R Notebook

Ecommerce Customer - Unsupervised Learnining

Definining The Question

Specifying the Question

1. Perform clustering stating insights drawn from your analysis and visualizations.
2. Upon implementation, provide comparisons between the approaches learned this week i.e. K-Means clustering vs Hierarchical clustering highlighting the strengths and limitations of each approach in the context of your analysis.

Metric of success

- Importing the data
- Cleaning the data
- performing a thorough EDA
- Build K Means and Hierarchical model structures.

Data relevance

The dataset for this Independent project can be found here [http://bit.ly/EcommerceCustomersDataset].

- The dataset consists of **10 numerical** and **8 categorical attributes**. The **'Revenue'** attribute can be used as the class label. **"Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration"** represents the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories. The values of these features are derived from the URL information of the pages visited by the user and updated in real-time when a user takes an action, e.g. moving from one page to another.
- The **"Bounce Rate", "Exit Rate" and "Page Value"** features represent the metrics measured by "Google Analytics" for each page in the e-commerce site.
- The value of the **"Bounce Rate"** feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.
- The value of the **"Exit Rate"** feature for a specific web page is calculated as for all pageviews to the page, the percentage that was the last in the session.
- The **"Page Value"** feature represents the average value for a web page that a user visited before completing an e-commerce transaction.
- The **"Special Day"** feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with the transaction. The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentina's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.

- The dataset also includes the **operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.**

Understanding the context

Kira Plastinina 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

Experimental design

The experimental design will involve the following steps: Creating a Markdown which will comprise the following sections.

- Problem Definition
- Data Sourcing
- Check the Data
- Perform Data Cleaning
- Perform Exploratory Data Analysis (Univariate, Bivariate & Multivariate)
- Implement the Solution
- Challenge the Solution
- Follow up Questions

Reading The Data

```
#Install packages needed
#install.packages("naniar")
#pkgs <- c("factoextra",  "NbClust")
#install.packages(pkgs)
#install.packages("superml")
```

```
# Importing Libraries
library (tidyr)
library(naniar)
library (ggplot2)
library (e1071)
library (corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(factoextra)
```

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

```
library(NbClust)
library(superml)
```

```
## Loading required package: R6
```

```r
#installing packages
library(data.table)
#
#Loading the dataset
df <- fread("http://bit.ly/EcommerceCustomersDataset")
```

Checking The Data

```r
# Preview the data
head(df)
```

```
##    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
```

```r
# Preview the data
tail(df)
```

```
##    Administrative Administrative_Duration Informational Informational_Duration
## 1:              0                       0             1                       0
## 2:              3                     145             0                       0
## 3:              0                       0             0                       0
## 4:              0                       0             0                       0
## 5:              4                      75             0                       0
## 6:              0                       0             0                       0
##    ProductRelated ProductRelated_Duration BounceRates  ExitRates PageValues
```

```
## 1:             16                503.000 0.000000000 0.03764706    0.00000
## 2:             53               1783.792 0.007142857 0.02903061   12.24172
## 3:              5                465.750 0.000000000 0.02133333    0.00000
## 4:              6                184.250 0.083333333 0.08666667    0.00000
## 5:             15                346.000 0.000000000 0.02105263    0.00000
## 6:              3                 21.250 0.000000000 0.06666667    0.00000
##    SpecialDay Month OperatingSystems Browser Region TrafficType
## 1:          0   Nov                2       2      1           1
## 2:          0   Dec                4       6      1           1
## 3:          0   Nov                3       2      1           8
## 4:          0   Nov                3       2      1          13
## 5:          0   Nov                2       2      3          11
## 6:          0   Nov                3       2      1           2
##          VisitorType Weekend Revenue
## 1: Returning_Visitor   FALSE   FALSE
## 2: Returning_Visitor    TRUE   FALSE
## 3: Returning_Visitor    TRUE   FALSE
## 4: Returning_Visitor    TRUE   FALSE
## 5: Returning_Visitor   FALSE   FALSE
## 6:       New_Visitor    TRUE   FALSE
```

```
# Dimensionanity of the data
dim(df)
```

```
## [1] 12330    18
```

The dataframe has 12330 rows and 18 columns

Tidying The Dataset

```
# check the column names.
colnames(df)
```

```
##  [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 have a mixture of uppercase and lowercase charachers we should correct that and
#make all the characters lowercase.
names(df) <- tolower(names(df))
# Confirmation
colnames(df)
```

```
##  [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"
```

```
# Let us find the datatypes of the data
str(df)
```

```
## 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" "Returni
##  $ weekend                : logi  FALSE FALSE FALSE FALSE TRUE FALSE ...
##  $ revenue                : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

There are a number of datatypes. Most of the data is in numerical format The Month and Visitor type columns are in character The weekend and revenue columns are have logical values

```
#Finding the total number of missing values in each column
colSums(is.na(df))
```

```
##          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
```

There are several columns with missing values. They will affect the data analysis and will have to be dropped.

```
# Dropping null values.
df <- na.omit(df)

# Confirming that the nulls have been removed
colSums(is.na(df))
```

```
##          administrative administrative_duration           informational
##                       0                       0                       0
##  informational_duration          productrelated productrelated_duration
##                       0                       0                       0
##             bouncerates               exitrates              pagevalues
##                       0                       0                       0
##              specialday                   month        operatingsystems
##                       0                       0                       0
##                 browser                  region             traffictype
##                       0                       0                       0
##             visitortype                 weekend                 revenue
##                       0                       0                       0
```

Now there are no more nulls in the dataframe.

```
# Cheking for duplicates
df_dup <- df[duplicated(df),]
df_dup
```

```
##      administrative administrative_duration informational
##   1:              0                       0             0
##   2:              0                       0             0
##   3:              0                       0             0
##   4:              0                       0             0
##   5:              0                       0             0
##  ---
## 113:              0                       0             0
## 114:              0                       0             0
## 115:              0                       0             0
## 116:              0                       0             0
## 117:              0                       0             0
##      informational_duration productrelated productrelated_duration bouncerates
##   1:                      0              1                       0         0.2
##   2:                      0              1                       0         0.2
##   3:                      0              1                       0         0.2
##   4:                      0              1                       0         0.2
##   5:                      0              1                       0         0.2
##  ---
## 113:                      0              1                       0         0.2
## 114:                      0              1                       0         0.2
## 115:                      0              1                       0         0.2
## 116:                      0              1                       0         0.2
## 117:                      0              1                       0         0.2
##      exitrates pagevalues specialday month operatingsystems browser region
##   1:       0.2          0          0   Feb                1       1      1
##   2:       0.2          0          0   Feb                3       2      3
##   3:       0.2          0          0   Mar                1       1      1
```

```
##   4:       0.2          0          0   Mar                2       2      4
##   5:       0.2          0          0   Mar                3       2      3
##  ---
## 113:       0.2          0          0   Dec                1       1      1
## 114:       0.2          0          0   Dec                1       1      4
## 115:       0.2          0          0   Dec                1       1      1
## 116:       0.2          0          0   Dec                1      13      9
## 117:       0.2          0          0   Dec                8      13      9
##      traffictype       visitortype weekend revenue
##   1:           3 Returning_Visitor   FALSE   FALSE
##   2:           3 Returning_Visitor   FALSE   FALSE
##   3:           1 Returning_Visitor    TRUE   FALSE
##   4:           1 Returning_Visitor   FALSE   FALSE
##   5:           1 Returning_Visitor   FALSE   FALSE
##  ---
## 113:           2       New_Visitor   FALSE   FALSE
## 114:           1 Returning_Visitor    TRUE   FALSE
## 115:           3 Returning_Visitor   FALSE   FALSE
## 116:          20 Returning_Visitor   FALSE   FALSE
## 117:          20             Other   FALSE   FALSE
```

There are a number of duplicated records. The duplicated records sum up to 117 records.

```
# Shape before
dim(df)
```

```
## [1] 12316    18
```

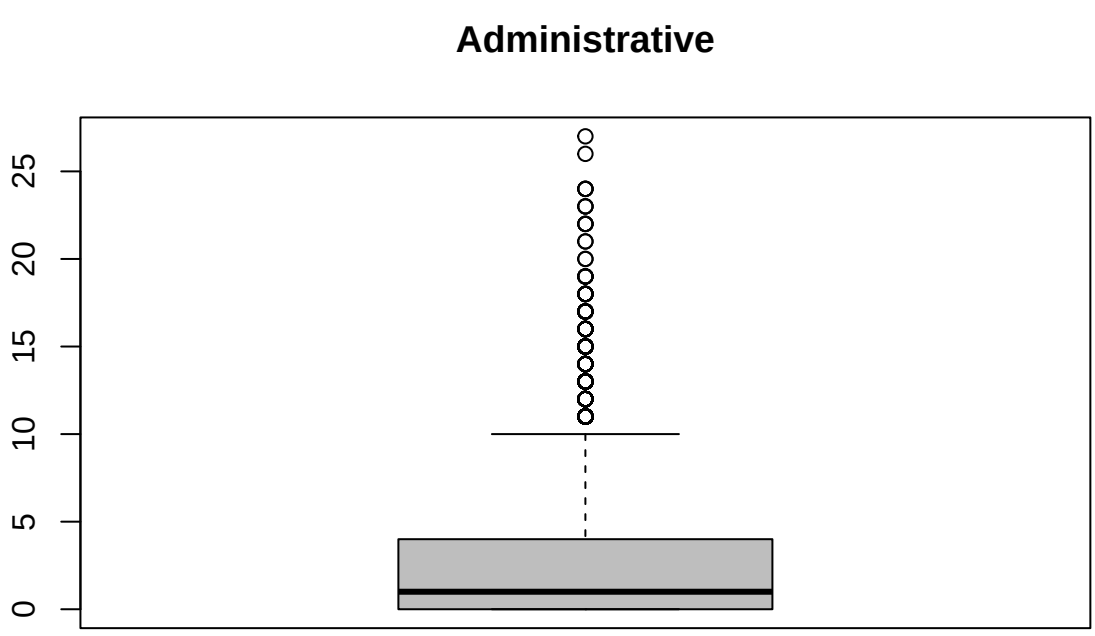
```
# Removing duplicates
df <- unique(df)

# Shape after(confirmation)
dim (df)
```

```
## [1] 12199    18
```

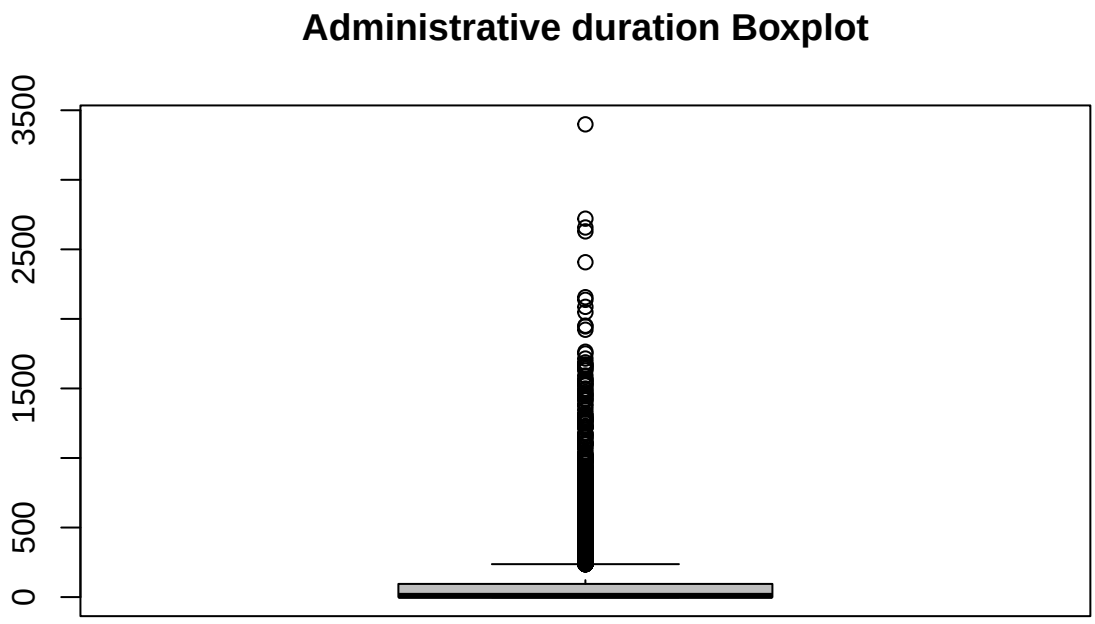
There are no more duplicates

Checking for outliers

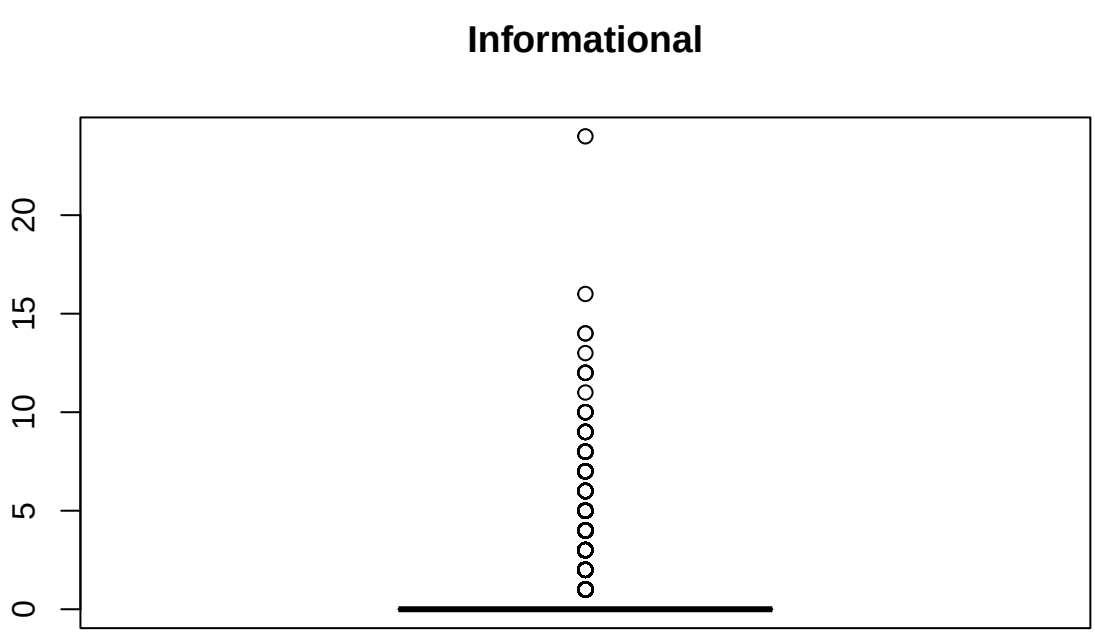
```
# Plotting boxplots to check for outliers
boxplot(df$administrative,col='grey', main = 'Administrative')
```



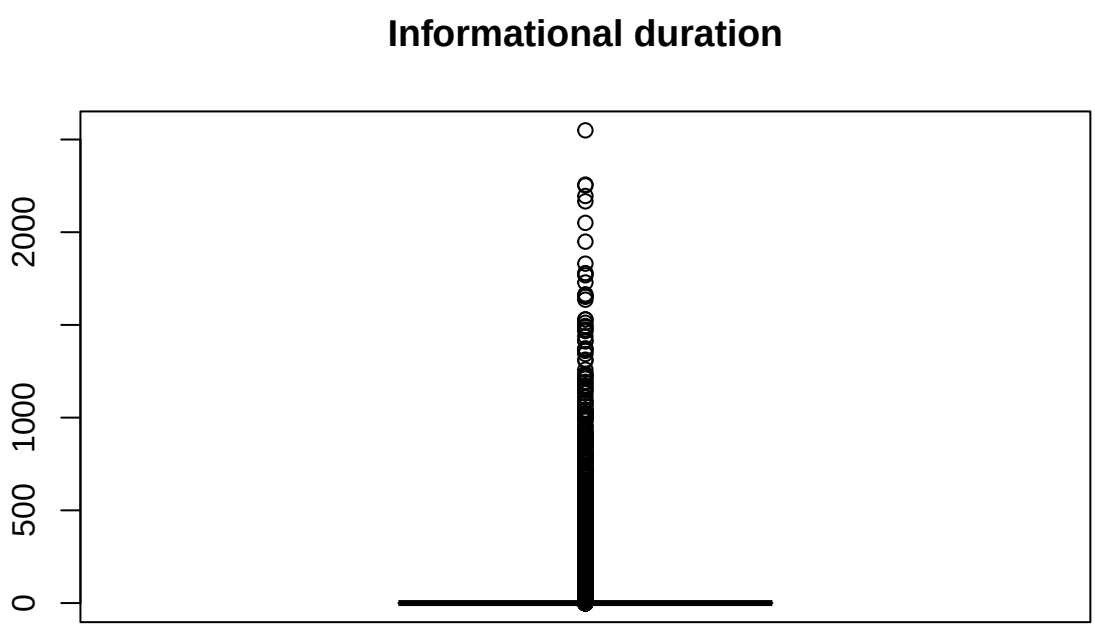

```
boxplot(df$administrative_duration,col='grey', main = 'Administrative duration Boxplot')
```



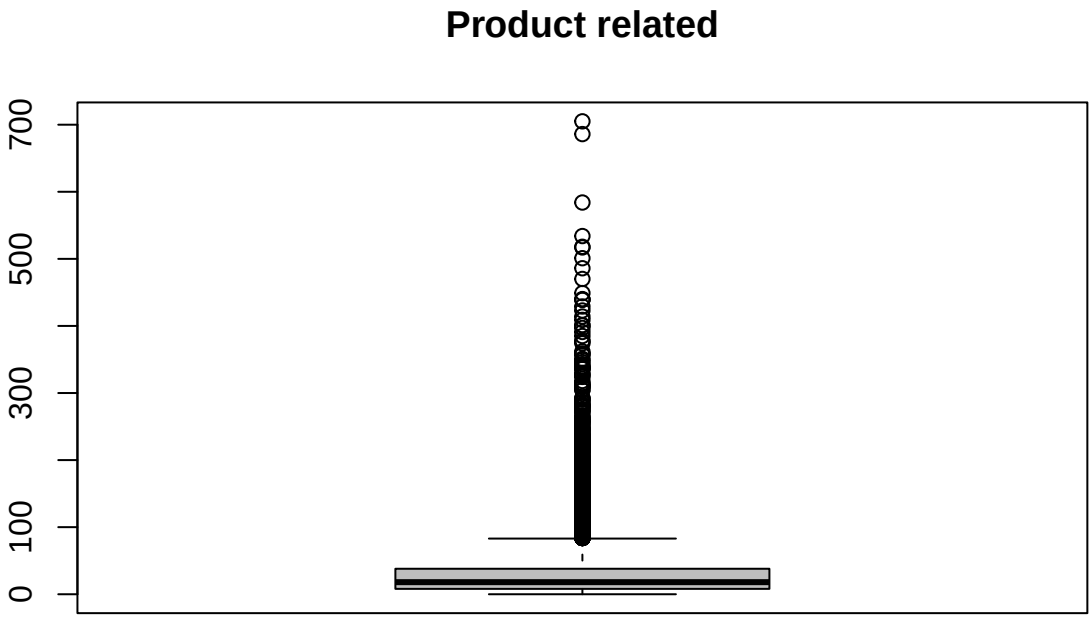

