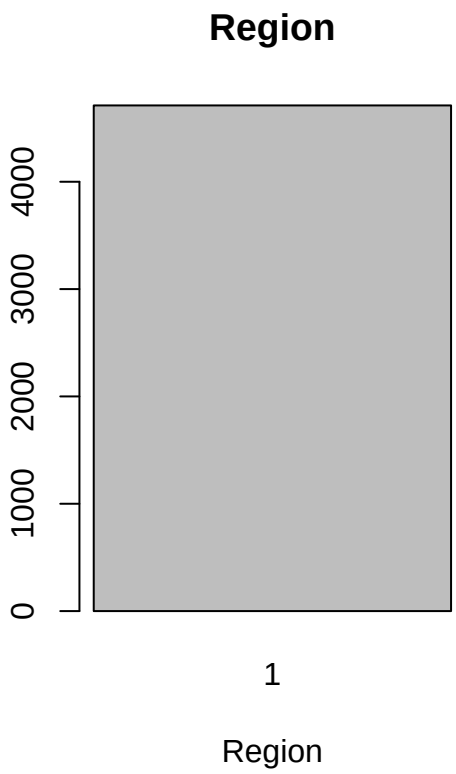

PageValues
0
500
1500
2500
3500
0
7.9755
23.13
64.788
PageValues

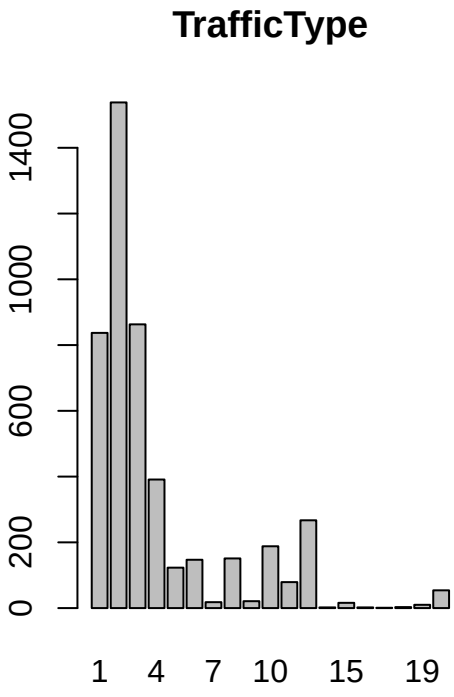

SpecialDay
0
1000
2000
3000
4000
0
0.2
0.6
1
SpecialDay

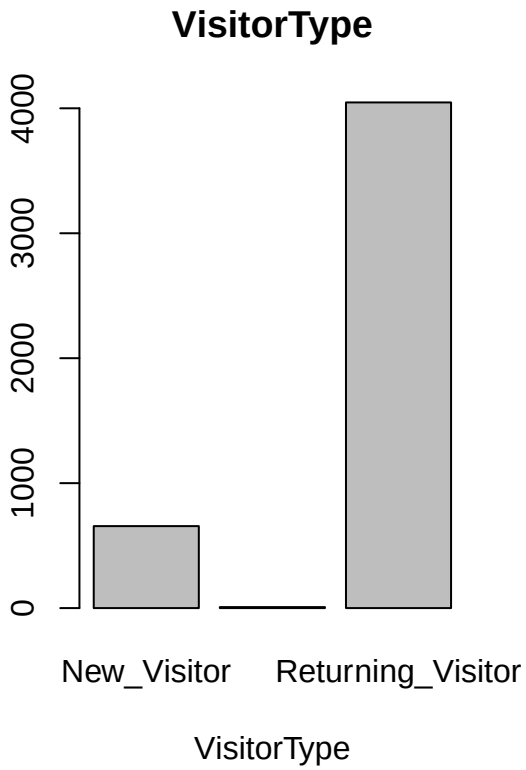

Month
0
200
600
1000
Aug
Jul
May
Sep
Month

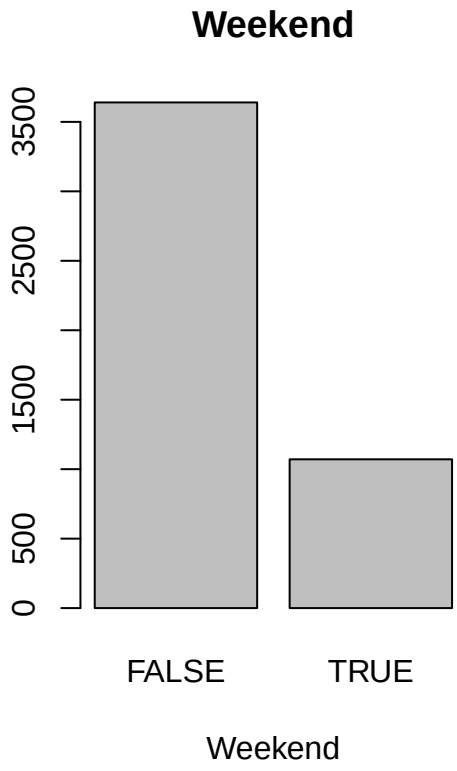

OperatingSystems
0
500
1000
1500
2000
1
2
3
4
5
6
7
8
OperatingSystems

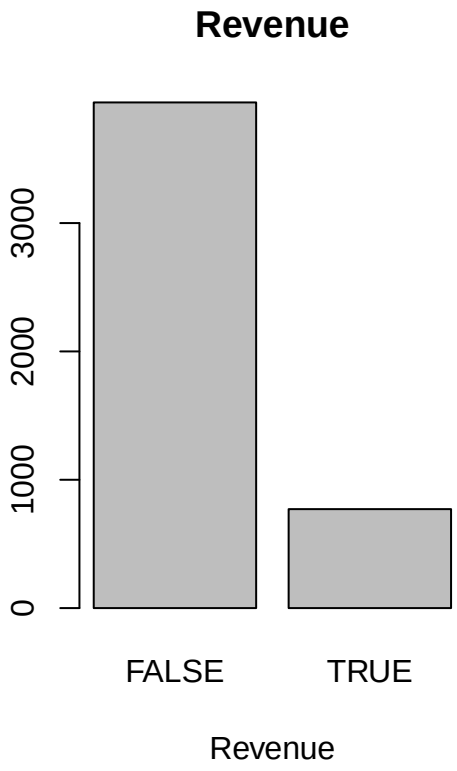

Browser
0
500
1500
2500
1
3
5
7
10
12
Browser


Region
0
1000
2000
3000
4000
1
Region


TrafficType
0
200
600
1000
1400
1
4
7
10
15
19
TrafficType


VisitorType
0
1000
2000
3000
4000
New_Visitor
Returning_Visitor
VisitorType






```
describe(region)
```