```
boxplot(df$informational,col='grey', main = 'Informational')
```



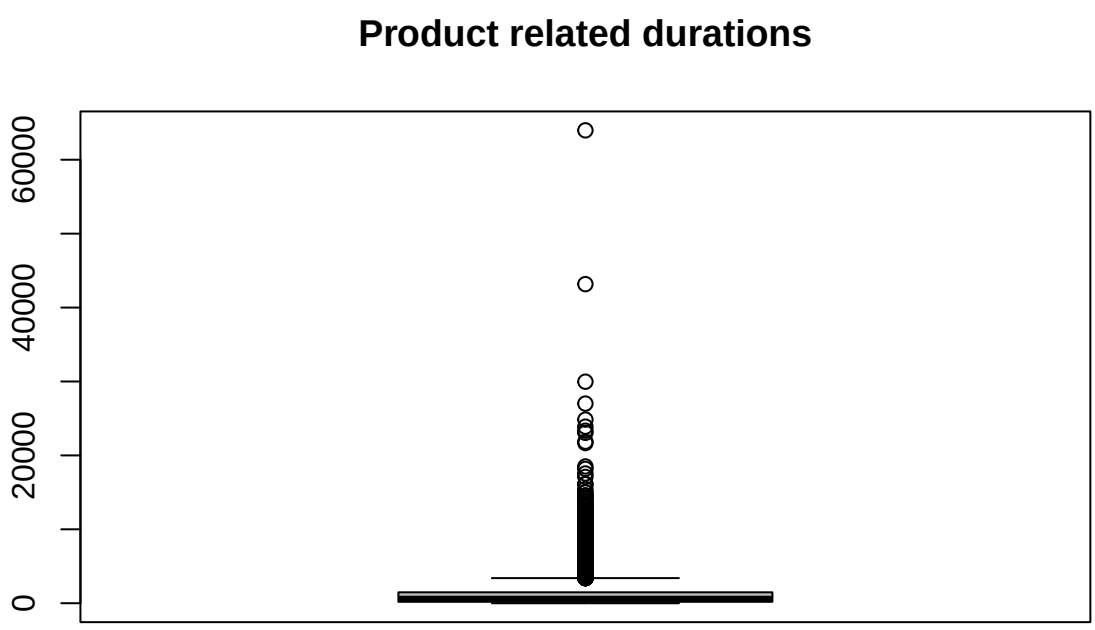

```
boxplot(df$informational_duration,col='grey', main = 'Informational duration')
```



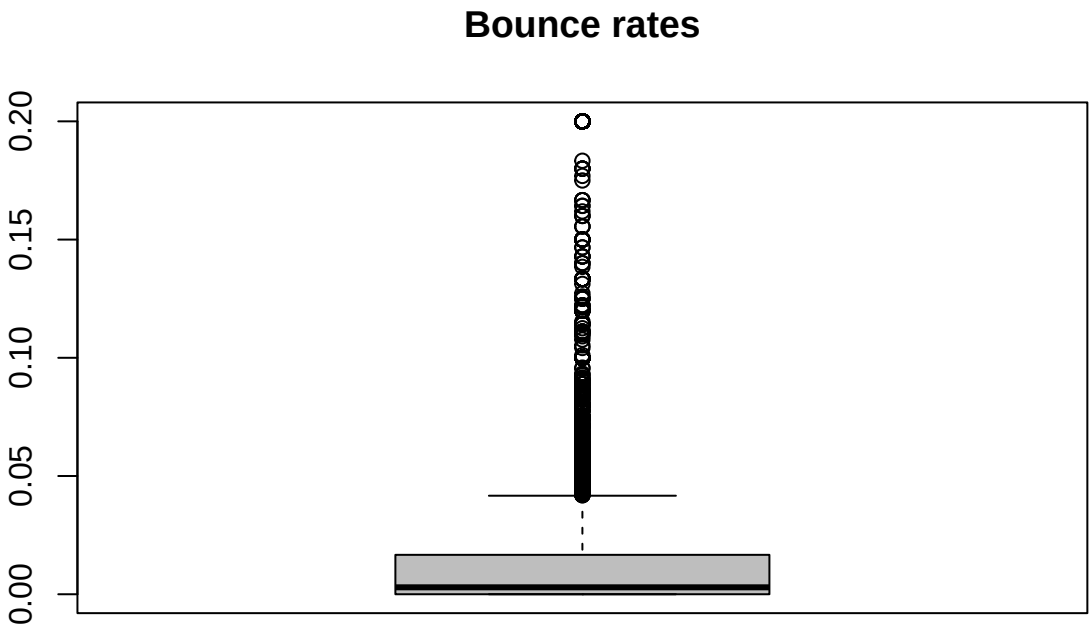

```
boxplot(df$productrelated,col='grey', main = 'Product related')
```



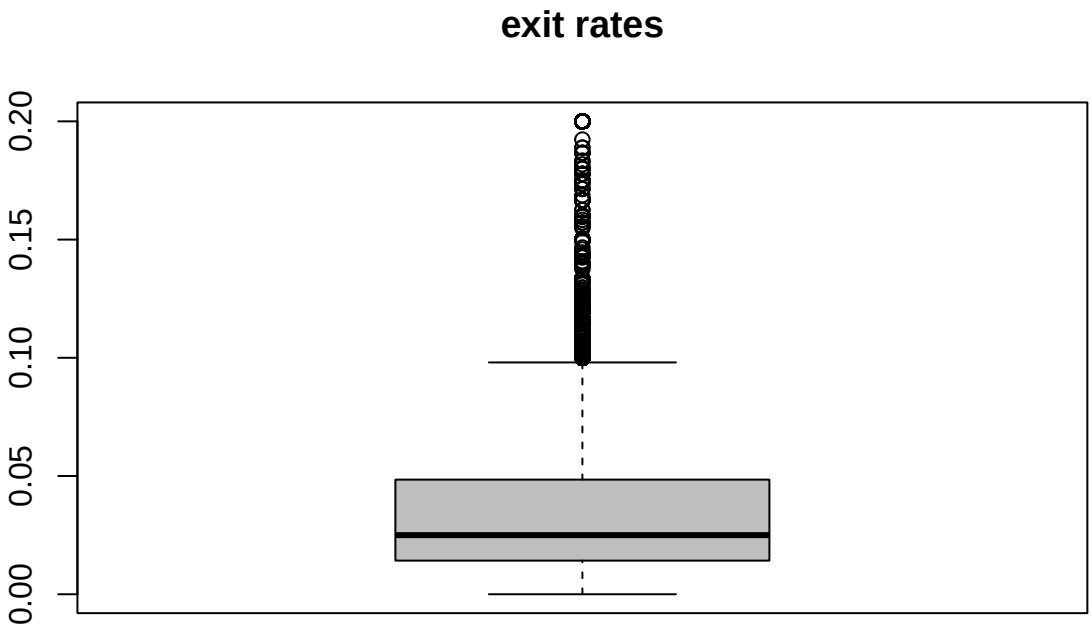

```
boxplot(df$productrelated_duration,col='grey', main = 'Product related durations')
```



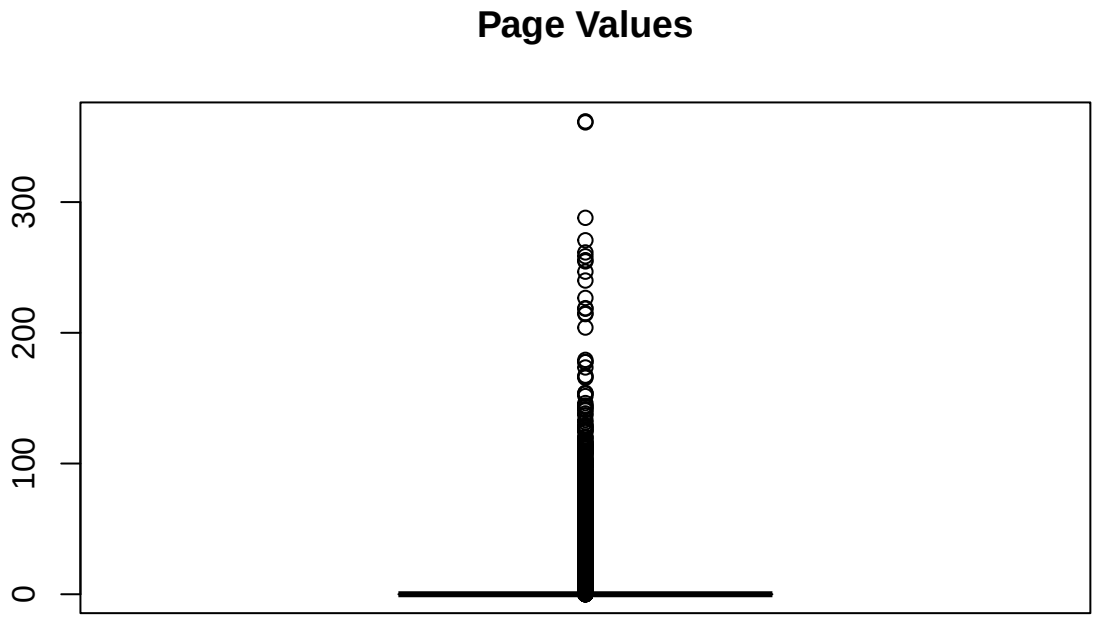

```
boxplot(df$bouncerates,col='grey', main = 'Bounce rates')
```



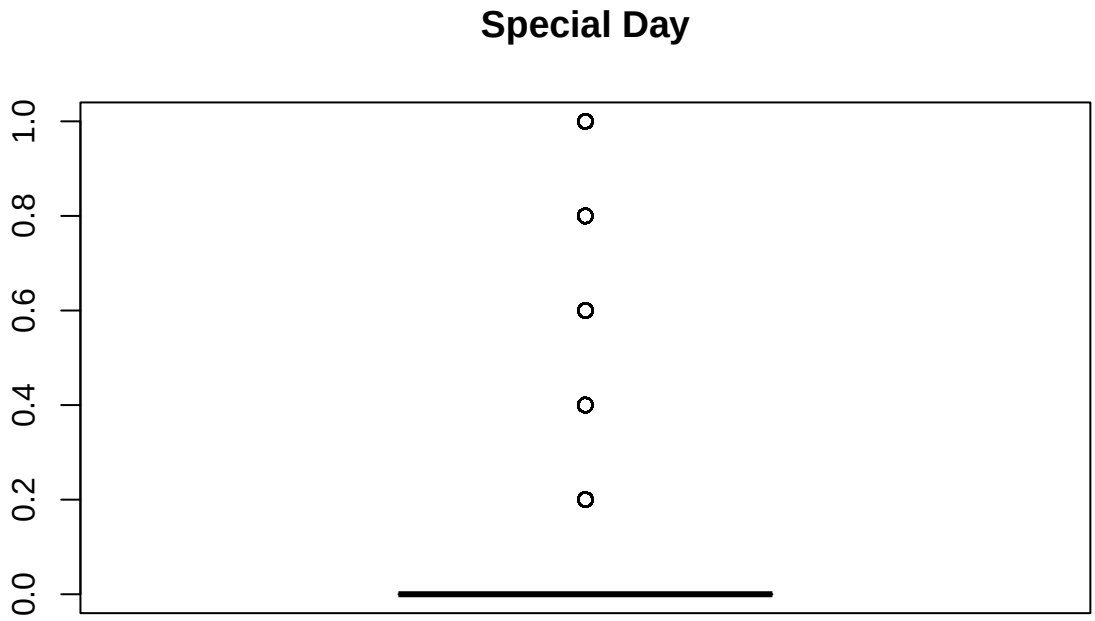

```
boxplot(df$exitrates,col='grey', main = 'exit rates')
```



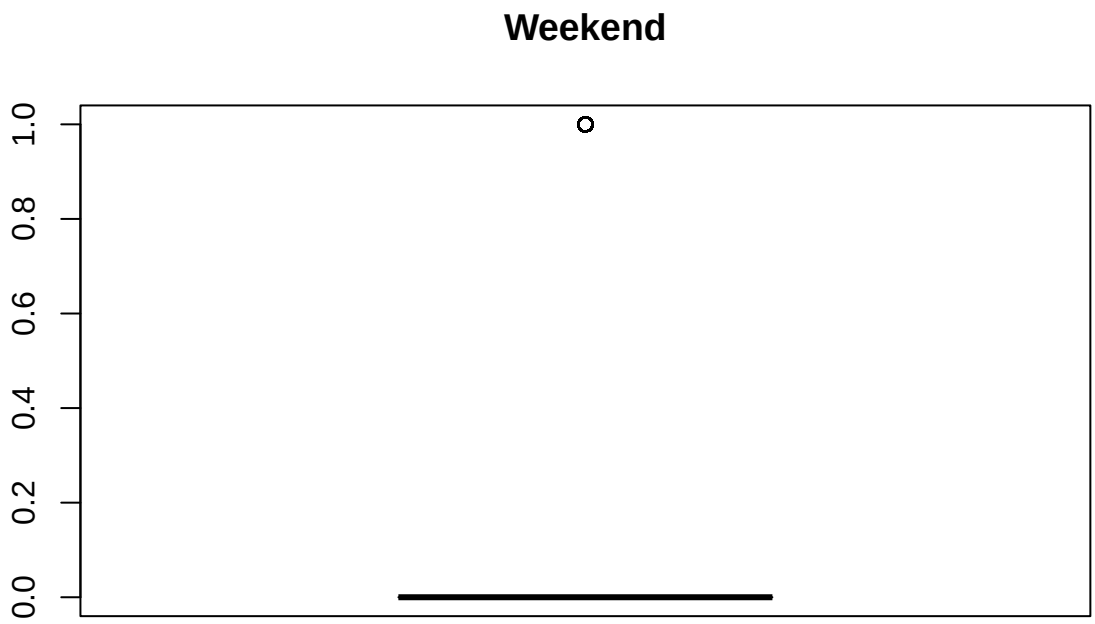

```
boxplot(df$pagevalues,col='grey', main = 'Page Values')
```



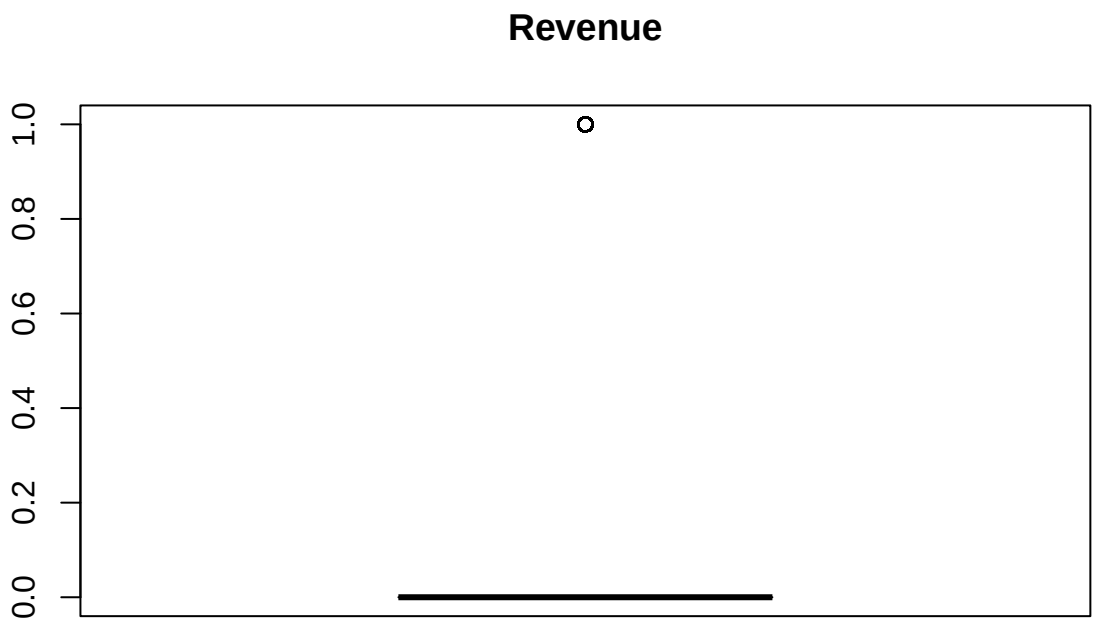

```
boxplot(df$specialday,col='grey', main = 'Special Day')
```




```
boxplot(df$weekend, col='grey', main = 'Weekend')
```




```
boxplot(df$revenue,col='grey', main = 'Revenue')
```



From the boxplots we can see that all the columns have too many outliers to remove.

Exploratory Data Analysis

Univariate Analysis

Numerical Variables Measures of dispersion

```
# Run the summary function that returns the Minimum, maximum, mean and the quantile data
summary (df)
```

```
##  administrative  administrative_duration informational
##  Min.   : 0.00   Min.   :  -1.00         Min.   : 0.0000
##  1st Qu.: 0.00   1st Qu.:   0.00         1st Qu.: 0.0000
##  Median : 1.00   Median :   9.00         Median : 0.0000
##  Mean   : 2.34   Mean   :  81.68         Mean   : 0.5088
##  3rd Qu.: 4.00   3rd Qu.:  94.75         3rd Qu.: 0.0000
##  Max.   :27.00   Max.   :3398.75         Max.   :24.0000
##  informational_duration productrelated   productrelated_duration
##  Min.   :  -1.00        Min.   :  0.00   Min.   :   -1.0
##  1st Qu.:   0.00        1st Qu.:  8.00   1st Qu.:  193.6
##  Median :   0.00        Median : 18.00   Median :  609.5
##  Mean   :  34.84        Mean   : 32.06   Mean   : 1207.5
##  3rd Qu.:   0.00        3rd Qu.: 38.00   3rd Qu.: 1477.6
##  Max.   :2549.38        Max.   :705.00   Max.   :63973.5
```

```
##   bouncerates        exitrates         pagevalues        specialday
##  Min.   :0.00000   Min.   :0.00000   Min.   :  0.000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.01422   1st Qu.:  0.000   1st Qu.:0.00000
##  Median :0.00293   Median :0.02500   Median :  0.000   Median :0.00000
##  Mean   :0.02045   Mean   :0.04150   Mean   :  5.952   Mean   :0.06197
##  3rd Qu.:0.01667   3rd Qu.:0.04848   3rd Qu.:  0.000   3rd Qu.:0.00000
##  Max.   :0.20000   Max.   :0.20000   Max.   :361.764   Max.   :1.00000
##     month           operatingsystems    browser           region
##  Length:12199       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   : 2.358   Mean   :3.153
##                     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.   : 1.000   Length:12199       Mode :logical   Mode :logical
##  1st Qu.: 2.000   Class :character   FALSE:9343      FALSE:10291
##  Median : 2.000   Mode  :character   TRUE :2856      TRUE :1908
##  Mean   : 4.075
##  3rd Qu.: 4.000
##  Max.   :20.000
```

Most of the columns are categorical and hence make it the value of the Variance will not be useful.

Variance

```
paste("The variance for Administrative Duration is" , (var(df$administrative_duration)), sep = " ")
```

```
## [1] "The variance for Administrative Duration is 31516.2503598087"
```

```
paste("The variance for Informational Duration is" , (var(df$informational_duration)), sep = " ")
```

```
## [1] "The variance for Informational Duration is 20010.5068641882"
```

```
paste("The variance for Product Related  is" , (var(df$productrelated)), sep = " ")
```

```
## [1] "The variance for Product Related  is 1989.24129586768"
```

```
paste("The variance for Product Related Duration is" , (var(df$productrelated_duration  )), sep = " ")
```

```
## [1] "The variance for Product Related Duration is 3686121.49674345"
```

```
paste("The variance for Bounce Rates  is" , (var(df$bouncerates)), sep = " ")
```

```
## [1] "The variance for Bounce Rates  is 0.0020613872549165"
```

```
paste("The variance for Exit rates is" , (var(df$exitrates)), sep = " ")
```

```
## [1] "The variance for Exit rates is 0.00213879984993362"
```

```
paste("The variance for Page Values is" , (var(df$pagevalues)), sep = " ")
```

```
## [1] "The variance for Page Values is 348.113183761479"
```

```
paste("The variance for Special day  is" , (var(df$specialday)), sep = " ")
```

```
## [1] "The variance for Special day  is 0.0398843209243294"
```

```
paste("The variance for Operating System  is" , (var(df$operatingsystems)), sep = " ")
```

```
## [1] "The variance for Operating System  is 0.822622912882059"
```

```
paste("The variance for Browser is" , (var(df$browser)), sep = " ")
```

```
## [1] "The variance for Browser is 2.92607545357061"
```

```
paste("The variance for Region  is" , (var(df$region)), sep = " ")
```

```
## [1] "The variance for Region  is 5.77171219512844"
```

```
paste("The variance for Traffic type is" , (var(df$traffictype)), sep = " ")
```

```
## [1] "The variance for Traffic type is 16.1267517123029"
```

Standard deviation

```
paste("The Standard Deviation for Administrative Duration is" , (sd(df$administrative_duration)), sep =
```

```
## [1] "The Standard Deviation for Administrative Duration is 177.528167792632"
```

```
paste("The Standard Deviation for Informational Duration is" , (sd(df$informational_duration)), sep = "
```

```
## [1] "The Standard Deviation for Informational Duration is 141.458498734393"
```

```
paste("The Standard Deviation for Product Related  is" , (sd(df$productrelated)), sep = " ")
```

```
## [1] "The Standard Deviation for Product Related  is 44.6009113793394"
```

```
paste("The Standard Deviation for Product Related Duration is" , (sd(df$productrelated_duration  )), se
```

```
## [1] "The Standard Deviation for Product Related Duration is 1919.92747174039"
```

```
paste("The Standard Deviation for Bounce Rates  is" , (sd(df$bouncerates)), sep = " ")
```

```
## [1] "The Standard Deviation for Bounce Rates  is 0.0454025027384669"
```

```
paste("The Standard Deviation for Exit rates is" , (sd(df$exitrates)), sep = " ")
```

```
## [1] "The Standard Deviation for Exit rates is 0.046247160452655"
```

```
paste("The Standard Deviation for Page Values is" , (sd(df$pagevalues)), sep = " ")
```

```
## [1] "The Standard Deviation for Page Values is 18.6577915027872"
```

```
paste("The Standard Deviation for Special day  is" , (sd(df$specialday)), sep = " ")
```

```
## [1] "The Standard Deviation for Special day  is 0.199710592919678"
```

```
paste("The Standard Deviation for Operating System  is" , (sd(df$operatingsystems)), sep = " ")
```

```
## [1] "The Standard Deviation for Operating System  is 0.906985618894842"
```

```
paste("The Standard Deviation for Browser is" , (sd(df$browser)), sep = " ")
```

```
## [1] "The Standard Deviation for Browser is 1.71057752047974"
```

```
paste("The Standard Deviation for Region  is" , (sd(df$region)), sep = " ")
```

```
## [1] "The Standard Deviation for Region  is 2.40243880153656"
```

```
paste("The Standard Deviation for Traffic type is" , (sd(df$traffictype)), sep = " ")
```

```
## [1] "The Standard Deviation for Traffic type is 4.01581270881785"
```

Kurtosis

```
paste("The kurtosis for Administrative Duration is" , (kurtosis(df$administrative_duration)), sep = " "
```

```
## [1] "The kurtosis for Administrative Duration is 50.085181158405"
```

```
paste("The kurtosis for Informational Duration is" , (kurtosis(df$informational_duration)), sep = " ")
```

```
## [1] "The kurtosis for Informational Duration is 75.4512246461555"
```

```
paste("The kurtosis for Product Related  is" , (kurtosis(df$productrelated)), sep = " ")
```

```
## [1] "The kurtosis for Product Related  is 31.0434452164828"
```

```
paste("The kurtosis for Product Related Duration is" , (kurtosis(df$productrelated_duration  )), sep =
```

```
## [1] "The kurtosis for Product Related Duration is 136.56790690015"
```

```
paste("The kurtosis for Bounce Rates  is" , (kurtosis(df$bouncerates)), sep = " ")
```

```
## [1] "The kurtosis for Bounce Rates  is 9.25305503204776"
```

```
paste("The kurtosis for Exit rates is" , (kurtosis(df$exitrates)), sep = " ")
```

```
## [1] "The kurtosis for Exit rates is 4.62300253695039"
```

```
paste("The kurtosis for Page Values is" , (kurtosis(df$pagevalues)), sep = " ")
```

```
## [1] "The kurtosis for Page Values is 64.9291691002094"
```

```
paste("The kurtosis for Special day  is" , (kurtosis(df$specialday)), sep = " ")
```

```
## [1] "The kurtosis for Special day  is 9.78395765970814"
```

```
paste("The kurtosis for Operating System  is" , (kurtosis(df$operatingsystems)), sep = " ")
```

```
## [1] "The kurtosis for Operating System  is 10.266692831375"
```

```
paste("The kurtosis for Browser is" , (kurtosis(df$browser)), sep = " ")
```

```
## [1] "The kurtosis for Browser is 12.534043219127"
```

```
paste("The kurtosis for Region  is" , (kurtosis(df$region)), sep = " ")
```

```
## [1] "The kurtosis for Region  is -0.160270347887721"
```

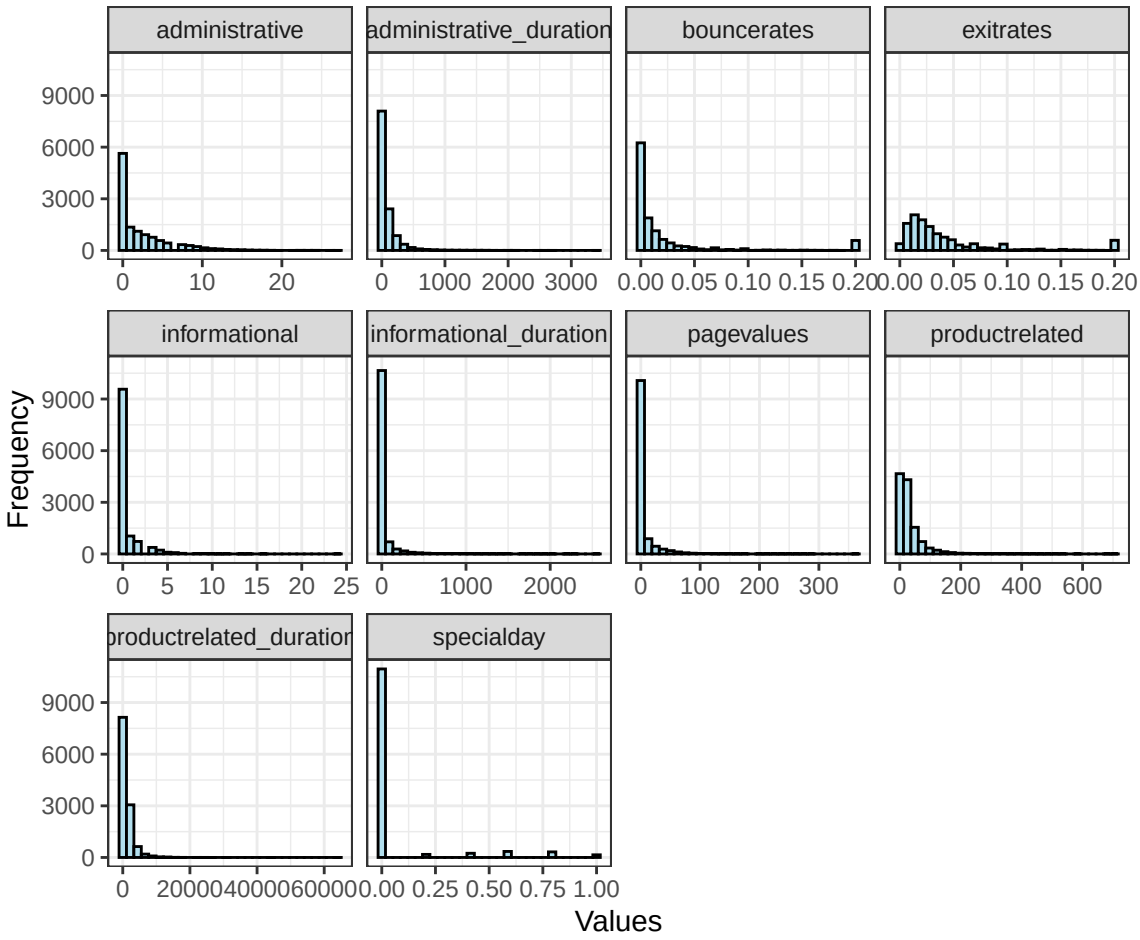
```
paste("The kurtosis for Traffic type is" , (kurtosis(df$traffictype)), sep = " ")
```

```
## [1] "The kurtosis for Traffic type is 3.46506739912327"
```

Histograms of the different columns

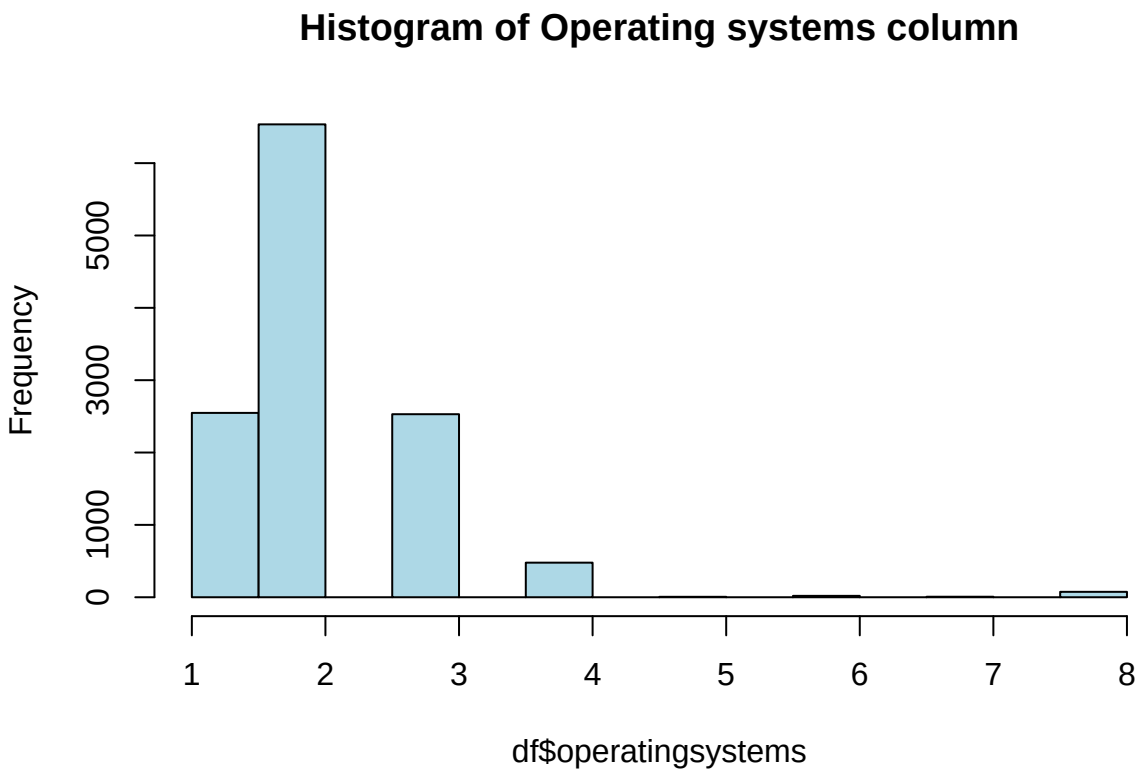
```
df %>%
  gather(attributes, value, 1:10) %>%
  ggplot(aes(x = value)) +
  geom_histogram(fill = 'lightblue2', color = 'black') +
  facet_wrap(~attributes, scales = 'free_x') +
  labs(x="Values", y="Frequency") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



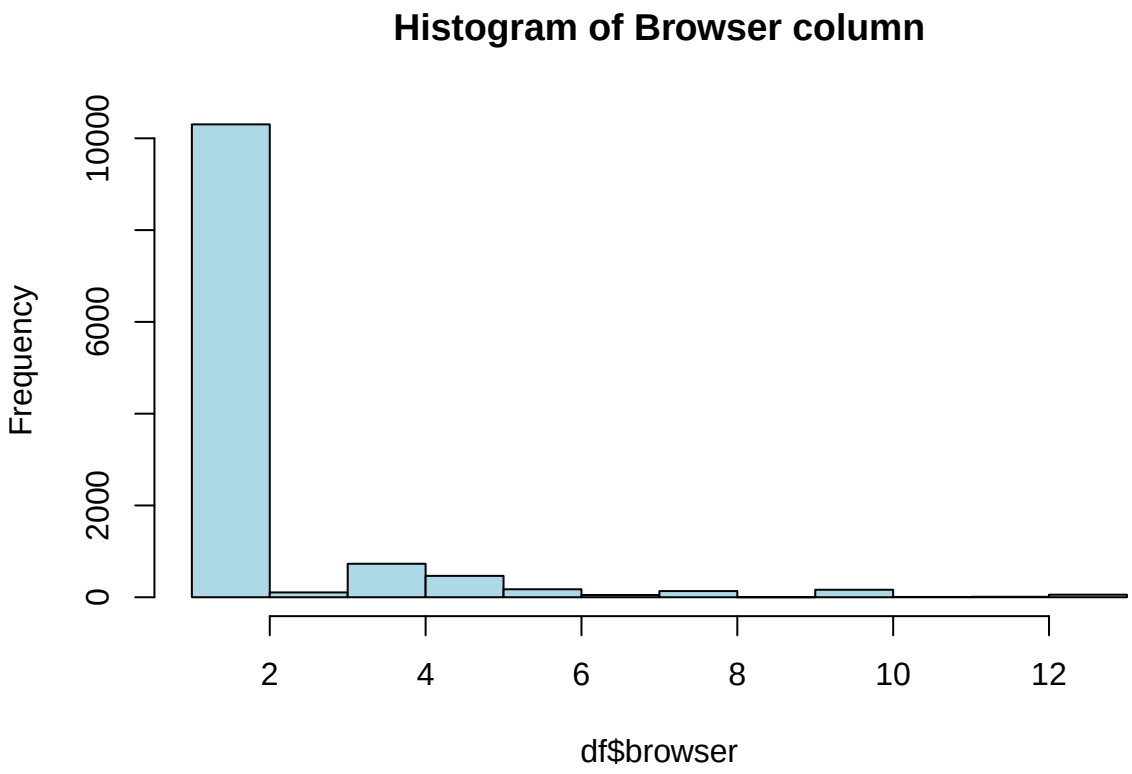
All columns are rightly skewed with the majority of data at the 0 mark. ‘

```
# Histogram for operating systemms
hist(df$operatingsystems, main = 'Histogram of Operating systems column',col="lightblue")
```



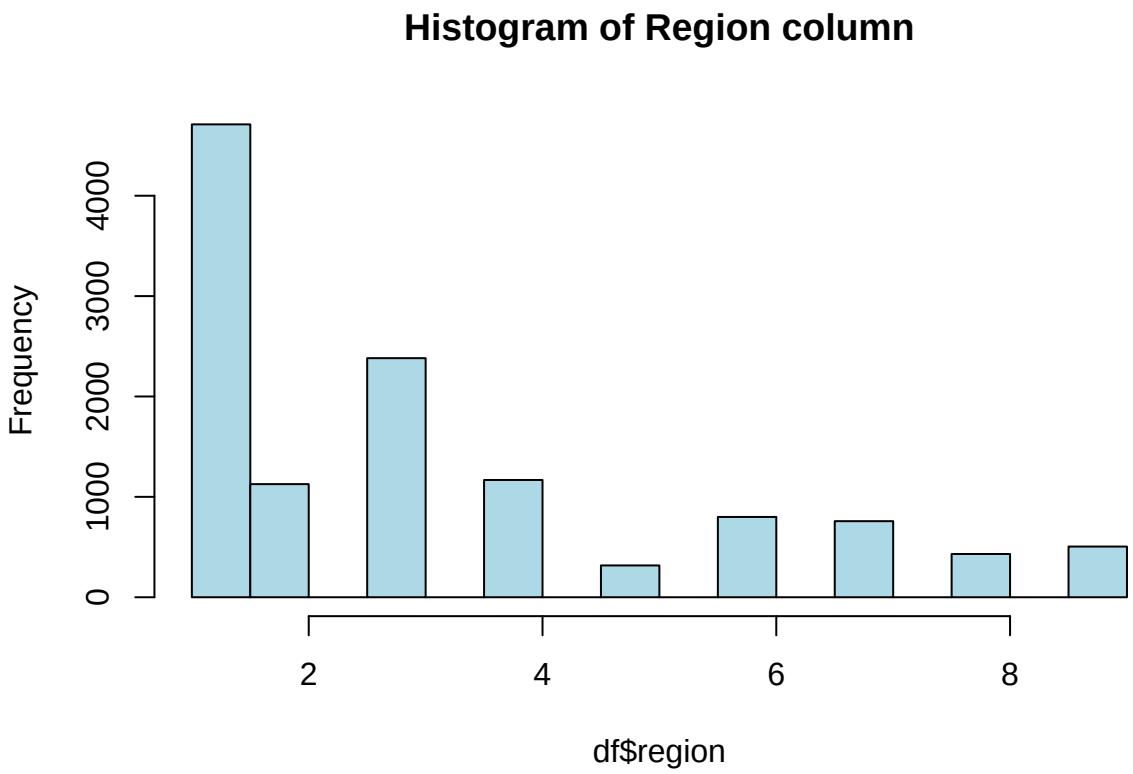
Most people use an operating system of 1

```
# Histogram for operating systemms

hist(df$browser, main = 'Histogram of Browser column',col="lightblue")
```



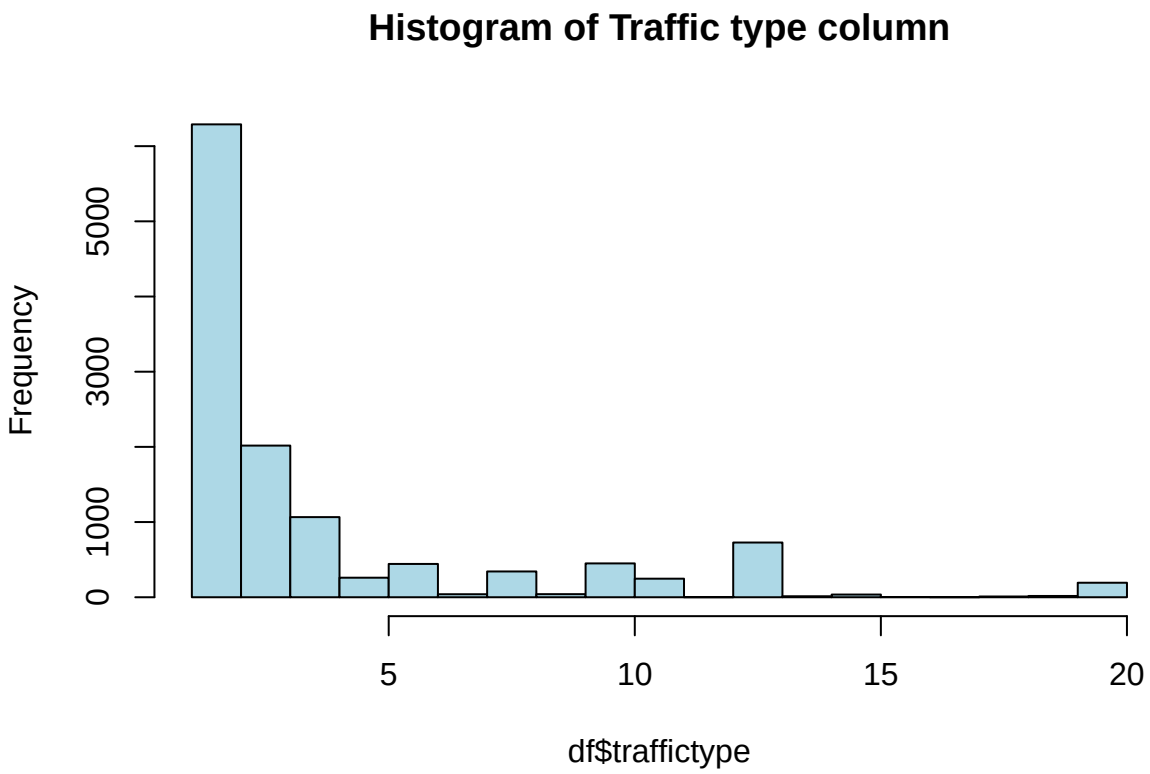
Majority of people visiting the site use browser 1

```
# Histogram for operating systemms

hist(df$region, main = 'Histogram of Region column',col="lightblue")
```