descriptive statistics for region

```
## region
##
##  18  Variables      4711  Observations
## ---------------------------------------------------------------------------
## Administrative
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0       24    0.894    2.305    3.174        0        0
##      .25      .50      .75      .90      .95
##        0        1        4        7        9
##
## lowest :  0  1  2  3  4, highest: 19 20 21 22 23
## ---------------------------------------------------------------------------
## Administrative_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0     1550     0.89    80.81    126.5      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      7.0     88.0    219.3    349.6
##
## lowest :   -1.000000    0.000000    1.333333    2.000000    3.000000
```

```
## highest: 2407.423810 2629.253968 2657.318056 2720.500000 3398.750000
## ---------------------------------------------------------------------------
## Informational
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0       17     0.54   0.5534   0.9431        0        0
##      .25      .50      .75      .90      .95
##        0        0        0        2        3
##
## lowest :  0  1  2  3  4, highest: 12 13 14 16 24
##
## Value          0     1     2     3     4     5     6     7     8     9    10
## Frequency   3633   400   303   176    97    38    29    15     4     6     4
## Proportion 0.771 0.085 0.064 0.037 0.021 0.008 0.006 0.003 0.001 0.001 0.001
##
## Value         11    12    13    14    16    24
## Frequency      1     1     1     1     1     1
## Proportion 0.000 0.000 0.000 0.000 0.000 0.000
## ---------------------------------------------------------------------------
## Informational_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0      634    0.504    37.82    70.18      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0      0.0     80.5    211.4
##
## lowest :   -1.000    0.000    1.000    1.500    2.000
## highest: 1949.167 2050.433 2166.500 2256.917 2549.375
## ---------------------------------------------------------------------------
## ProductRelated
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0      257    0.999    34.28    40.11        2        3
##      .25      .50      .75      .90      .95
##        8       19       39       81      121
##
## lowest :   0   1   2   3   4, highest: 486 501 517 686 705
## ---------------------------------------------------------------------------
## ProductRelated_Duration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0     3978        1     1295     1649      5.0     46.0
##      .25      .50      .75      .90      .95
##    188.2    606.8   1507.9   3181.5   4784.6
##
## lowest :    -1.00     0.00     1.00     4.00     5.00
## highest: 23342.08 27009.86 29970.47 43171.23 63973.52
## ---------------------------------------------------------------------------
## BounceRates
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0      992    0.912   0.0197  0.03167 0.000000 0.000000
##      .25      .50      .75      .90      .95
## 0.000000 0.003044 0.015385 0.050000 0.120000
##
## lowest : 0.000000000 0.000033500 0.000039400 0.000070900 0.000123762
## highest: 0.155555556 0.160000000 0.166666667 0.176923077 0.200000000
## ---------------------------------------------------------------------------
## ExitRates
```

```
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0     2205        1  0.04096  0.04153 0.004797 0.007882
##      .25      .50      .75      .90      .95
## 0.014286 0.025000 0.047066 0.100000 0.152778
##
## lowest : 0.000000000 0.000262123 0.000468384 0.000505051 0.000586510
## highest: 0.175000000 0.177777778 0.180000000 0.192307692 0.200000000
## ------------------------------------------------------------------------------
## PageValues
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0     1023    0.521    5.989    10.72     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00     0.00    20.30    40.08
##
## lowest :   0.0000000   0.0986214   0.1206999   0.1318370   0.1392006
## highest: 214.3066627 254.6071579 261.4912857 270.7846931 287.9537928
## ------------------------------------------------------------------------------
## SpecialDay
##        n  missing distinct     Info     Mean      Gmd
##     4711        0        6    0.282  0.06262   0.1151
##
## lowest : 0.0 0.2 0.4 0.6 0.8, highest: 0.2 0.4 0.6 0.8 1.0
##
## Value        0.0   0.2   0.4   0.6   0.8   1.0
## Frequency   4219    67    96   151   127    51
## Proportion 0.896 0.014 0.020 0.032 0.027 0.011
## ------------------------------------------------------------------------------
## Month
##        n  missing distinct
##     4711        0       10
##
## lowest : Aug  Dec  Feb  Jul  June, highest: Mar  May  Nov  Oct  Sep
##
## Value        Aug   Dec   Feb   Jul  June   Mar   May   Nov   Oct   Sep
## Frequency    157   617    84   148    98   746  1208  1282   215   156
## Proportion 0.033 0.131 0.018 0.031 0.021 0.158 0.256 0.272 0.046 0.033
## ------------------------------------------------------------------------------
## OperatingSystems
##        n  missing distinct     Info     Mean      Gmd
##     4711        0        8    0.837    2.095   0.8465
##
## lowest : 1 2 3 4 5, highest: 4 5 6 7 8
##
## Value          1     2     3     4     5     6     7     8
## Frequency   1049  2465   954   226     2     2     3    10
## Proportion 0.223 0.523 0.203 0.048 0.000 0.000 0.001 0.002
## ------------------------------------------------------------------------------
## Browser
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0       11    0.729    2.246    1.175        1        1
##      .25      .50      .75      .90      .95
##        2        2        2        4        5
##
## lowest :  1  2  3  4  5, highest:  7  8 10 11 12
```

```
##
## Value          1     2     3     4     5     6     7     8    10    11    12
## Frequency   1028  3008    42   274   174    75    11    41    53     2     3
## Proportion 0.218 0.639 0.009 0.058 0.037 0.016 0.002 0.009 0.011 0.000 0.001
## ---------------------------------------------------------------------------
## Region
##        n  missing distinct     Info     Mean      Gmd
##     4711        0        1        0        1        0
##
## Value         1
## Frequency  4711
## Proportion    1
## ---------------------------------------------------------------------------
## TrafficType
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     4711        0       19    0.953    4.022    3.585        1        1
##      .25      .50      .75      .90      .95
##        2        2        4       10       13
##
## lowest :  1  2  3  4  5, highest: 16 17 18 19 20
##
## Value          1     2     3     4     5     6     7     8     9    10    11
## Frequency    837  1538   863   391   123   147    18   151    21   188    79
## Proportion 0.178 0.326 0.183 0.083 0.026 0.031 0.004 0.032 0.004 0.040 0.017
##
## Value         13    14    15    16    17    18    19    20
## Frequency    267     2    16     2     1     3    10    54
## Proportion 0.057 0.000 0.003 0.000 0.000 0.001 0.002 0.011
## ---------------------------------------------------------------------------
## VisitorType
##        n  missing distinct
##     4711        0        3
##
## Value            New_Visitor             Other Returning_Visitor
## Frequency                656                 8              4047
## Proportion             0.139             0.002             0.859
## ---------------------------------------------------------------------------
## Weekend
##        n  missing distinct
##     4711        0        2
##
## Value      FALSE  TRUE
## Frequency   3640  1071
## Proportion 0.773 0.227
## ---------------------------------------------------------------------------
## Revenue
##        n  missing distinct
##     4711        0        2
##
## Value      FALSE  TRUE
## Frequency   3940   771
## Proportion 0.836 0.164
## ---------------------------------------------------------------------------
```

Here's our findings from the most popular region 1 on customer behavior: Most popular months for visitors from region 1 are March, May and November.