Majority of people visiting the site are from region 1

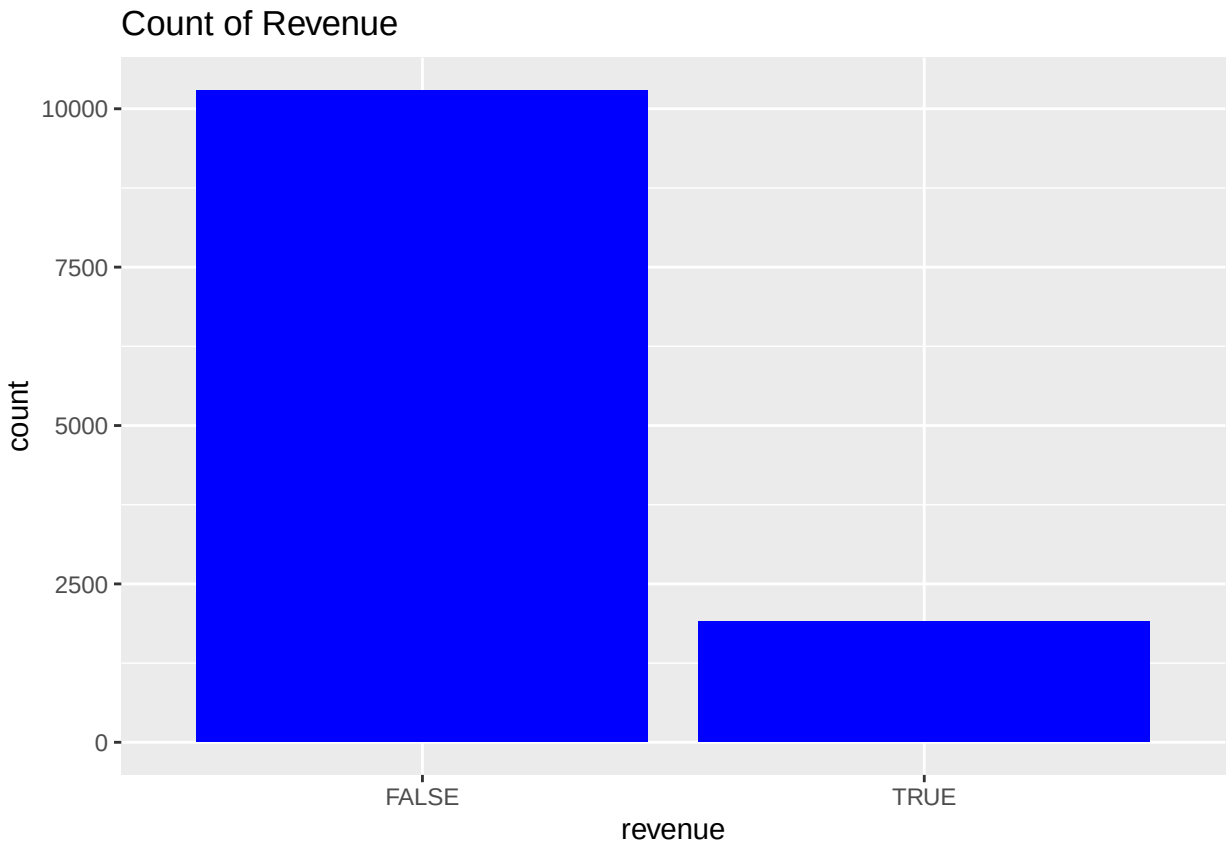
```
# Histogram for operating systemms

hist(df$traffictype, main = 'Histogram of Traffic type column',col="lightblue")
```



The majority of people visiting the use do so with traffic type 1

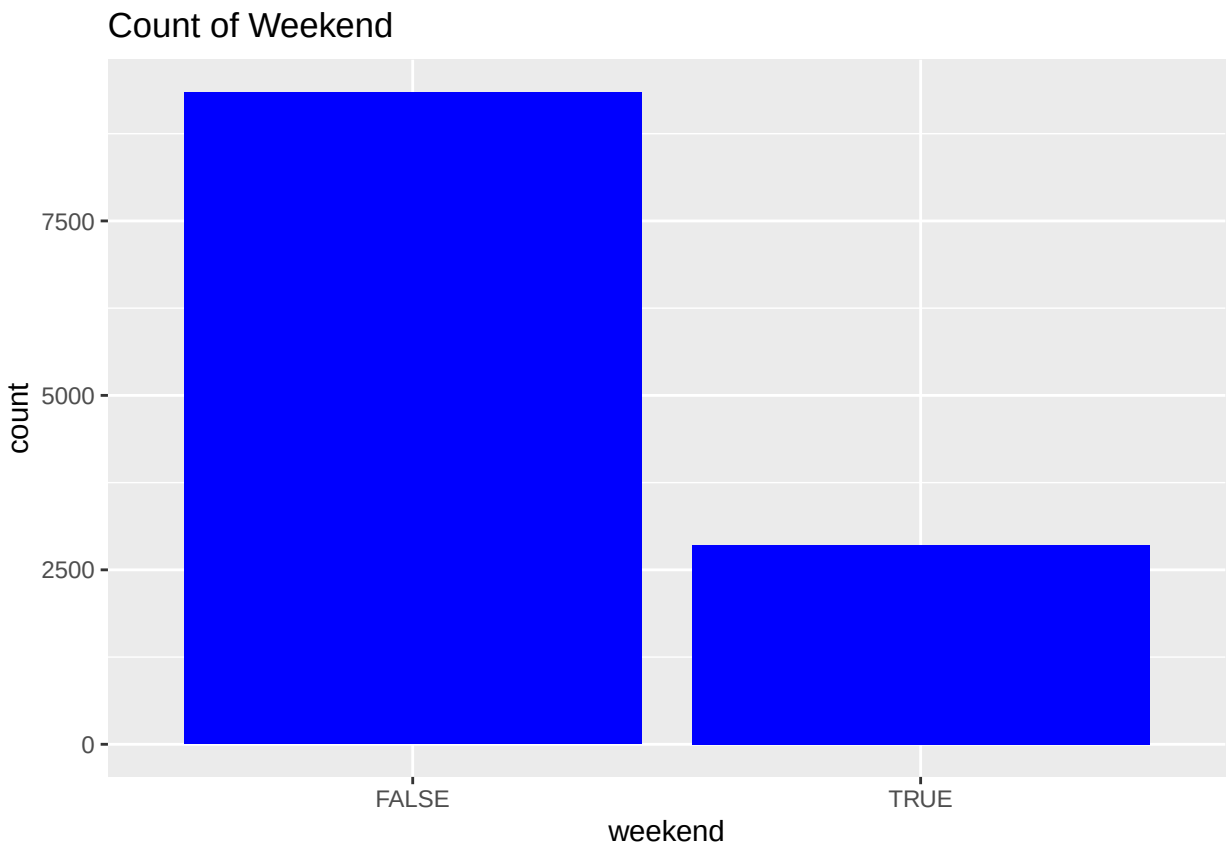
Categorical Variables

```
# Countplot for Revenue column
# This is to see the number of people that added revenue.

ggplot(df, aes(x = revenue)) +
  geom_bar(fill="blue") +  ggtitle('Count of Revenue')
```



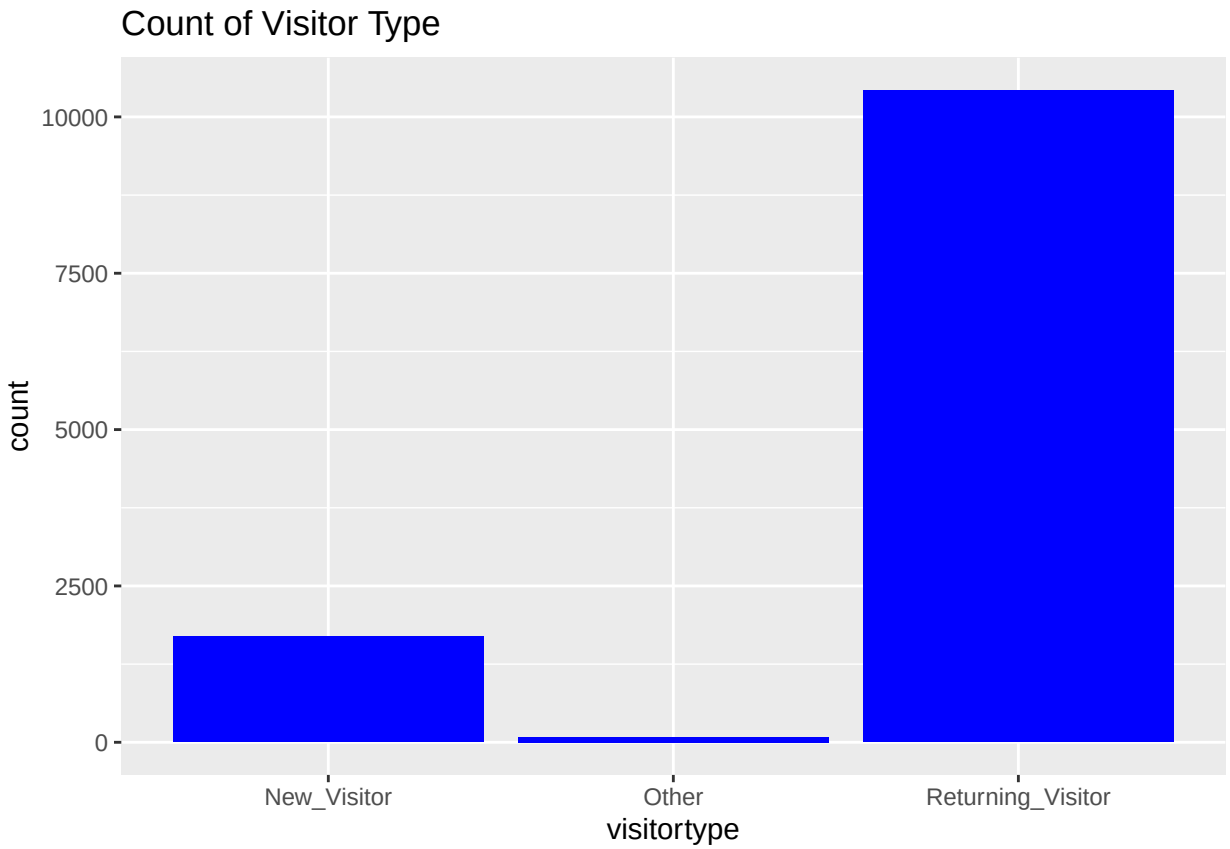
Most individuals who visited the site did not buy any merchandise

```
# Countplot for Weekend column
# This is to see how the visitors to the site are distributed between weekday and aweekend
ggplot(df, aes(x = weekend)) +
  geom_bar(fill="blue") +  ggtitle('Count of Weekend')
```



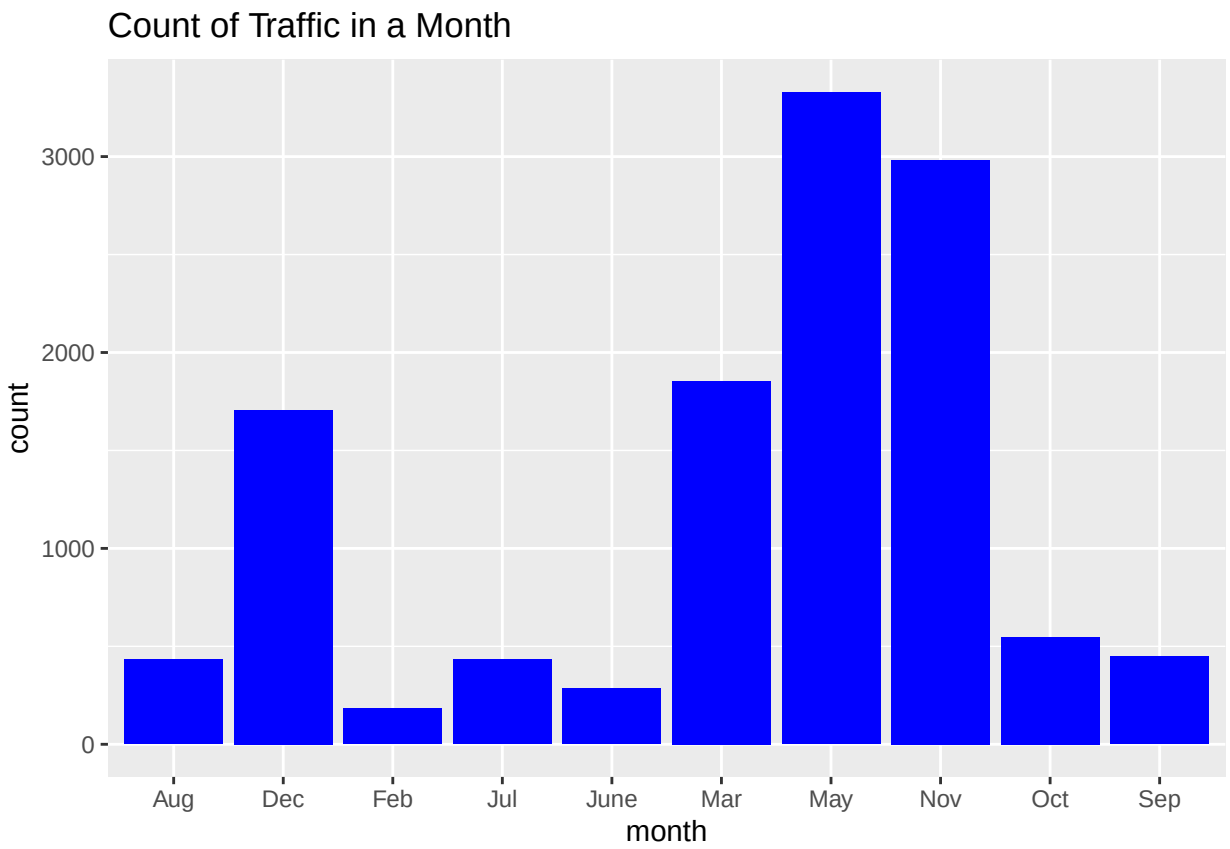
It is safe to say that visitors to the site visited on weekdays and weekends

```
# Countplot for Weekend column
# This is to see who the visitors to the site are
ggplot(df, aes(x = visitortype)) +
  geom_bar(fill="blue") +  ggtitle('Count of Visitor Type')
```



Most visitors to the site were returning visitors. We can infer that they had bought from the site and were satisfied with the products.

```
# Countplot for month column
# This is to see the month with more traffic
ggplot(df, aes(x = month)) +
  geom_bar(fill="blue") +  ggtitle('Count of Traffic in a Month')
```

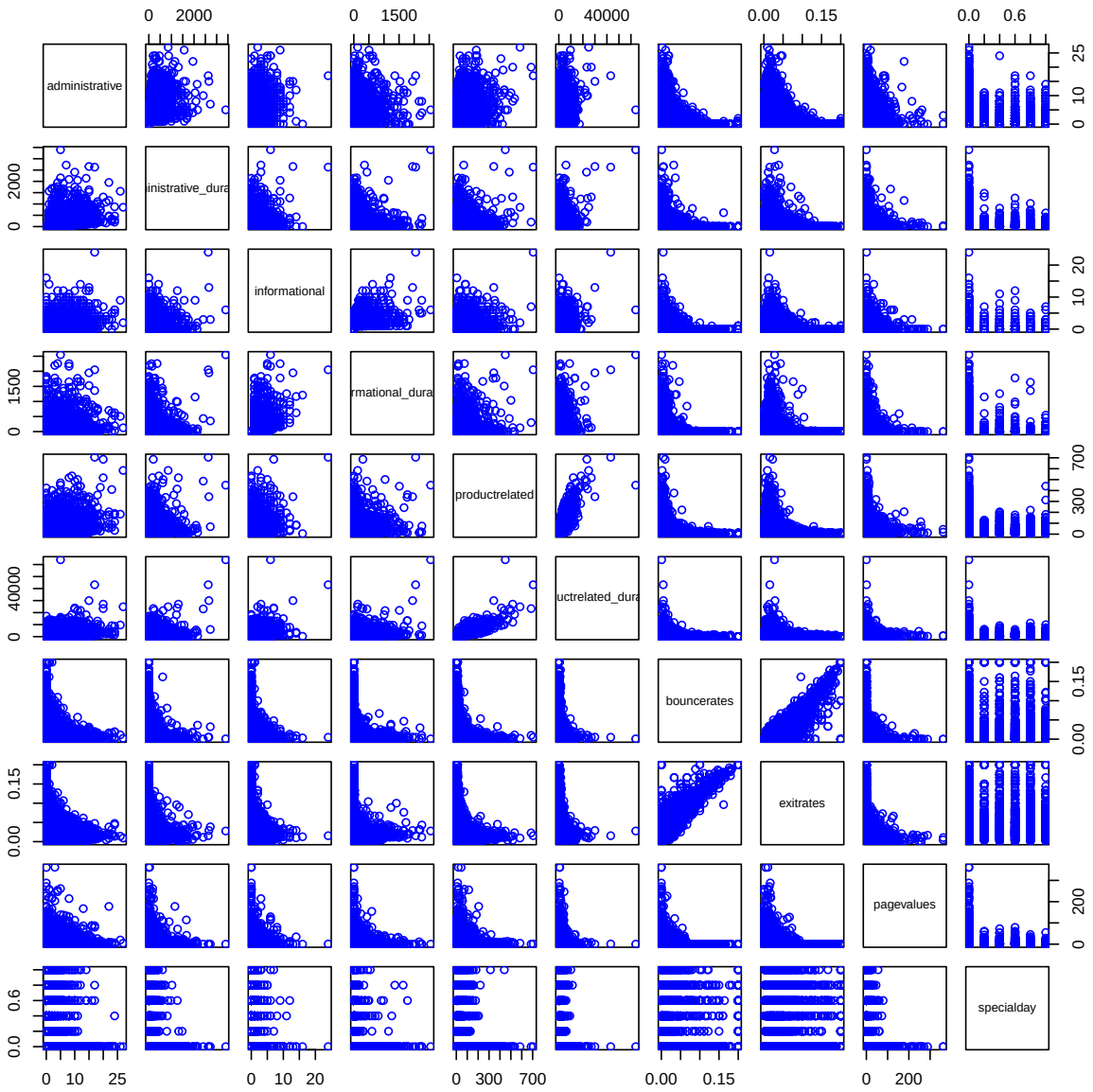


The month of May has the most traffic followed by November while the least is February. The dataset does not have 2 months January and April.

Bivariate Analysis

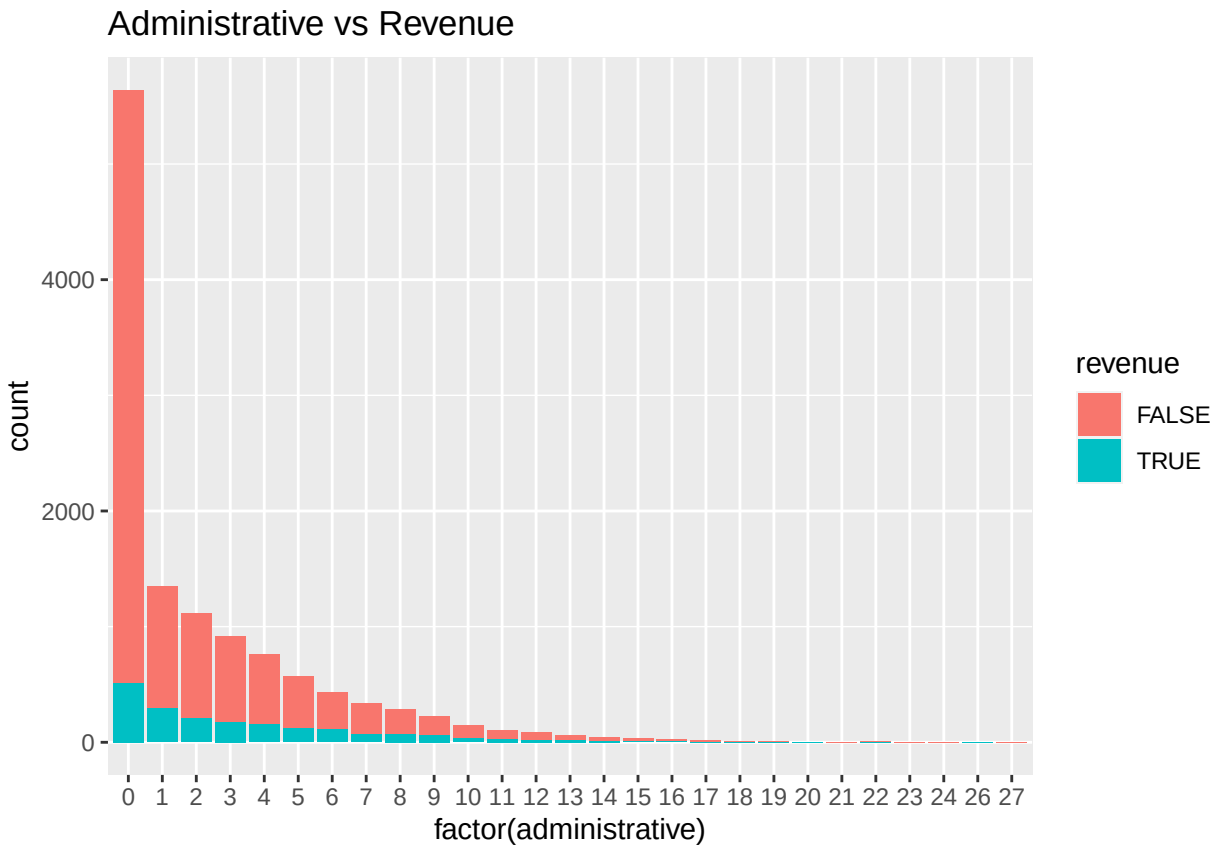
Numerical columns

```
# The following is a pair plot of all the number variables
pairs(df[,1:10], col = "blue")
```



Administrative vs Revenue

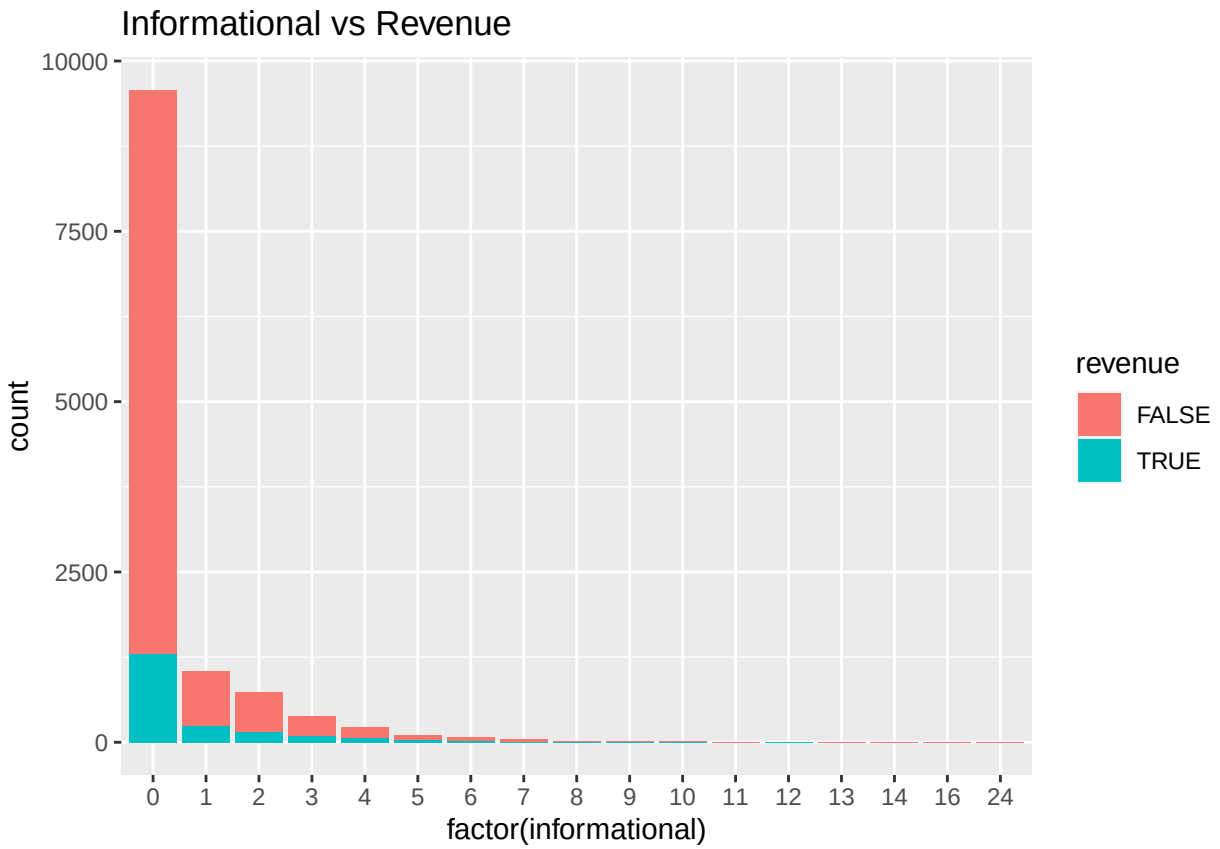
```
# Check how administrative and revenue relate in this dataset
ggplot(df, aes(x = factor(administrative), fill = revenue)) +
  geom_bar() +  ggtitle('Administrative vs Revenue')
```



We can see that a lot individuals visited administrative page 0 and it has the best turnover for revenue. The further back the administrative page the less an individual will buy something from the site.

Informational vs Revenue

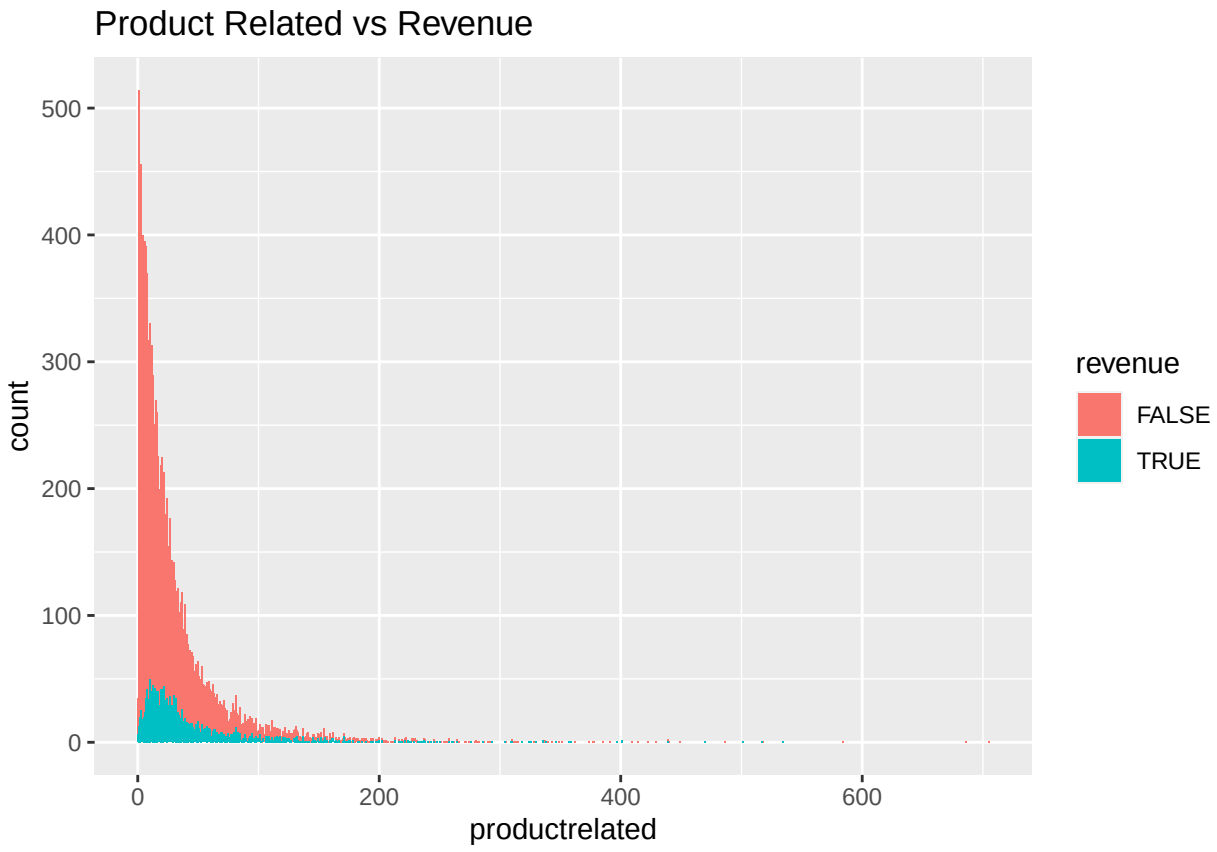
```
# Check how informational and revenue relate in this dataset

ggplot(df, aes(x = factor(informational), fill = revenue)) +
  geom_bar() +  ggtitle('Informational vs Revenue')
```



We can see that majority of the individuals visited informational page 0 and it has the best turnover for revenue. The further back the informational page the less an individual will buy something from the site.

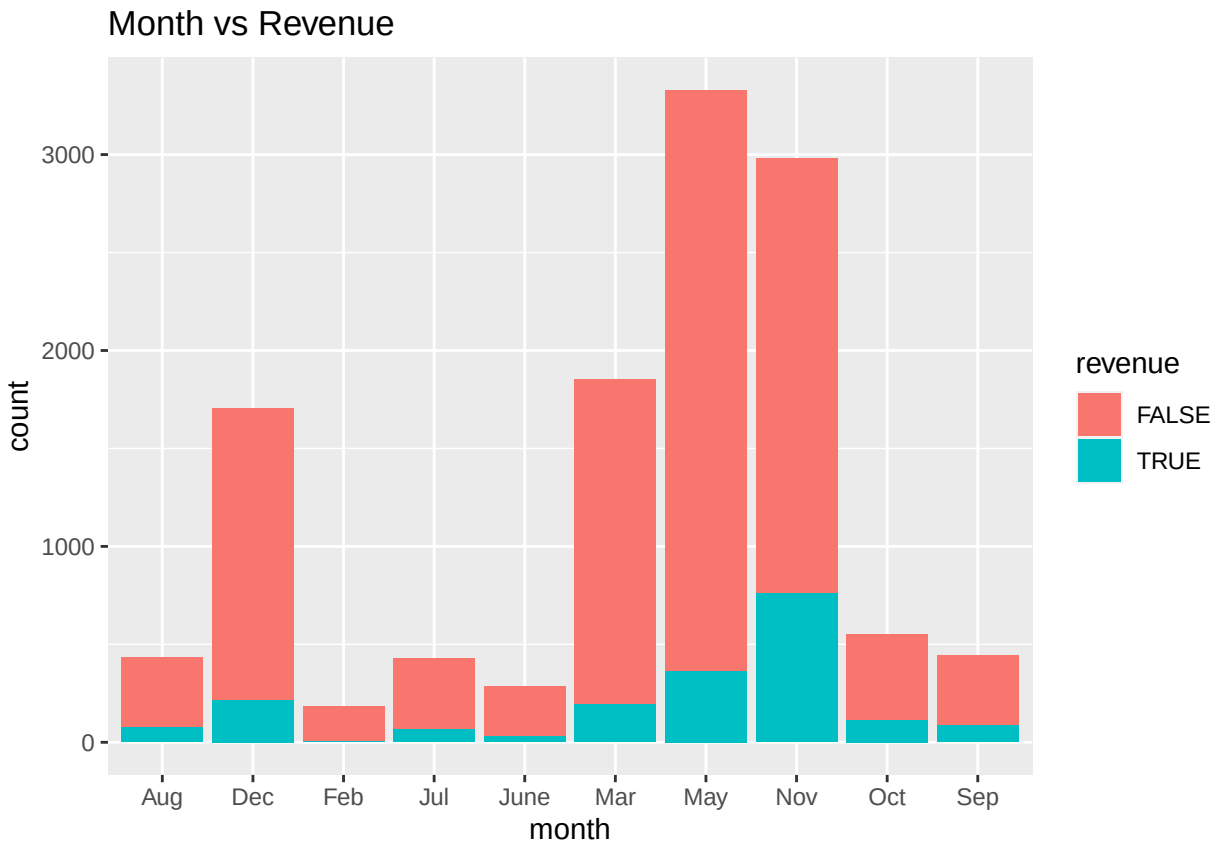
Product Related vs Revenue

```
# Check how Product Related and revenue relate in this dataset
ggplot(df, aes(x = productrelated, fill = revenue)) +
  geom_bar() +  ggtitle('Product Related vs Revenue')
```



Month vs Revenue

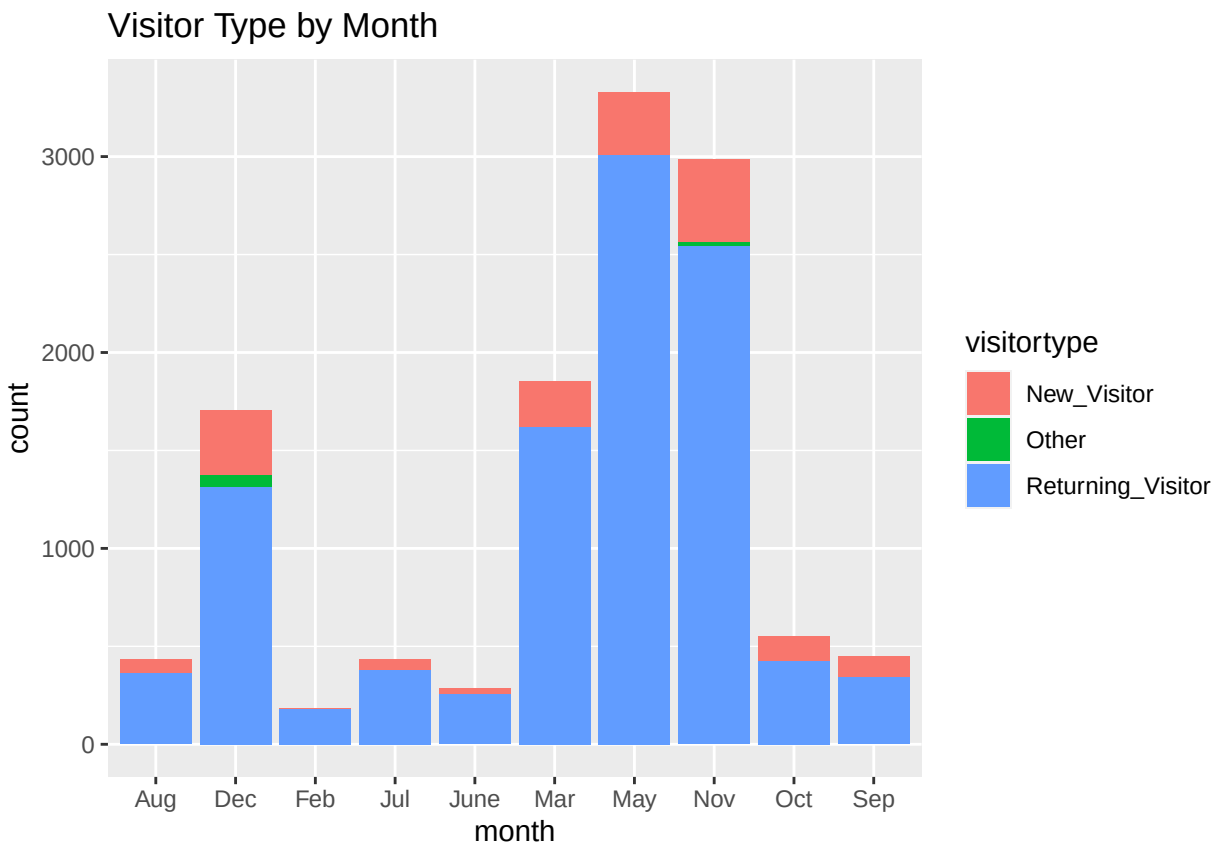
```
# Check how Month and revenue relate in this dataset
ggplot(df, aes(x = month, fill = revenue)) +
  geom_bar() +  ggtitle('Month vs Revenue')
```



Although May has the highest number of visitors, November has the highest revenue followed by May. This can be because black Friday is celebrated in November which led to a higher income.

Visitor Type by Month

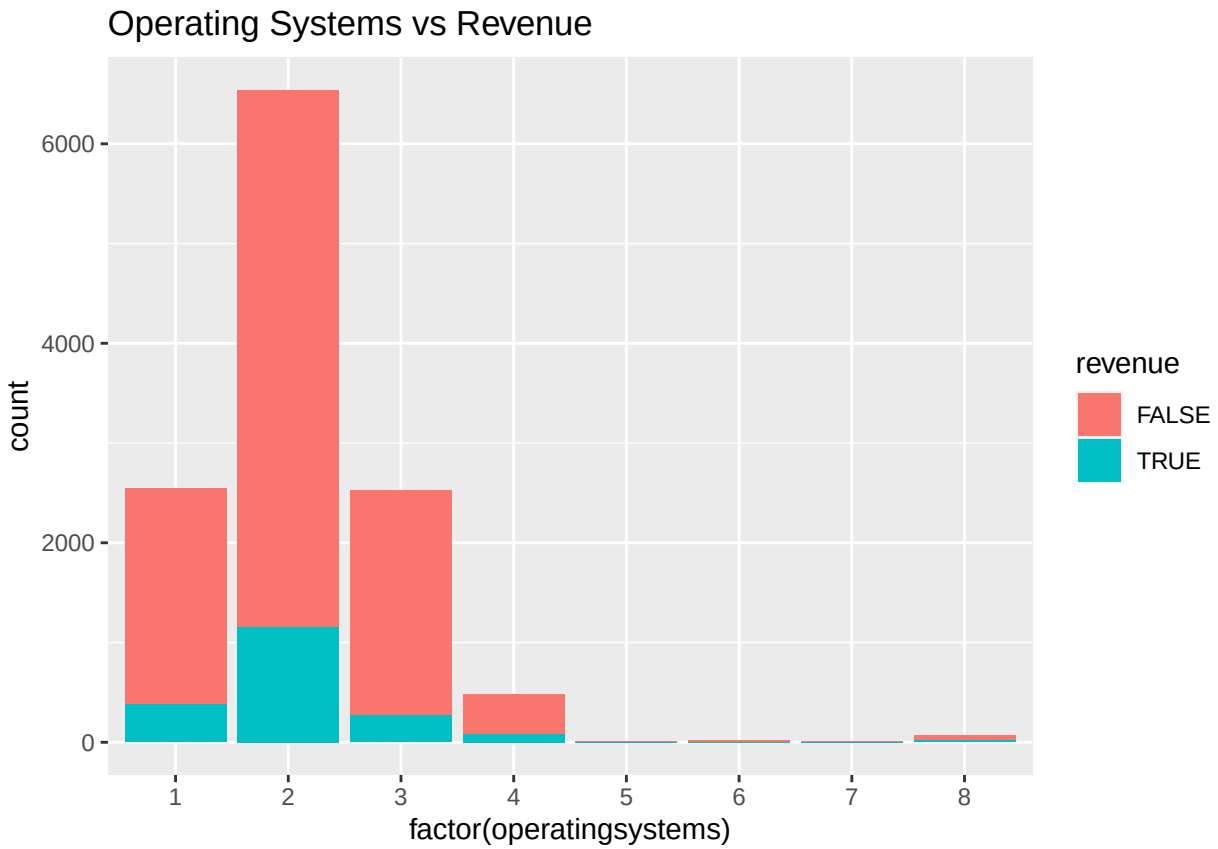
```
# Lets look at the months which have a lot of customers. and the type of customers.
df %>%
    ggplot(aes(month)) +
    geom_bar(aes(fill = visitortype))+
    labs(title = "Visitor Type by Month")
```



It can be observed that the months with the most number of customers are: May, November, March and December. This can be as a result of the various celebrations in these months. For instance, in march, most people are usually shopping foe Easter products. In November, the black Friday sales generate a lot of traffic for the site. They are also shopping for Thanksgiving, Christmas and Hanukkah. It is also clear that the site gets a lot of returning visitors. New visitors are usually found in the months that there is an increase in customer traffic.