52.3% of all visitors using operating system type 2 while 22.3% were using type 1

Browser 2 had brought in the most visitors with 62.9% while browser 1 brough in 14.3% of the visitors.

16.4% of all visitors from region one brought in revenue.

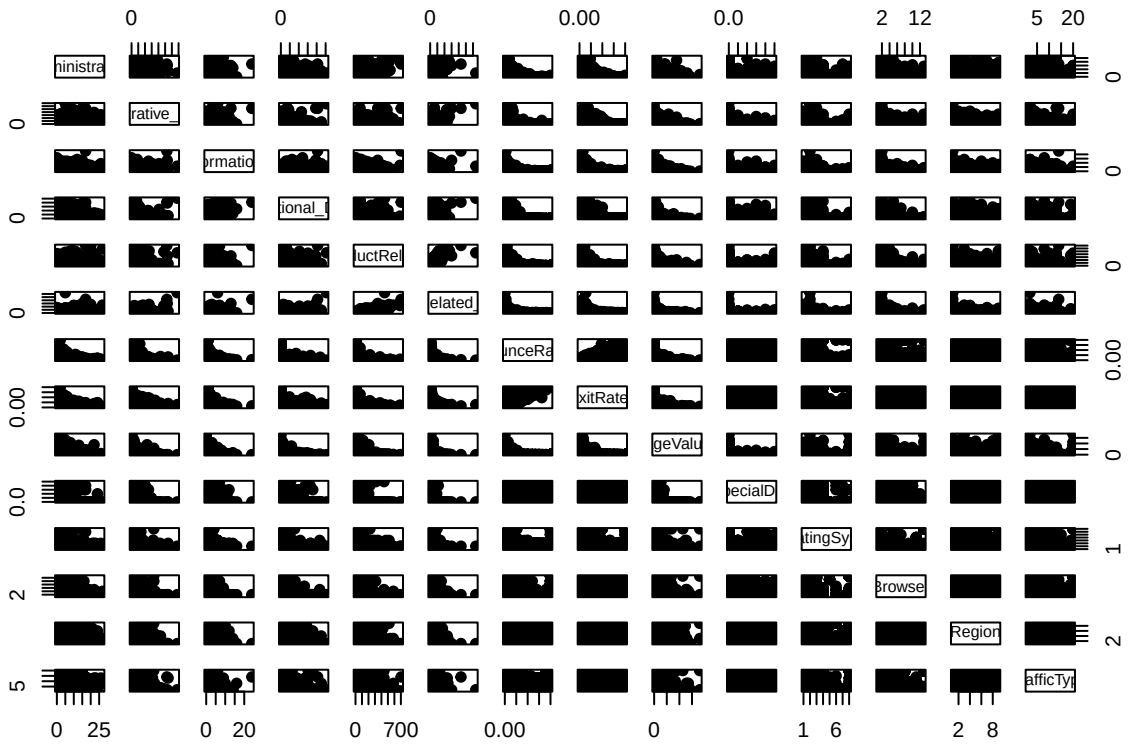
Trffic type 1,2 and 3 bring in the most number of visitors from region 1 with type 2 being the highest at 32.6%.

13.9% of the visitors was from new visitors while 85.9% were return visitors.

22.7% of site visitors was gained during the weekend while the rest 77.3% was gained during the weekdays.