Operating Systems vs Revenue

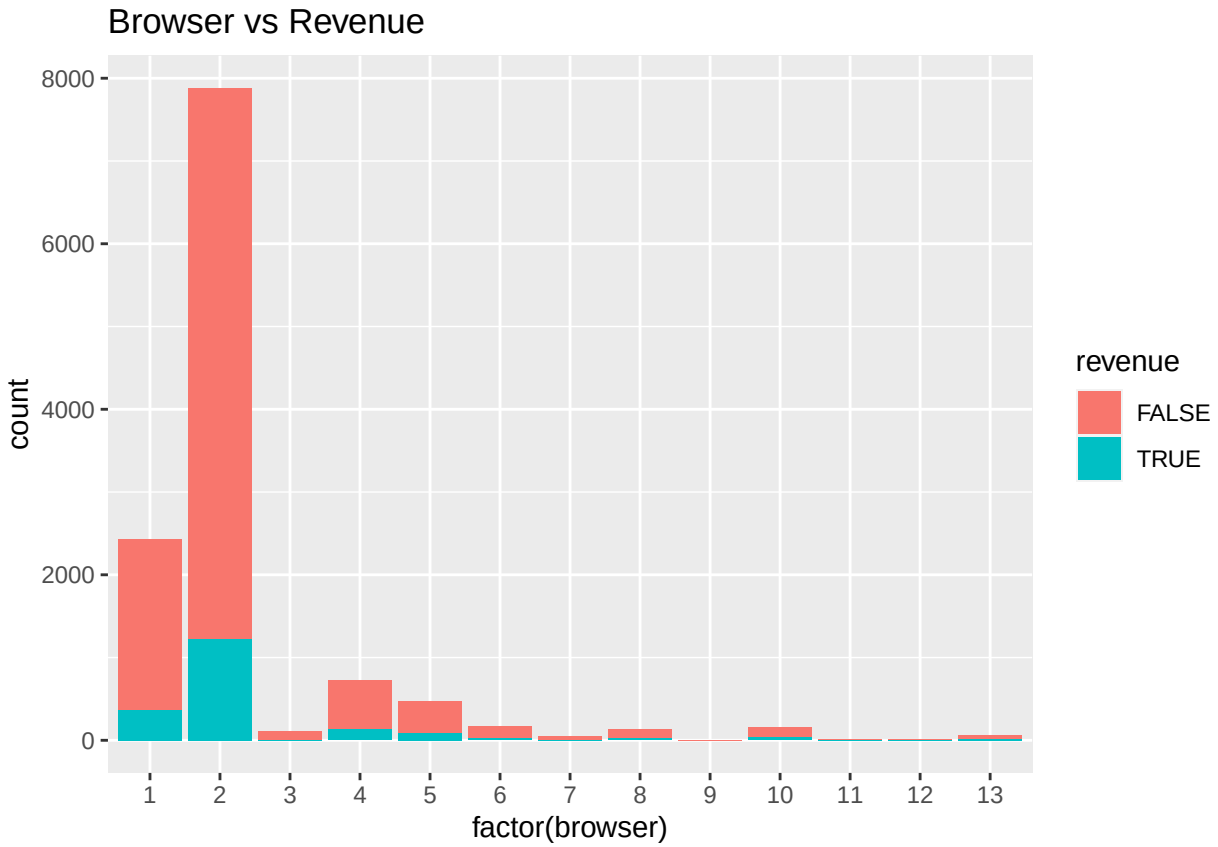
```
# Check how Operating system and revenue relate in this dataset
ggplot(df, aes(x = factor(operatingsystems), fill = revenue)) +
  geom_bar() +  ggtitle('Operating Systems vs Revenue')
```



Majority of the people visiting the site use operating system 2 which also has the highest revenue return

Browser vs Revenue

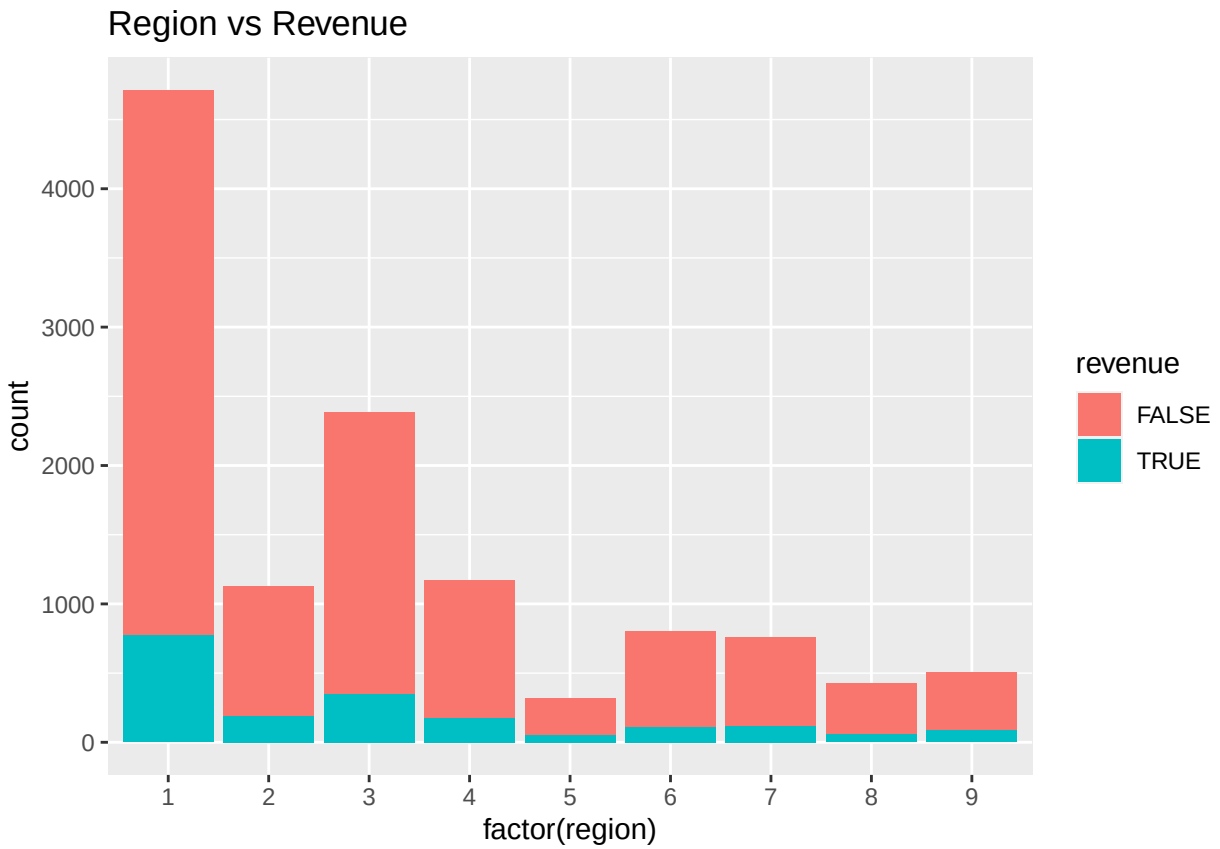
```
# Check how Browser and revenue relate in this dataset

ggplot(df, aes(x = factor(browser), fill = revenue)) +
  geom_bar() +  ggtitle('Browser vs Revenue')
```



Majority of people visiting the site use browser 2 which also has the highest revenue return

Region vs Revenue

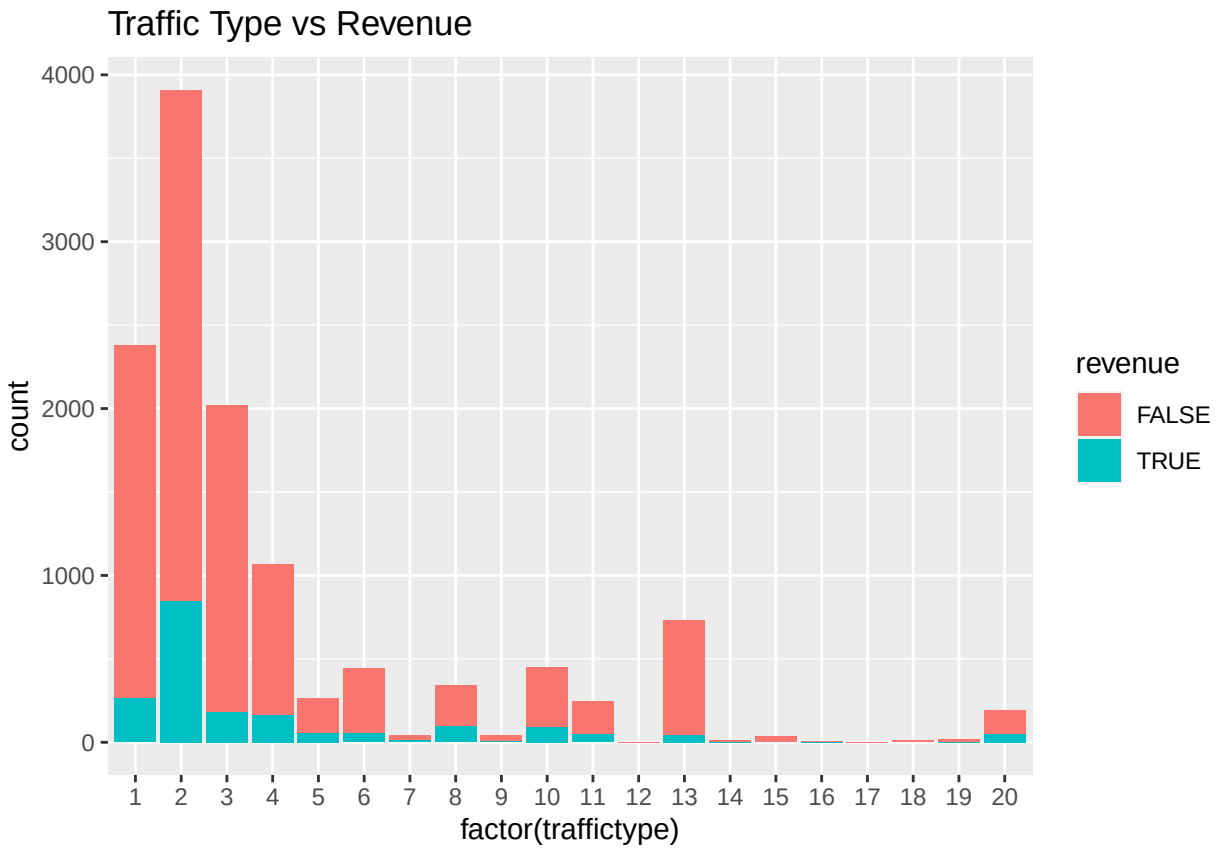
```
# Check how region and revenue relate in this dataset

ggplot(df, aes(x = factor(region), fill = revenue)) +
  geom_bar() +  ggtitle('Region vs Revenue')
```



Majority of individuals visit from region1 and has the highest revenue return (Region 1 — United States, Canada, Bermuda, Caribbean, U.S. territories) followed by Region 3 (Region 3 — Southeast Asia, Hong Kong, South Korea, Macau, Taiwan)

Traffic Type vs Revenue

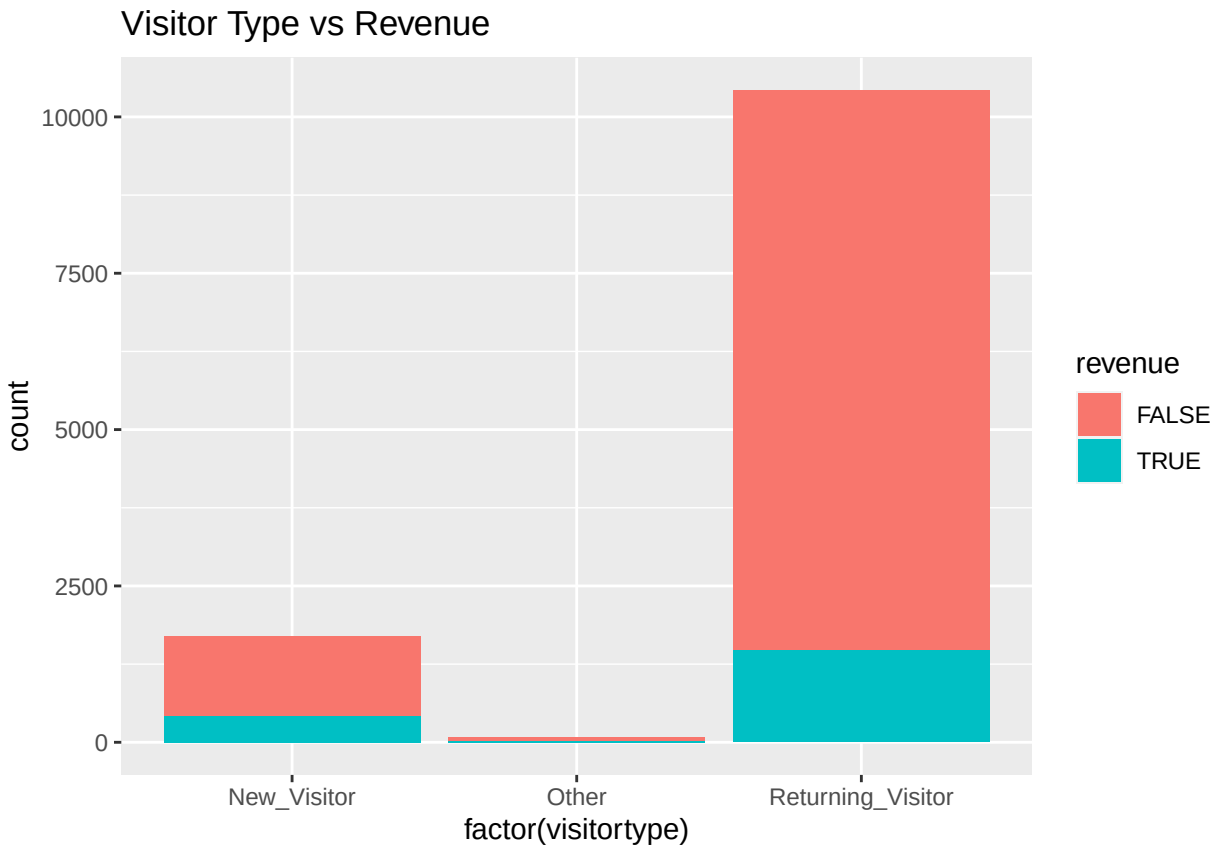
```
# Check how traffic type and revenue relate in this dataset

ggplot(df, aes(x = factor(traffictype), fill = revenue)) +
  geom_bar() +  ggtitle('Traffic Type vs Revenue')
```



Traffic type 2 refers a lot of individuals to the site and also has the highest revenue return followed by traffic type 1. So the company should focus on taffic type 2 and 1.

Visitor Type vs Revenue

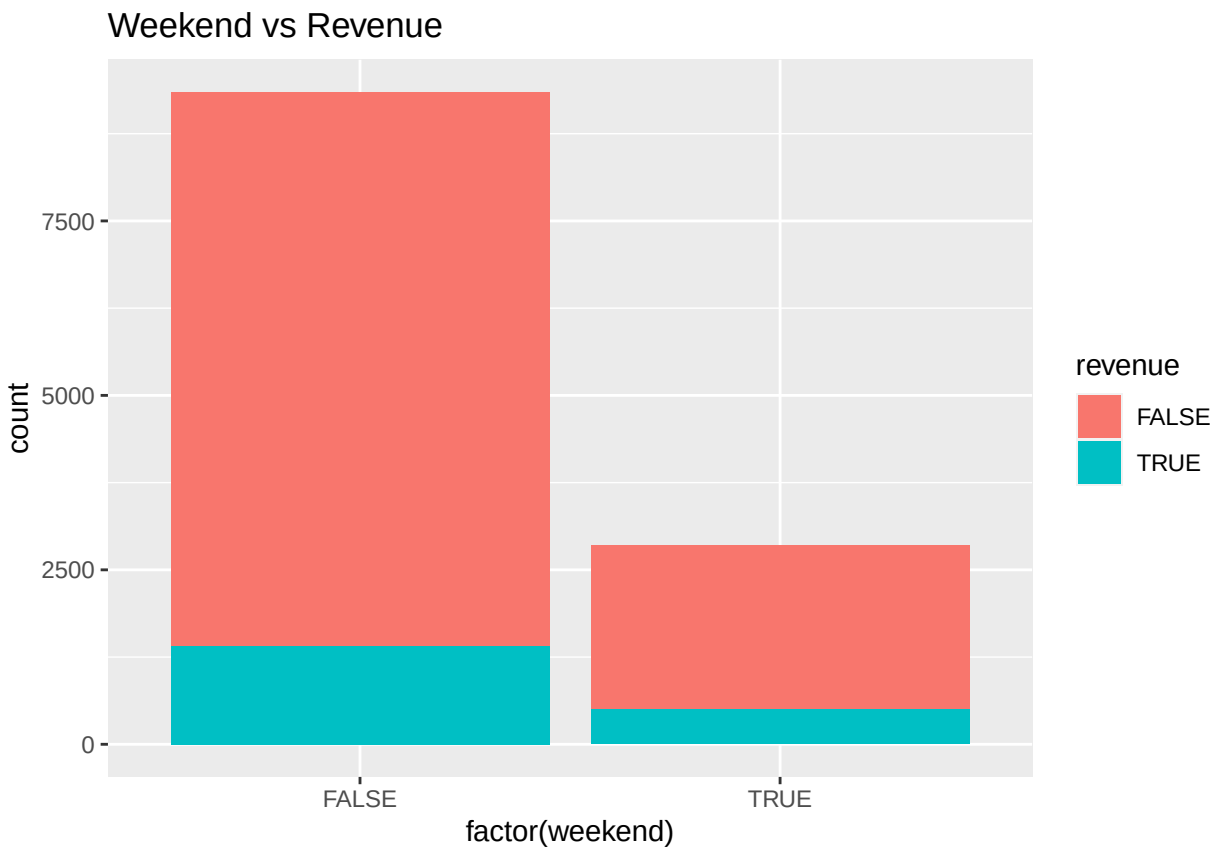
```
# Check how visitor type and revenue relate in this dataset

ggplot(df, aes(x = factor(visitortype), fill = revenue)) +
  geom_bar() +  ggtitle('Visitor Type vs Revenue')
```



The company gets higher revenue from return visitors to the site.

Weekend vs Revenue

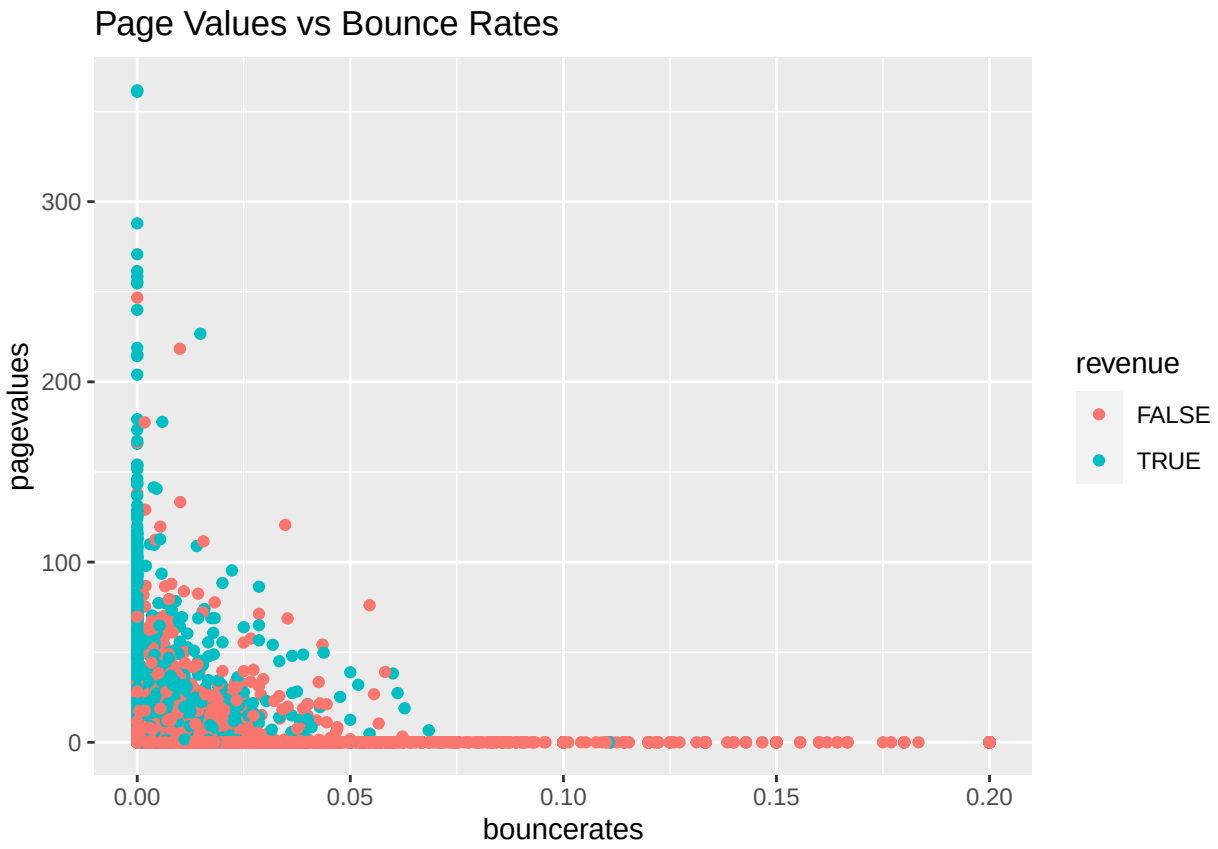
```
ggplot(df, aes(x = factor(weekend), fill = revenue)) +
  geom_bar() +  ggtitle('Weekend vs Revenue')
```



People visit the site both on weekdays and weekends. Weekdays produce a higher revenue.