Correlation Analysis

```
pairs(num_col, pch = 19)
```



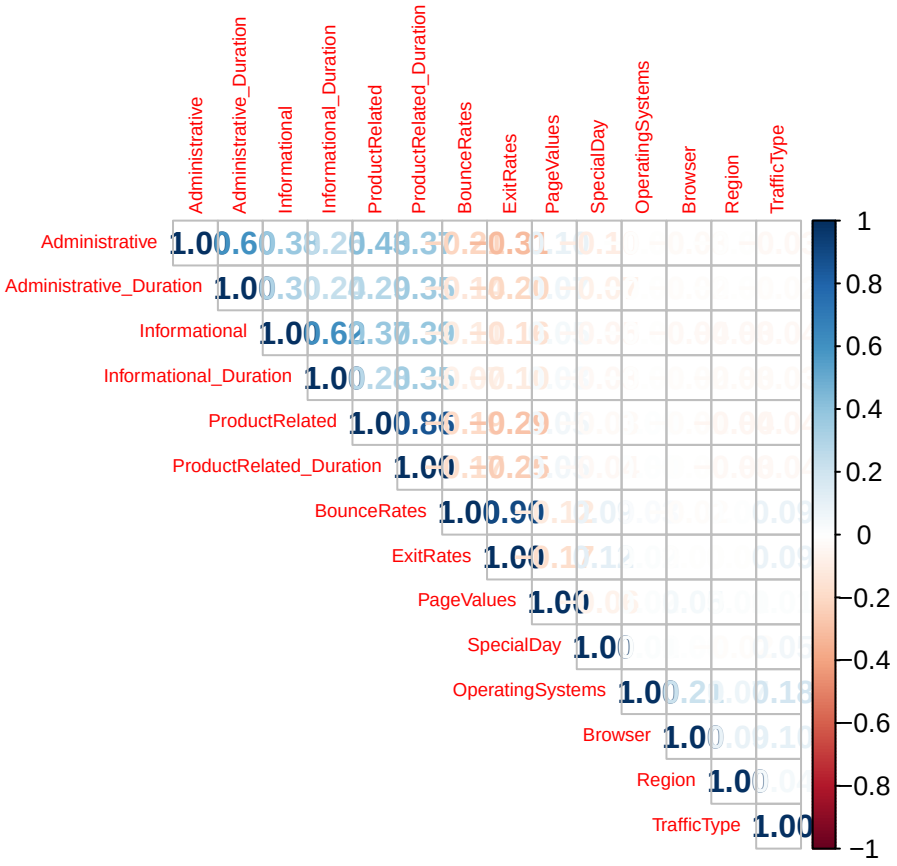
let's get the Correlation Coefficients

```
cor(num_col)
```

```
##                          Administrative Administrative_Duration Informational
## Administrative             1.000000000             0.600409653    0.37528761
## Administrative_Duration    0.600409653             1.000000000    0.30143630
## Informational              0.375287611             0.301436296    1.00000000
## Informational_Duration     0.254786021             0.237189860    0.61867795
## ProductRelated             0.428191515             0.286783914    0.37260472
## ProductRelated_Duration    0.371027224             0.353513793    0.38608372
## BounceRates               -0.213666635            -0.137333397   -0.10950530
## ExitRates                 -0.311274132            -0.202024452   -0.15956681
## PageValues                 0.096920968             0.066168365    0.04739015
## SpecialDay                -0.097072098            -0.074736885   -0.04937677
## OperatingSystems          -0.006697922            -0.007610715   -0.00962587
## Browser                   -0.025763658            -0.015833675   -0.03876681
## Region                    -0.007262053            -0.006723711   -0.03047732
## TrafficType               -0.034784126            -0.015075015   -0.03518669
##                          Informational_Duration ProductRelated
## Administrative                      0.254786021    0.428191515
## Administrative_Duration             0.237189860    0.286783914
## Informational                       0.618677947    0.372604721
## Informational_Duration              1.000000000    0.279061948
## ProductRelated                      0.279061948    1.000000000
## ProductRelated_Duration             0.346580691    0.860308186
## BounceRates                        -0.070159472   -0.193515772
## ExitRates                          -0.102932678   -0.286163211
## PageValues                          0.030064160    0.054115494
## SpecialDay                         -0.031293040   -0.025930622
## OperatingSystems                   -0.009749983    0.004090351
## Browser                            -0.019609349   -0.013706213
## Region                             -0.027920098   -0.040106501
## TrafficType                        -0.025163571   -0.044344333
##                          ProductRelated_Duration  BounceRates    ExitRates
## Administrative                      0.371027224 -0.213666635 -0.311274132
## Administrative_Duration             0.353513793 -0.137333397 -0.202024452
## Informational                       0.386083717 -0.109505298 -0.159566815
## Informational_Duration              0.346580691 -0.070159472 -0.102932678
## ProductRelated                      0.860308186 -0.193515772 -0.286163211
## ProductRelated_Duration             1.000000000 -0.174375499 -0.245334012
## BounceRates                        -0.174375499  1.000000000  0.903358192
## ExitRates                          -0.245334012  0.903358192  1.000000000
## PageValues                          0.050840624 -0.115991977 -0.173571542
## SpecialDay                         -0.038210652  0.087839995  0.116783762
## OperatingSystems                    0.002775788  0.026839839  0.016482012
## Browser                            -0.007838332 -0.016018380 -0.003565541
## Region                             -0.034862498  0.001432015 -0.001837556
## TrafficType                        -0.037506944  0.089199039  0.087386232
##                           PageValues   SpecialDay OperatingSystems      Browser
## Administrative            0.09692097 -0.097072098     -0.006697922 -0.025763658
## Administrative_Duration   0.06616837 -0.074736885     -0.007610715 -0.015833675
## Informational             0.04739015 -0.049376774     -0.009625870 -0.038766808
## Informational_Duration    0.03006416 -0.031293040     -0.009749983 -0.019609349
## ProductRelated            0.05411549 -0.025930622      0.004090351 -0.013706213
## ProductRelated_Duration   0.05084062 -0.038210652      0.002775788 -0.007838332
## BounceRates              -0.11599198  0.087839995      0.026839839 -0.016018380
## ExitRates                -0.17357154  0.116783762      0.016482012 -0.003565541
```

```
## PageValues                1.00000000 -0.064532709      0.018583782  0.045845065
## SpecialDay               -0.06453271  1.000000000      0.012757766  0.003465984
## OperatingSystems          0.01858378  0.012757766      1.000000000  0.212244823
## Browser                   0.04584506  0.003465984      0.212244823  1.000000000
## Region                    0.01059087 -0.016452464      0.071953240  0.091889464
## TrafficType               0.01223694  0.052827944      0.182874100  0.102886237
##                                Region TrafficType
## Administrative           -0.007262053 -0.03478413
## Administrative_Duration  -0.006723711 -0.01507502
## Informational            -0.030477323 -0.03518669
## Informational_Duration   -0.027920098 -0.02516357
## ProductRelated           -0.040106501 -0.04434433
## ProductRelated_Duration  -0.034862498 -0.03750694
## BounceRates               0.001432015  0.08919904
## ExitRates                -0.001837556  0.08738623
## PageValues                0.010590868  0.01223694
## SpecialDay               -0.016452464  0.05282794
## OperatingSystems          0.071953240  0.18287410
## Browser                   0.091889464  0.10288624
## Region                    1.000000000  0.04252523
## TrafficType               0.042525234  1.00000000
```

```
corrplot(cor(num_col), type = 'upper', method = 'number', tl.cex = 0.6)
```



Corrplot

The variables can be seen to have very weak correlations

Let's Perform Clustering

```
# normalizing the data
web_traffic <- num_col
web_traffic[,c('Revenue')]<- list(NULL)
```

removing revenue from the clustering dataset as it is the class label

```
## Warning in `[<-.data.table`(`*tmp*`, , c("Revenue"), value = list(NULL)): Column
## 'Revenue' does not exist to remove
```

```
# previewing
head(web_traffic)
```

```
##    Administrative Administrative_Duration Informational Informational_Duration
## 1:              0                       0             0                      0
## 2:              0                       0             0                      0
## 3:              0                      -1             0                     -1
## 4:              0                       0             0                      0
## 5:              0                       0             0                      0
## 6:              0                       0             0                      0
##    ProductRelated ProductRelated_Duration BounceRates ExitRates PageValues
## 1:              1                0.000000  0.20000000 0.2000000          0
## 2:              2               64.000000  0.00000000 0.1000000          0
## 3:              1               -1.000000  0.20000000 0.2000000          0
## 4:              2                2.666667  0.05000000 0.1400000          0
## 5:             10              627.500000  0.02000000 0.0500000          0
## 6:             19              154.216667  0.01578947 0.0245614          0
##    SpecialDay OperatingSystems Browser Region TrafficType
## 1:          0                1       1      1           1
## 2:          0                2       2      1           2
## 3:          0                4       1      9           3
## 4:          0                3       2      2           4
## 5:          0                3       3      1           4
## 6:          0                2       2      1           3
```

```
web_traffic <- scale(web_traffic)
head(web_traffic)
```

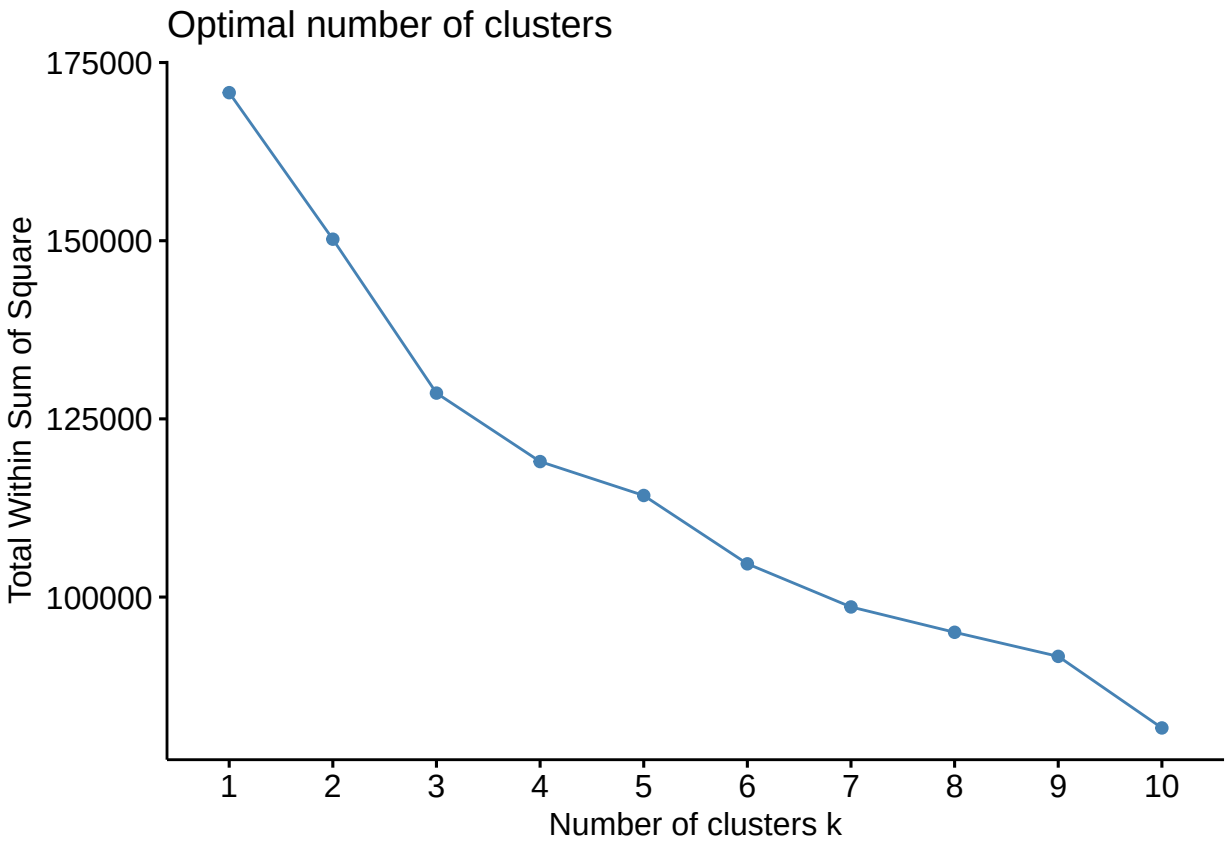
```
##      Administrative Administrative_Duration Informational
## [1,]     -0.7025315              -0.4601081    -0.3988128
## [2,]     -0.7025315              -0.4601081    -0.3988128
## [3,]     -0.7025315              -0.4657410    -0.3988128
## [4,]     -0.7025315              -0.4601081    -0.3988128
## [5,]     -0.7025315              -0.4601081    -0.3988128
## [6,]     -0.7025315              -0.4601081    -0.3988128
##      Informational_Duration ProductRelated ProductRelated_Duration  BounceRates
```

```
## [1,]             -0.2462725     -0.6963635             -0.6289343  3.954699721
## [2,]             -0.2462725     -0.6739424             -0.5955997 -0.450343788
## [3,]             -0.2533417     -0.6963635             -0.6294551  3.954699721
## [4,]             -0.2462725     -0.6739424             -0.6275453  0.650917089
## [5,]             -0.2462725     -0.4945739             -0.3020990 -0.009839437
## [6,]             -0.2462725     -0.2927843             -0.5486101 -0.102577188
##       ExitRates PageValues SpecialDay OperatingSystems    Browser     Region
## [1,]  3.4273070 -0.3190356 -0.3103105       -1.2396607 -0.7939682 -0.8962939
## [2,]  1.2650121 -0.3190356 -0.3103105       -0.1371074 -0.2093703 -0.8962939
## [3,]  3.4273070 -0.3190356 -0.3103105        2.0679992 -0.7939682  2.4336556
## [4,]  2.1299300 -0.3190356 -0.3103105        0.9654459 -0.2093703 -0.4800502
## [5,]  0.1838646 -0.3190356 -0.3103105        0.9654459  0.3752276 -0.8962939
## [6,] -0.3661929 -0.3190356 -0.3103105       -0.1371074 -0.2093703 -0.8962939
##      TrafficType
## [1,] -0.76562243
## [2,] -0.51660683
## [3,] -0.26759123
## [4,] -0.01857564
## [5,] -0.01857564
## [6,] -0.26759123
```

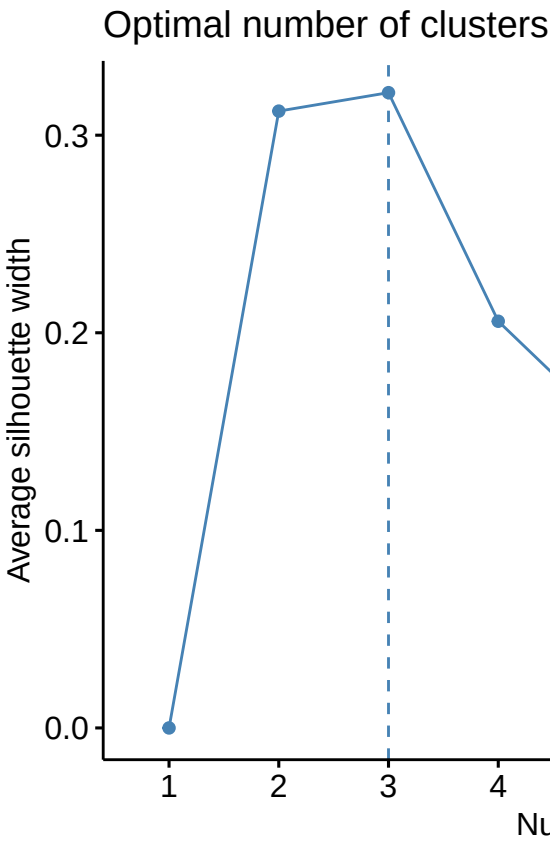
KMeans Clustering

let's find the best value for k

```
fviz_nbclust(x = web_traffic,FUNcluster = kmeans, method = 'wss' )
```




```
# optimal values using average silhouette method

fviz_nbclust(x = web_traffic,FUNcluster = kmeans, method = 'silhouette' )
```



let's find the optimal clusters k using (average silhouette method)