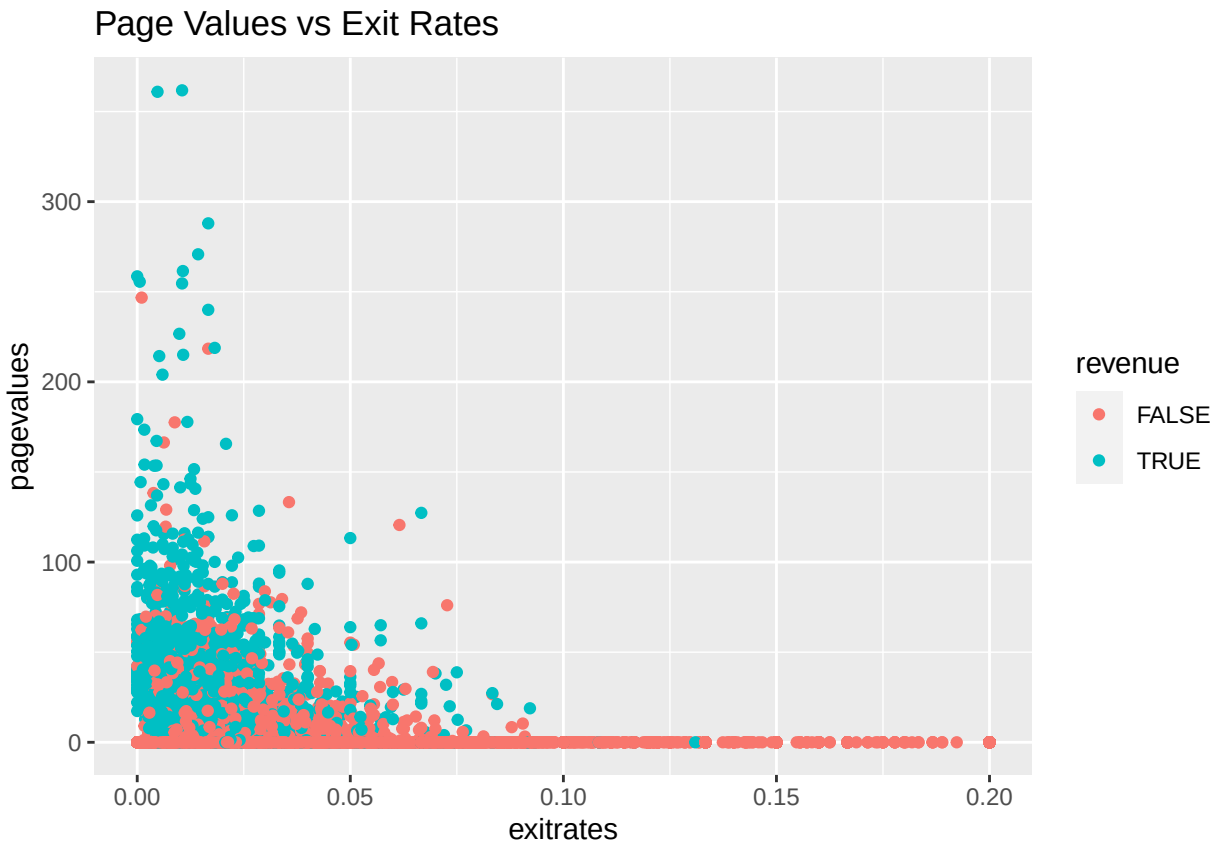
Page Values vs Bounce Rates

```
ggplot(df, aes(x = bouncerates, y = pagevalues , col = revenue)) +
  geom_point() +  ggtitle('Page Values vs Bounce Rates')
```



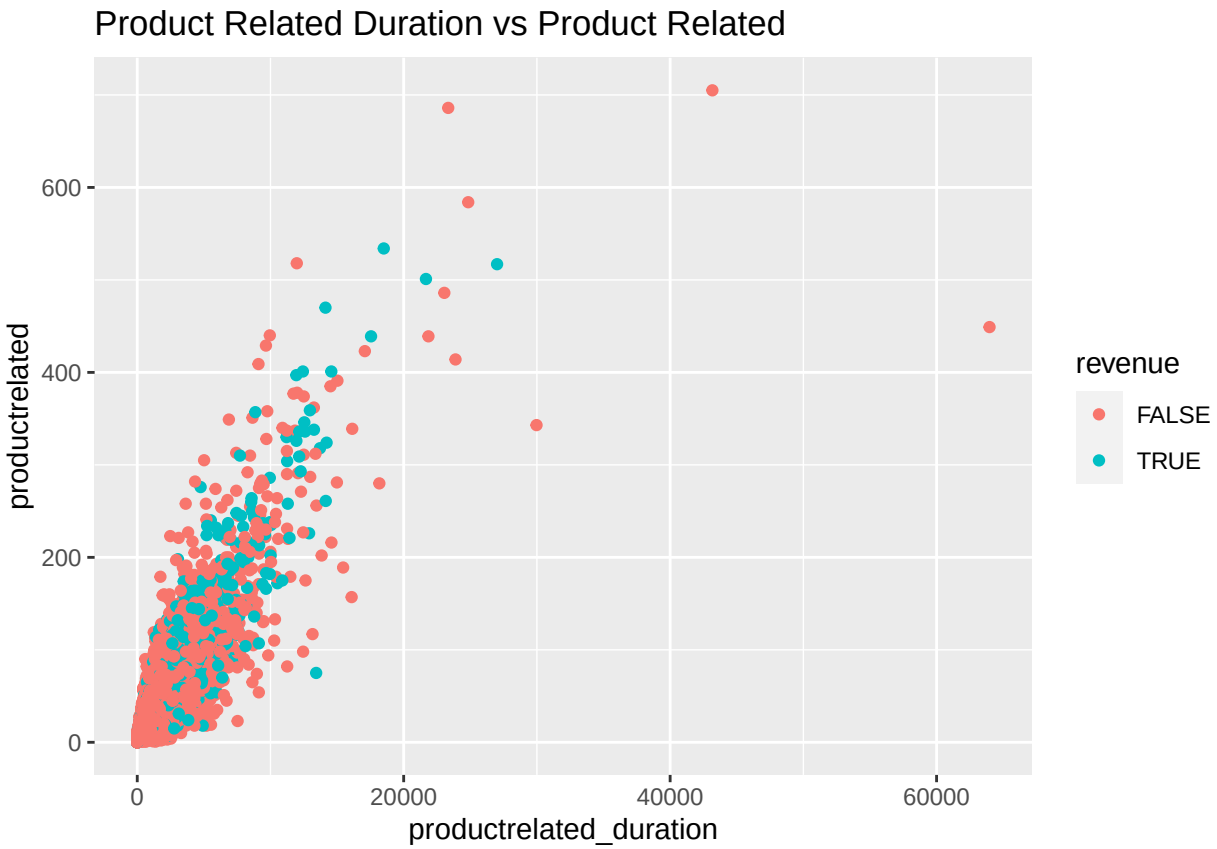
Page Values vs Exit Rates

```
ggplot(df, aes(x = exitrates, y = pagevalues , col = revenue)) +
  geom_point() +  ggtitle('Page Values vs Exit Rates')
```



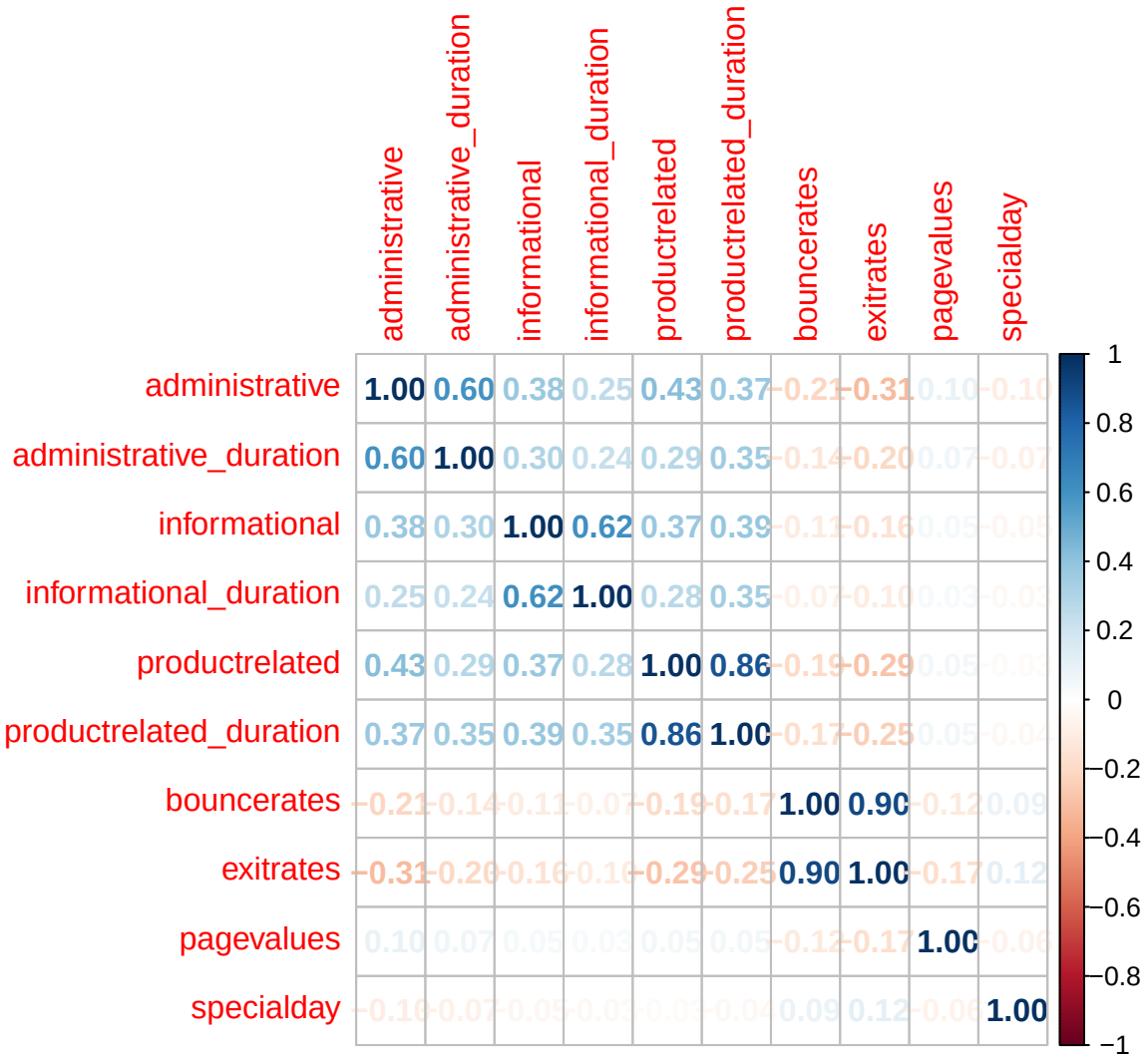
Product Related Duration vs Product Related

```
ggplot(df, aes(x = productrelated_duration, y = productrelated , col = revenue)) +
  geom_point() +  ggtitle('Product Related Duration vs Product Related')
```



Correlation We should find the correlation of the data

```
# calculate correlations
correlations <- cor(df[,1:10])
# create correlation plot
corrplot(correlations, method="number")
```



From the correlation plot that there are only few columns that are correlated. The correlations is usually from the duration columns. However, the bounce rate and exit rates are also correlated

Unsupervised machine learning models

Preparation of the data

```
#Create a copy of our dataframe
df_copy <- data.frame(df)             # Create copy of data

# Preprocessing the dataset
# ---
# Since clustering is a type of Unsupervised Learning,
# we would not require Class Label(output) during execution of our algorithm.
# We will, therefore, remove Class Attribute "revenue" and store it in another variable.
# We would then normalize the attributes between 0 and 1 using our own function.
# ---
```

```
#
df.class<- df_copy[, "revenue"]
df.model <- df[, c(1, 2, 3, 4,5,6,7,8,9,10,11,12,13,14,15,16,17)]
head(df.model)
```

```
##    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
## 1: Returning_Visitor   FALSE
## 2: Returning_Visitor   FALSE
## 3: Returning_Visitor   FALSE
## 4: Returning_Visitor   FALSE
## 5: Returning_Visitor    TRUE
## 6: Returning_Visitor   FALSE
```

```
#Check for unique values in revenue column
uniq_rev <- unique(df_copy$revenue, )
length(uniq_rev)
```

```
## [1] 2
```

```
# Use label encoder to change logical and categorical variable to numerical for the model
# Initialise the instance of the encoder
lbl <- LabelEncoder$new()

# Month
lbl$fit(df.model$month)
df.model$month <- lbl$fit_transform(df.model$month)

#Visitortype
lbl$fit(df.model$visitortype)
df.model$visitortype <- lbl$fit_transform(df.model$visitortype)
```

```
#Weekend
df.model$weekend <- as.integer(df.model$weekend)

head(df.model)
```

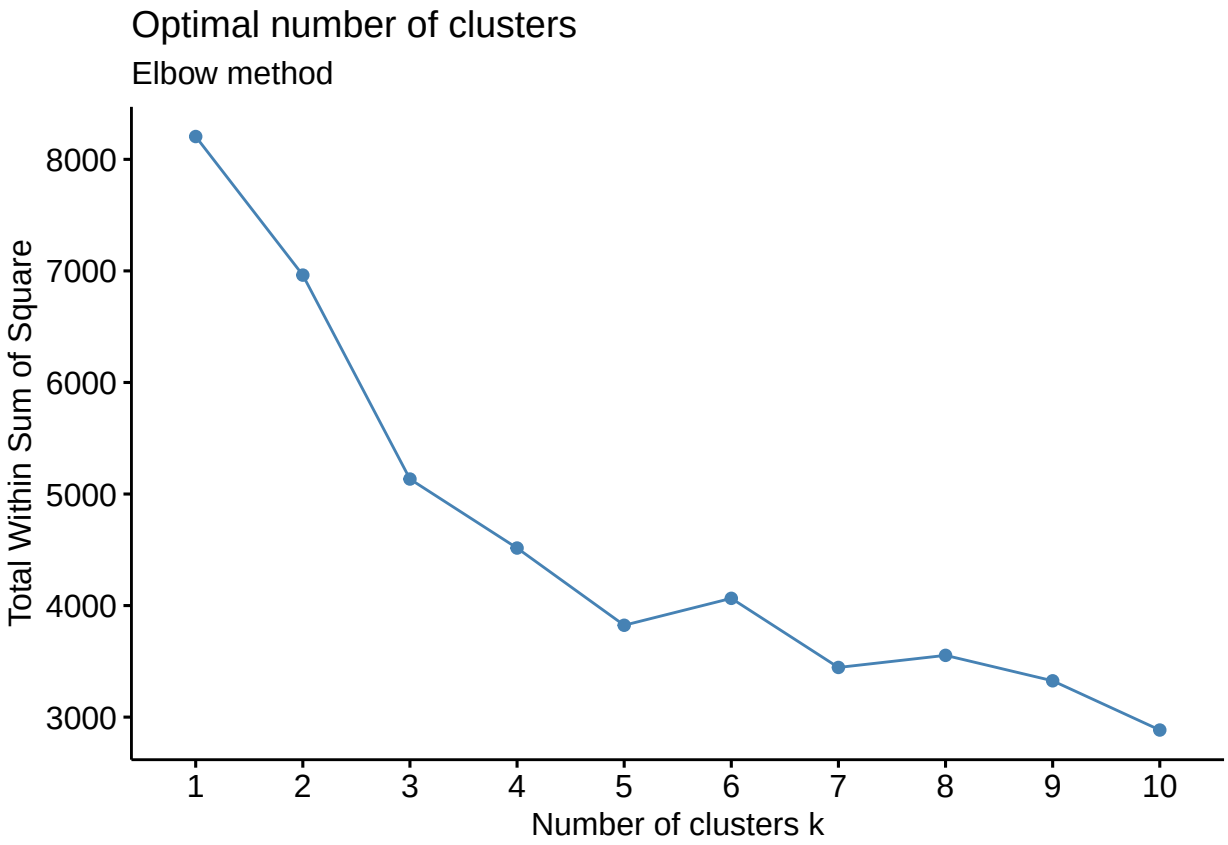
```
##    administrative administrative_duration informational informational_duration
## 1:              0                       0             0                      0
## 2:              0                       0             0                      0
## 3:              0                      -1             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 visitortype
## 1:          0     0                1       1      1           1           0
## 2:          0     0                2       2      1           2           0
## 3:          0     0                4       1      9           3           0
## 4:          0     0                3       2      2           4           0
## 5:          0     0                3       3      1           4           0
## 6:          0     0                2       2      1           3           0
##    weekend
## 1:       0
## 2:       0
## 3:       0
## 4:       0
## 5:       1
## 6:       0
```

Now we can normalize the data.

```
df_norm <- as.data.frame(apply(df.model, 2, function(x) (x - min(x))/(max(x)-min(x))))
```

Determining the optimal number of clusters for the KMeans function

```
fviz_nbclust(df_norm, kmeans, method = "wss") +
   labs(subtitle = "Elbow method")
```