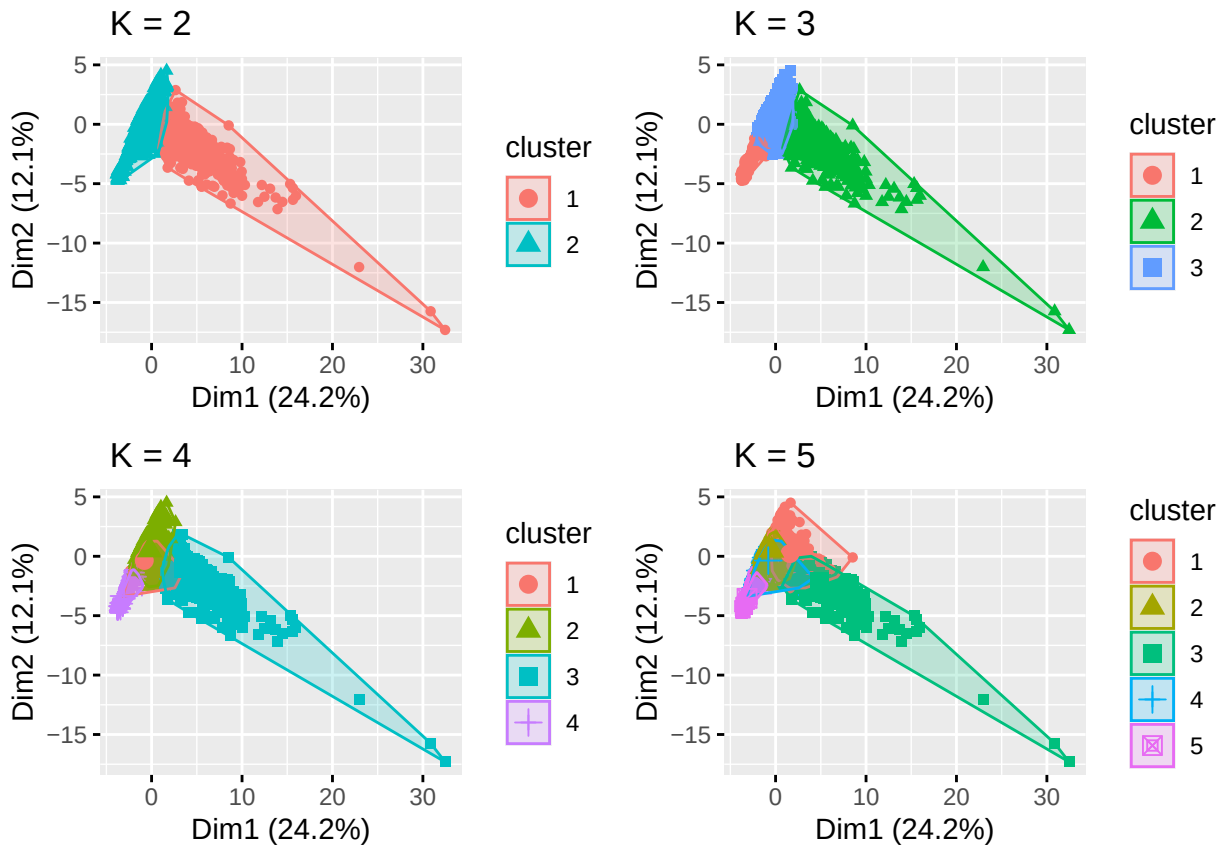
The best value for k from the graph is 3

```
# let's test with a few values for k
data_K3 <- kmeans(web_traffic, centers = 2, nstart = 50)
data_K4 <- kmeans(web_traffic, centers = 3, nstart = 50)
data_K5 <- kmeans(web_traffic, centers = 4, nstart = 50)
data_K6 <- kmeans(web_traffic, centers = 5, nstart = 50)

#lets plot these clusters for different K value to compare.
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
p1 <- fviz_cluster(data_K3, geom = "point", data = web_traffic) + ggtitle(" K = 2")
p2 <- fviz_cluster(data_K4, geom = "point", data = web_traffic) + ggtitle(" K = 3")
p3 <- fviz_cluster(data_K5, geom = "point", data = web_traffic) + ggtitle(" K = 4")
p4 <- fviz_cluster(data_K6, geom = "point", data = web_traffic) + ggtitle(" K = 5")

grid.arrange(p1, p2, p3, p4, nrow = 2)
```




```
set.seed(123)
# let's compute kmeans clusters
data_kmns1 <- kmeans(num_col, centers = 1, nstart = 25)
print(data_kmns1)
```

```
## K-means clustering with 1 clusters of sizes 12199
##
## Cluster means:
##   Administrative Administrative_Duration Informational Informational_Duration
## 1       2.340028                81.68214     0.5088122                34.83734
##   ProductRelated ProductRelated_Duration BounceRates  ExitRates PageValues
## 1       32.05845                1207.508  0.02044674 0.04149678     5.9525
##   SpecialDay OperatingSystems  Browser   Region TrafficType
## 1 0.06197229         2.124354 2.358144 3.153291    4.074596
##
## Clustering vector:
##     [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
##    [37] 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
##    [73] 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
##   [109] 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
##   [145] 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
##   [181] 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
##   [217] 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
##   [253] 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
##   [289] 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
```

```
##   [325] 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
##   [361] 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
##   [397] 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
##   [433] 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
##   [469] 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
##   [505] 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
##   [541] 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
##   [577] 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
##   [613] 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
##   [649] 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
##   [685] 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
##   [721] 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
##   [757] 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
##   [793] 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
##   [829] 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
##   [865] 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
##   [901] 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
##   [937] 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
##   [973] 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
##  [1009] 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
##  [1045] 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
##  [1081] 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
##  [1117] 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
##  [1153] 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
##  [1189] 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
##  [1225] 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
##  [1261] 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
##  [1297] 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
##  [1333] 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
##  [1369] 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
##  [1405] 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
##  [1441] 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
##  [1477] 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
##  [1513] 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
##  [1549] 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
##  [1585] 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
##  [1621] 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
##  [1657] 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
##  [1693] 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
##  [1729] 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
##  [1765] 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
##  [1801] 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
##  [1837] 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
##  [1873] 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
##  [1909] 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
##  [1945] 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
##  [1981] 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
##  [2017] 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
##  [2053] 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
##  [2089] 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
##  [2125] 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
##  [2161] 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
##  [2197] 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
##  [2233] 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
```

```
##  [2269] 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
##  [2305] 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
##  [2341] 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
##  [2377] 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
##  [2413] 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
##  [2449] 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
##  [2485] 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
##  [2521] 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
##  [2557] 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
##  [2593] 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
##  [2629] 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
##  [2665] 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
##  [2701] 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
##  [2737] 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
##  [2773] 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
##  [2809] 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
##  [2845] 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
##  [2881] 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
##  [2917] 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
##  [2953] 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
##  [2989] 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
##  [3025] 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
##  [3061] 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
##  [3097] 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
##  [3133] 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
##  [3169] 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
##  [3205] 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
##  [3241] 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
##  [3277] 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
##  [3313] 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
##  [3349] 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
##  [3385] 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
##  [3421] 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
##  [3457] 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
##  [3493] 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
##  [3529] 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
##  [3565] 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
##  [3601] 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
##  [3637] 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
##  [3673] 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
##  [3709] 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
##  [3745] 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
##  [3781] 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
##  [3817] 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
##  [3853] 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
##  [3889] 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
##  [3925] 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
##  [3961] 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
##  [3997] 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
##  [4033] 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
##  [4069] 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
##  [4105] 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
##  [4141] 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
##  [4177] 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
```

```
##  [4213] 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
##  [4249] 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
##  [4285] 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
##  [4321] 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
##  [4357] 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
##  [4393] 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
##  [4429] 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
##  [4465] 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
##  [4501] 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
##  [4537] 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
##  [4573] 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
##  [4609] 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
##  [4645] 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
##  [4681] 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
##  [4717] 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
##  [4753] 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
##  [4789] 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
##  [4825] 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
##  [4861] 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
##  [4897] 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
##  [4933] 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
##  [4969] 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
##  [5005] 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
##  [5041] 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
##  [5077] 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
##  [5113] 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
##  [5149] 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
##  [5185] 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
##  [5221] 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
##  [5257] 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
##  [5293] 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
##  [5329] 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
##  [5365] 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
##  [5401] 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
##  [5437] 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
##  [5473] 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
##  [5509] 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
##  [5545] 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
##  [5581] 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
##  [5617] 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
##  [5653] 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
##  [5689] 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
##  [5725] 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
##  [5761] 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
##  [5797] 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
##  [5833] 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
##  [5869] 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
##  [5905] 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
##  [5941] 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
##  [5977] 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
##  [6013] 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
##  [6049] 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
##  [6085] 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
##  [6121] 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
```

```
##  [6157] 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
##  [6193] 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
##  [6229] 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
##  [6265] 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
##  [6301] 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
##  [6337] 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
##  [6373] 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
##  [6409] 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
##  [6445] 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
##  [6481] 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
##  [6517] 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
##  [6553] 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
##  [6589] 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
##  [6625] 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
##  [6661] 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
##  [6697] 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
##  [6733] 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
##  [6769] 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
##  [6805] 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
##  [6841] 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
##  [6877] 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
##  [6913] 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
##  [6949] 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
##  [6985] 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
##  [7021] 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
##  [7057] 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
##  [7093] 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
##  [7129] 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
##  [7165] 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
##  [7201] 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
##  [7237] 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
##  [7273] 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
##  [7309] 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
##  [7345] 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
##  [7381] 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
##  [7417] 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
##  [7453] 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
##  [7489] 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
##  [7525] 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
##  [7561] 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
##  [7597] 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
##  [7633] 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
##  [7669] 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
##  [7705] 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
##  [7741] 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
##  [7777] 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
##  [7813] 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
##  [7849] 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
##  [7885] 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
##  [7921] 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
##  [7957] 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
##  [7993] 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
##  [8029] 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
##  [8065] 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
```

```
##  [8101] 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
##  [8137] 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
##  [8173] 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
##  [8209] 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
##  [8245] 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
##  [8281] 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
##  [8317] 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
##  [8353] 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
##  [8389] 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
##  [8425] 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
##  [8461] 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
##  [8497] 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
##  [8533] 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
##  [8569] 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
##  [8605] 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
##  [8641] 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
##  [8677] 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
##  [8713] 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
##  [8749] 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
##  [8785] 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
##  [8821] 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
##  [8857] 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
##  [8893] 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
##  [8929] 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
##  [8965] 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
##  [9001] 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
##  [9037] 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
##  [9073] 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
##  [9109] 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
##  [9145] 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
##  [9181] 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
##  [9217] 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
##  [9253] 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
##  [9289] 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
##  [9325] 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
##  [9361] 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
##  [9397] 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
##  [9433] 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
##  [9469] 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
##  [9505] 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
##  [9541] 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
##  [9577] 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
##  [9613] 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
##  [9649] 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
##  [9685] 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
##  [9721] 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
##  [9757] 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
##  [9793] 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
##  [9829] 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
##  [9865] 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
##  [9901] 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
##  [9937] 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
##  [9973] 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
## [10009] 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
```

```
## [10045] 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
## [10081] 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
## [10117] 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
## [10153] 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
## [10189] 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
## [10225] 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
## [10261] 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
## [10297] 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
## [10333] 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
## [10369] 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
## [10405] 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
## [10441] 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
## [10477] 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
## [10513] 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
## [10549] 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
## [10585] 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
## [10621] 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
## [10657] 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
## [10693] 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
## [10729] 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
## [10765] 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
## [10801] 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
## [10837] 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
## [10873] 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
## [10909] 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
## [10945] 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
## [10981] 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
## [11017] 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
## [11053] 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
## [11089] 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
## [11125] 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
## [11161] 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
## [11197] 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
## [11233] 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
## [11269] 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
## [11305] 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
## [11341] 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
## [11377] 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
## [11413] 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
## [11449] 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
## [11485] 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
## [11521] 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
## [11557] 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
## [11593] 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
## [11629] 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
## [11665] 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
## [11701] 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
## [11737] 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
## [11773] 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
## [11809] 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
## [11845] 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
## [11881] 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
## [11917] 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
## [11953] 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
```

```
## [11989] 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
## [12025] 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
## [12061] 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
## [12097] 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
## [12133] 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
## [12169] 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
##
## Within cluster sum of squares by cluster:
## [1] 45620813020
##  (between_SS / total_SS =   0.0 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
compilation <- shop %>%
    mutate(cluster = data_kmns1$cluster) %>%
    select(Revenue, cluster)
head(compilation)
```

let's add clusters to our dataset as cols

```
##    Revenue cluster
## 1:   FALSE       1
## 2:   FALSE       1
## 3:   FALSE       1
## 4:   FALSE       1
## 5:   FALSE       1
## 6:   FALSE       1
```

```
# let's match the cluster categories to the revenue for comparison
compilation$cluster[compilation$cluster == 1] <- 'FALSE'
head(compilation)
```

```
##    Revenue cluster
## 1:   FALSE   FALSE
## 2:   FALSE   FALSE
## 3:   FALSE   FALSE
## 4:   FALSE   FALSE
## 5:   FALSE   FALSE
## 6:   FALSE   FALSE
```

```
# let's compare the revenue to the model cluster
table(compilation$cluster == compilation$Revenue)
```

```
##
## FALSE  TRUE
##  1908 10291
```

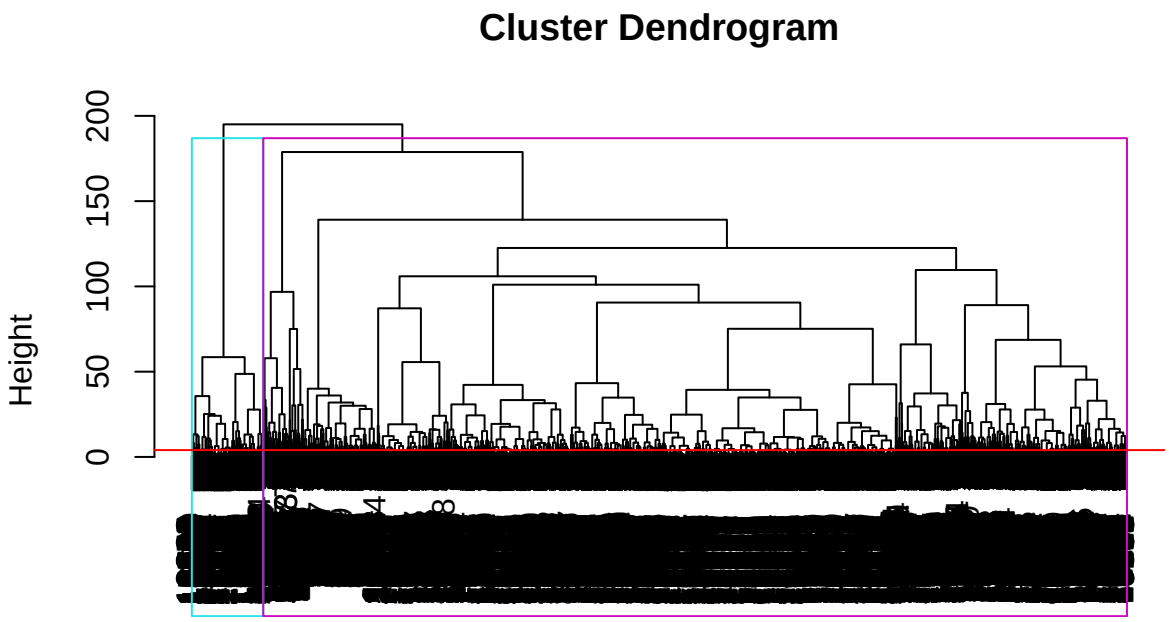
Using the revenue column for comparison, the model was able to match the revenue column with an accuracy of 83.3% which constituted 10,291 columns correctly matched.

Hierarchical Clustering

```
# let's calculate the distance

## methods to assess
#m <- c( "average", "single", "complete", "ward")
#names(m) <- c( "average", "single", "complete", "ward")
#
#library(cluster)
#library(purrr)
#
#library(factoextra)
# computing the coefficient with a function
#ac <- function(x) {
#    agnes(web_traffic, method = x)
#}
#
#map_dbl(m, ac)
```

```
# calculating distances between observations
dist_euc <- dist(web_traffic, method = 'euclidean')

# computing the manhattan distance between observarions
dist_man <- dist(web_traffic, method = "manhattan")
```

```
#use R's cutree() function to cut the tree with hclust_avg as one parameter and the other parameter as
hc.avg <- hclust(dist_euc, method = "ward.D2")
cut_avg <- cutree(hc.avg, k = 2)

plot(hc.avg)
rect.hclust(hc.avg , k = 2, border = 5:6)
abline(h = 4, col = 'red')
```




```
# using ward's method
hc5 <- hclust(dist_euc, method = "ward.D2" )

# Cut tree into 2 groups
sub_grp <- cutree(hc5, k = 2)

# Number of members in each cluster
table(sub_grp)
```

```
## sub_grp
##     1     2
##   932 11267
```

```
# Ward's method
hc_complete <- hclust(dist_euc, method = "complete" )

# Cut tree into 2 groups
sub_grp3 <- cutree(hc_complete, k = 2)

# Number of members in each cluster
table(sub_grp3)
```

```
## sub_grp3
##     1     2
## 12196     3
```

```r
# Ward's method
hc_avg <- hclust(dist_euc, method = "average" )

# Cut tree into 2 groups
sub_grp2 <- cutree(hc_avg, k = 2)

# Number of members in each cluster
table(sub_grp2)
```

```
## sub_grp2
##     1     2
## 12196     3
```

```r
# Ward's method
hc_complete <- hclust(dist_man, method = "ward.D2" )

# Cut tree into 2 groups
sub_grp3 <- cutree(hc_complete, k = 2)

# Number of members in each cluster
table(sub_grp3)
```

```
## sub_grp3
##     1     2
## 11268   931
```

```r
hcomp <- shop %>%
  mutate(cluster = sub_grp) %>%
  head
```

```r
# Adding the clusters as a column to our original dataset

hcomp <- shop %>%
  mutate(cluster2 = sub_grp) %>%
  select(Revenue, cluster2)
tail(hcomp, n= 10)
```

```
##     Revenue cluster2
##  1:   FALSE        2
##  2:   FALSE        1
##  3:   FALSE        2
##  4:   FALSE        2
##  5:   FALSE        2
##  6:   FALSE        2
##  7:   FALSE        2
##  8:   FALSE        2
##  9:   FALSE        2
## 10:   FALSE        2
```

```r
# Matching the cluster categories to the revenue for comparison
hcomp$cluster2[hcomp$cluster2 == 2] <- 'FALSE'
hcomp$cluster2[hcomp$cluster2 == 1] <- 'TRUE'
head(hcomp)
```

```
##    Revenue cluster2
## 1:   FALSE     TRUE
## 2:   FALSE    FALSE
## 3:   FALSE     TRUE
## 4:   FALSE     TRUE
## 5:   FALSE    FALSE
## 6:   FALSE    FALSE
```

```
#Comparing the revenue and the model cluster to see if the clusters match
table(hcomp$cluster == hcomp$Revenue)
```

```
##
## FALSE  TRUE
##  2824  9375
```

On model clustering using the revenue column as our comparison, the model was able to match the revenue column with a 76.8% accuracy score and we had 9375 columns matched correctly. This is a slight decline from the kmeans model.

Conclusion

- Mothers day holiday brought in more revenue than Valentines day.
- Most of the wed visits were during the month of May but November had more revenue than may.
- Most of the traffic and revenue was from region 1. During the holidays, more regions visit the site and contribute significantly to the total revenue.
- Traffic type 2 brought in the most visitors. Some of the traffic types did not bring in any visitors for all the 10 months under analysis.They should be eliminated when considering advertisement or re evaluated to find out the problem.
- Most of the revenue and visits was from return visitors. A good indicator of customer satisfaction.
- Return customers are the main source of revenue
- Bounce rate is high especial for new customers.
- Most of the site visits did not bring forth revenue