From the plot above, it suggests that optimal number of clusters is 3 but from going through our dataset we know that it is 2.. so we will go with 2

```
km.out <- kmeans(df_norm, centers = 2, nstart = 20, iter.max = 50)
```

```
# Viewing the cluster center datapoints by each attribute
km.out$centers
```

```
##   administrative administrative_duration informational informational_duration
## 1     0.08502702              0.02393269    0.02019337             0.01333872
## 2     0.09203496              0.02558730    0.02449521             0.01638451
##   productrelated productrelated_duration bouncerates exitrates pagevalues
## 1     0.04499291              0.01880571  0.10680370  0.214444 0.01614319
## 2     0.04704343              0.01916769  0.08728349  0.184715 0.01747125
##   specialday     month operatingsystems   browser    region traffictype
## 1 0.06391951 0.5078074        0.1604868 0.1162992 0.2694397   0.1621084
## 2 0.05560224 0.5124883        0.1610644 0.1029704 0.2682511   0.1608801
##   visitortype weekend
## 1  0.07288879       0
## 2  0.08630952       1
```

```
# Looking at how many data points are in each cluster
km.out$size
```

```
## [1] 9343 2856
```

```r
# Getting the cluster vector that shows the cluster where each record falls
# ---
#
km.out$cluster
```

```
##    [1] 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1
##   [37] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
##   [73] 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
##  [109] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 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 2 2 2 1 1 2 2 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 1 2 1
##  [181] 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1
##  [217] 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2 2 2 1 2 2 2 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1
##  [253] 2 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 2 2 1 1 1 2 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1
##  [289] 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [325] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2
##  [361] 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1
##  [397] 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 2 2 1 2 2 1 2 1 1 2 1 2 2 1 1 1 1 2 1 2 2 2
##  [433] 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 2 2 2 1 1 2 1 2 1 1 2 1 1 1 2 1 1
##  [469] 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 2 2 2 2 2 1 1 1 1 1 1 2 1 1 1 1
##  [505] 1 1 1 2 1 1 1 2 1 2 1 1 2 2 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 2 2
##  [541] 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 2 2 1 1 2 1 2 1 2 2 1 1 2 1 1
##  [577] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 2 1 2 1 1 1 2 1 1 1 1 1
##  [613] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1
##  [649] 1 1 1 2 1 2 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 2 1 1 1 2 1 1 2 1 2 1 1 1 1 1 2
##  [685] 1 1 2 2 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2
##  [721] 1 2 1 2 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1
##  [757] 2 1 1 1 1 1 2 2 1 2 1 1 2 2 1 1 2 1 1 1 1 1 2 2 1 1 1 2 1 1 2 1 1 1 1 1 1
##  [793] 2 1 1 1 1 1 1 2 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 2 2 2 1
##  [829] 2 1 2 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 2 1 1 1 2 1 2 2
##  [865] 1 2 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 2 1 2 2 1 1 2
##  [901] 1 1 2 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2
##  [937] 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2
##  [973] 2 1 1 2 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 1 1 2 1 1 1 2 1 1 2
## [1009] 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 2 2 2 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 1
## [1045] 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 2
## [1081] 1 1 1 2 1 2 2 1 2 2 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
## [1117] 1 2 2 1 1 1 1 1 1 1 1 2 1 2 2 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1
## [1153] 1 1 1 1 1 1 2 2 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1
## [1189] 1 1 2 1 2 1 2 2 1 2 1 1 1 1 1 1 1 2 2 2 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1
## [1225] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 2 2 1 1 2 1 1 1
## [1261] 2 1 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 2 2 1
## [1297] 2 1 2 1 1 1 1 2 1 1 1 1 1 2 2 2 1 1 2 2 2 1 1 1 1 1 2 1 2 1 1 2 1 1 1 1
## [1333] 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 2 1 2 1 1 1 1 1 2 2 1 2 1 1 2 2 2 1 1 1 1 2
## [1369] 1 2 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 2 1 1
## [1405] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 1 1 1 2 2 1 1 1
## [1441] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1
## [1477] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 1 2 1 1 1 1 1
## [1513] 2 2 1 2 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1
## [1549] 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 2 1 1 2 1 2
## [1585] 1 1 1 1 2 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1
## [1621] 2 2 2 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 1 1 1 1 2 1 2 2 1 2 1 1 1 2 2
## [1657] 1 1 1 2 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1
## [1693] 2 2 1 1 1 1 1 1 2 2 1 2 1 1 1 1 2 1 2 1 1 2 1 2 1 1 1 1 1 2 2 1 1 1 1 1
```

```
##  [1729] 1 1 2 1 2 1 1 2 2 2 2 1 1 2 1 1 2 1 2 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 2
##  [1765] 1 1 1 2 2 1 1 2 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1
##  [1801] 1 1 1 2 1 1 1 1 1 1 2 2 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1
##  [1837] 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 2 1 1 2 2 1 1 1 2 1 2 1
##  [1873] 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 2 2
##  [1909] 1 1 2 1 2 1 1 1 1 1 1 1 2 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1
##  [1945] 1 1 2 1 1 2 2 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 2 2 1 1 2 1 1 1 1
##  [1981] 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1
##  [2017] 2 1 1 1 1 1 1 1 2 1 1 1 2 2 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
##  [2053] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 2 2 2 1 1 1 1 1 1 2 1 2 1
##  [2089] 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 2 1 2 2 2 1 1 1 1 2 1 1 1 1 1
##  [2125] 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 2 2 1 2
##  [2161] 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 2 2 2 1 1 1 1
##  [2197] 1 1 2 1 2 1 2 1 1 2 1 2 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
##  [2233] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 2 1 2 2 1 1 1 2 1 2 1 1
##  [2269] 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1
##  [2305] 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
##  [2341] 1 2 2 1 1 1 1 1 2 2 2 1 1 1 1 2 1 1 2 2 1 1 2 2 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 2 1 2 1 1 1 2 1 2 1 1 1 2 1 2 1 1 2 1 1 1 1
##  [2413] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 2 2 2 1 2 1 1 1 1 1 1 1 1 2 1 1 2 2 2 1 1
##  [2449] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1
##  [2485] 1 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 1 1 1
##  [2521] 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 2 2 1 1 1 1 2 2 1 2
##  [2557] 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 1 1
##  [2593] 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1
##  [2629] 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
##  [2665] 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1
##  [2701] 1 1 1 2 1 1 2 2 2 1 2 2 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1
##  [2737] 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1
##  [2773] 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1
##  [2809] 1 1 1 2 1 2 1 2 2 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1
##  [2845] 1 2 1 2 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 2 1 1 2 1 1 1 2 1 1
##  [2881] 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 2 2 2 1 2 2 1 1 2 1 1
##  [2917] 2 1 1 2 1 1 2 2 2 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2
##  [2953] 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
##  [2989] 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [3025] 1 2 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 2 2 1 1 2 1 1
##  [3061] 1 1 1 2 1 1 1 2 1 2 2 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1
##  [3097] 1 2 1 2 2 1 1 2 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
##  [3133] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1
##  [3169] 1 1 1 2 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 2 1 1 1 1
##  [3205] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1
##  [3241] 2 2 1 1 2 2 1 1 1 1 1 2 2 1 1 1 1 2 2 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1
##  [3277] 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 2 2 2 2 1
##  [3313] 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 2 1 1
##  [3349] 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
##  [3385] 1 1 2 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2
##  [3421] 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 2 1 2 2 1 1 2 1 1 1 1 1 2 1 2
##  [3457] 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1
##  [3493] 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1
##  [3529] 1 1 1 1 2 2 2 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 2 1 2
##  [3565] 1 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2
##  [3601] 1 2 1 2 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
##  [3637] 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2
```

```
##  [3673] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1
##  [3709] 1 1 1 1 2 1 2 2 1 2 1 2 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2
##  [3745] 2 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1
##  [3781] 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1
##  [3817] 2 2 1 1 2 1 1 1 1 2 1 1 2 1 1 2 2 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 2 1 2 2
##  [3853] 1 1 1 2 1 2 2 1 2 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1
##  [3889] 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
##  [3925] 1 1 1 2 2 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 1
##  [3961] 1 1 1 2 2 2 2 1 1 2 2 1 1 1 1 1 2 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 1 1 2 1
##  [3997] 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 2
##  [4033] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
##  [4069] 1 1 1 1 1 2 1 1 1 1 2 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1
##  [4105] 1 1 2 1 1 2 1 2 1 2 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1
##  [4141] 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1
##  [4177] 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 2 1 1 2 1 2
##  [4213] 1 1 1 2 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 2 2 1 1 1 1 2 1 1 1
##  [4249] 1 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 2 1 1 1
##  [4285] 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1
##  [4321] 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1
##  [4357] 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 2 1 2
##  [4393] 2 1 2 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 2 1 1
##  [4429] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2
##  [4465] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 2 2 2 1 1 2 1 2 1 1 1 1 1 1 1
##  [4501] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1
##  [4537] 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 2 2 1 1 2 1 1 2 2 1 1
##  [4573] 2 1 2 1 1 1 2 2 1 2 1 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1
##  [4609] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 2 1 2 1 1 1 2 1 1 1 1 1
##  [4645] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1
##  [4681] 1 1 2 1 1 2 1 1 1 2 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2
##  [4717] 1 2 1 1 1 1 1 2 1 1 1 2 1 2 2 1 1 1 2 2 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1
##  [4753] 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
##  [4789] 2 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 2 2 1 2 1 1
##  [4825] 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1
##  [4861] 1 1 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1
##  [4897] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1
##  [4933] 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
##  [4969] 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
##  [5005] 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1
##  [5041] 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 2 1 1 2 1 1 1 1 1 1
##  [5077] 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 2 2 1 2 2 1 1 1 1 2 2 1 2 1 2 1 1
##  [5113] 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [5149] 1 2 1 2 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2 1 2 1 1 1 1 1 1 2
##  [5185] 2 1 1 1 2 1 1 2 2 1 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 2 1 1 1 2 1 1 1 2
##  [5221] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 2 1 1 1 2 2 1 1
##  [5257] 1 1 2 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1
##  [5293] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1
##  [5329] 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1
##  [5365] 1 1 1 1 1 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 1 1 2 1 1 1 1
##  [5401] 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 2 2 2 1 1 1 1 1 2 2 1 1 1 2 1 1 1 1 2
##  [5437] 1 1 1 1 2 1 1 1 2 2 1 2 1 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2
##  [5473] 1 1 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1
##  [5509] 1 1 2 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 2 1 1 1 2 2 2 1 1 2 2
##  [5545] 1 1 2 1 2 1 1 2 1 2 1 1 2 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [5581] 1 2 2 1 1 2 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 2 1 1 1 2 1 2 1
```

```
##  [5617] 2 1 1 2 1 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1
##  [5653] 1 2 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1
##  [5689] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 2 1 2 2 1 1 1 1 2 1
##  [5725] 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 2
##  [5761] 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 1
##  [5797] 1 2 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1 2 2 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1
##  [5833] 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1
##  [5869] 1 1 1 2 1 1 1 1 2 1 1 2 2 1 2 1 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 2 1 1 1 2
##  [5905] 1 1 1 2 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 2 1
##  [5941] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 1 2 1 1 1 1 1 1 1
##  [5977] 2 1 2 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2
##  [6013] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 2 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 2 2 1
##  [6049] 2 1 2 1 2 1 1 1 2 1 1 2 1 2 2 1 2 1 2 1 2 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1
##  [6085] 1 1 2 2 2 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1
##  [6121] 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 2
##  [6157] 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 2 1 1 2 2 1 1 1 1 1 1
##  [6193] 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1
##  [6229] 1 1 2 2 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [6265] 1 2 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 2 1 1 2 2 2 2 1 1 1 1 1 1 1 1 2
##  [6301] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 2 1 2 1 2 1 1 2 2 1 1 1 2
##  [6337] 1 2 2 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1
##  [6373] 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1
##  [6409] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 2 1 2 1 2 1 1 1
##  [6445] 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1
##  [6481] 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
##  [6517] 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 2 2
##  [6553] 1 1 2 2 1 2 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 2
##  [6589] 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
##  [6625] 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
##  [6661] 1 1 1 1 2 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1
##  [6697] 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 2 2 1 1 1 1 1 1
##  [6733] 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [6769] 1 2 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1
##  [6805] 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1
##  [6841] 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 1 2 1 1 2 1 1 1
##  [6877] 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1
##  [6913] 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 2 2 2 1 1 1 1 2 1 1 1 1 1 1 2 1
##  [6949] 1 1 2 1 1 2 2 1 2 2 1 2 2 1 1 2 1 1 1 2 1 1 1 1 2 2 2 1 1 2 1 1 1 2 1 1
##  [6985] 1 1 1 1 2 2 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1
##  [7021] 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
##  [7057] 1 1 1 1 1 1 1 2 1 1 2 2 1 1 2 1 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1
##  [7093] 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 2 2 1 1 1
##  [7129] 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 2 1 2 1 1 1 1 1 2 1 1 2 1 1 1
##  [7165] 1 2 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1
##  [7201] 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1
##  [7237] 1 1 2 1 1 1 2 1 1 2 2 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 2 1 1 1 2 1 2 1 1
##  [7273] 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 2 2 1 1 2 1 1 1 2
##  [7309] 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1
##  [7345] 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
##  [7381] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 1
##  [7417] 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 2 2 1
##  [7453] 1 1 1 2 1 1 1 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2
##  [7489] 1 2 1 1 1 1 1 2 2 1 1 1 2 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 2 1 1 2 1 1 2 1
##  [7525] 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 2 1 1
```

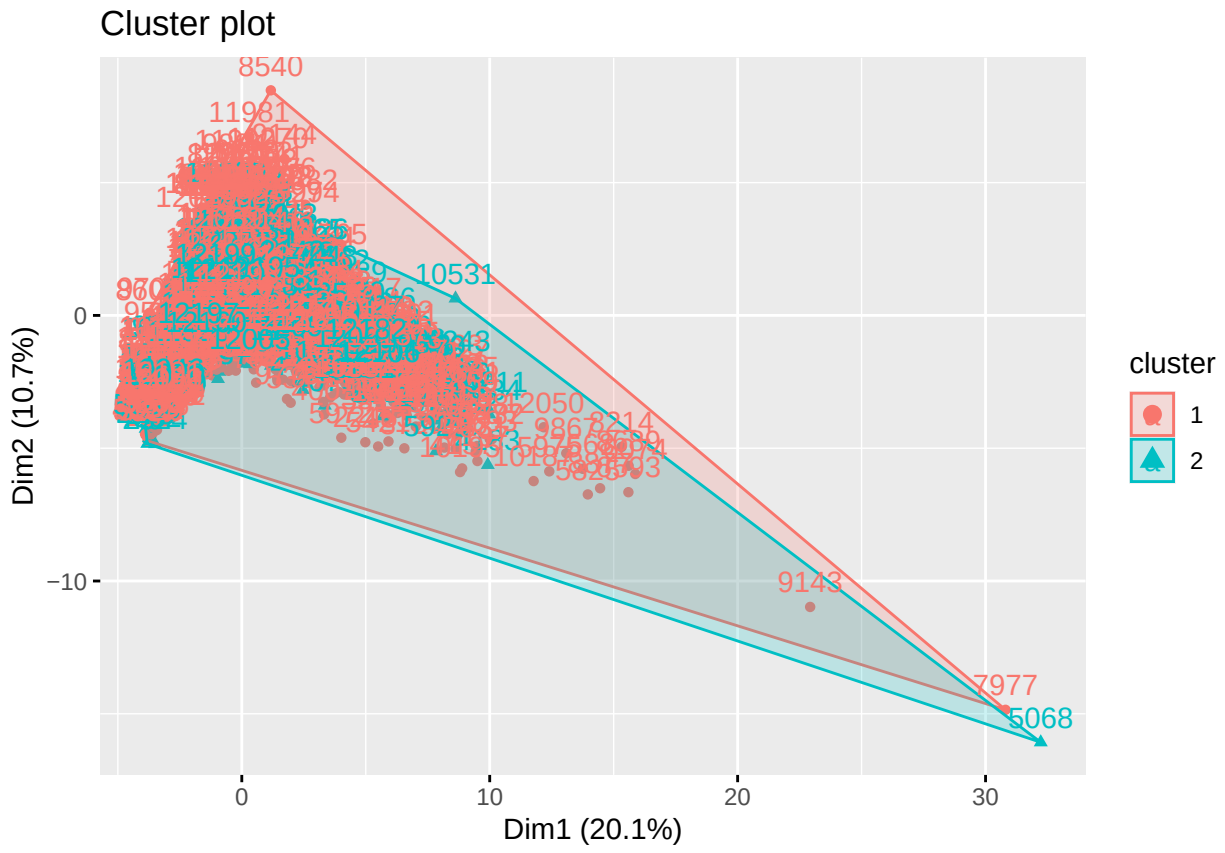
```
##  [7561] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 2 1 1 1 1 2
##  [7597] 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1
##  [7633] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1
##  [7669] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 2 1 1 2 1 1 2 1 1 1 1
##  [7705] 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1
##  [7741] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 2
##  [7777] 1 2 1 2 1 2 1 1 1 1 2 2 2 1 1 1 2 1 1 2 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 2
##  [7813] 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1 2 2 1 1 2 1 1 1
##  [7849] 2 2 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1
##  [7885] 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1 2 2 1 1
##  [7921] 2 2 1 2 1 2 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
##  [7957] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 2 1 1 1 1 1 2 1 1
##  [7993] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1
##  [8029] 1 1 1 1 2 1 2 1 1 1 2 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 2 1
##  [8065] 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1
##  [8101] 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [8137] 1 2 2 1 1 1 1 2 2 1 1 1 1 1 2 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1
##  [8173] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 2 1
##  [8209] 2 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1
##  [8245] 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1
##  [8281] 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 2 2 2 1 1 2 1 1 2
##  [8317] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 2 1 1 1 1 1 2 2 2 2 1 1 1
##  [8353] 1 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 2 2 2 1 2 2 1 1 2 1 2
##  [8389] 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1
##  [8425] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1
##  [8461] 2 1 1 1 2 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 2 1 1 1 1
##  [8497] 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [8533] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1
##  [8569] 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 2 2 1 1 2 1 1 2 2 1 1 1 1 2 1 2 1 1
##  [8605] 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 2 2 1 1 1 1 1
##  [8641] 1 1 1 1 2 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1
##  [8677] 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 2 2 1 1 1 2 1 1 1 2
##  [8713] 1 1 1 2 1 2 2 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 1
##  [8749] 2 2 2 1 1 1 1 1 1 2 2 2 1 1 1 1 2 1 2 1 2 1 1 1 2 2 2 1 2 1 1 1 1 2 2 2
##  [8785] 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 2 1 2 2 2 2 1 1 1 2 1 1 1 1 1 1 1
##  [8821] 1 2 2 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1
##  [8857] 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 2 1 2 1 1 2 1 2 1 1 1 1 1 1 1 2 2 1 2
##  [8893] 1 1 1 2 1 1 1 1 2 1 1 2 1 2 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1
##  [8929] 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [8965] 1 1 1 1 1 2 1 2 2 1 1 1 2 1 1 1 1 1 2 1 1 2 1 2 1 2 2 1 1 1 2 2 2 1 1 2
##  [9001] 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 2 2 1 1 2 1 1 1 1 1 2 1 1 2 1 1 1
##  [9037] 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 1 2 1 2 1 2 1 2 1 1
##  [9073] 1 2 2 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 1 1 1 2 2 1 1 2
##  [9109] 2 1 2 1 2 1 1 1 2 1 1 1 1 1 2 1 1 2 2 1 2 1 2 1 2 1 2 1 2 1 1 2 1 2 1 1
##  [9145] 2 2 1 1 1 2 1 1 2 1 1 1 2 1 2 1 1 2 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1
##  [9181] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1
##  [9217] 1 1 2 1 1 1 2 1 2 1 2 1 1 2 2 1 1 1 2 1 2 1 2 1 1 1 2 2 1 2 1 2 1 1 1 2
##  [9253] 2 1 1 1 2 1 2 1 2 2 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1
##  [9289] 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1
##  [9325] 2 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 2 2 1 2 1 1 2 1 1 1 1 2 2
##  [9361] 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 2 2 1 1 1 1 1 1 1 2 2 2 2 1 2 1
##  [9397] 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 2 1 2 1 1 2 2 1 1 1 1 2 2 1 1 1 1 1 2 1
##  [9433] 1 2 1 1 2 1 2 1 2 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 2
##  [9469] 1 1 2 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2 2 2 2 1 2 1 1 1 1 1 1 1 2
```

```
##  [9505] 1 1 2 2 1 1 2 2 1 1 2 2 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1
##  [9541] 1 2 1 2 1 1 2 1 2 1 1 1 2 1 1 1 2 2 2 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1
##  [9577] 1 2 1 1 2 2 1 2 2 1 1 2 2 1 1 1 1 1 2 2 2 1 2 1 1 2 1 1 1 1 2 2 1 2 1 1
##  [9613] 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 1 2 2 1 1 1 1 1 2 1 1 1 2 1 2 2 1 1
##  [9649] 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 1
##  [9685] 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2
##  [9721] 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2 2 1 1 1 1 2 2 1 1 1 1 2 1
##  [9757] 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
##  [9793] 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 2 1 2 1 1 1 2 1 1 2 1 1 2 1 2 1 1
##  [9829] 2 1 2 1 1 2 2 1 1 1 1 1 2 1 2 1 1 1 1 2 2 1 2 1 1 2 1 1 2 2 2 1 1 1 2 1
##  [9865] 1 1 1 1 2 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 2 1 1
##  [9901] 1 1 1 2 1 2 2 1 2 2 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 1
##  [9937] 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1
##  [9973] 1 2 1 1 1 2 2 1 1 2 1 2 2 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 2 1 2 2 1
## [10009] 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 2 1
## [10045] 1 1 1 1 1 2 1 2 1 1 2 2 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 2
## [10081] 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 2 1 1 1 1 1
## [10117] 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 2 1 2 2 1 1 1 1 2 1 1 1 2 1 2 1 1 2
## [10153] 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [10189] 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 2 1 2 1 1 2 1 1 1 2 1 1 1 1
## [10225] 2 2 1 1 1 1 1 1 1 2 2 2 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
## [10261] 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 1 1 1 2 1
## [10297] 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 2 1 2 2 1 1 2 1 1 2 1 1
## [10333] 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1
## [10369] 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1
## [10405] 2 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1
## [10441] 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1
## [10477] 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 2 1 1
## [10513] 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1
## [10549] 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 2 1 1 1 1 1 1 2 2 1 1 2
## [10585] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 2 1 2 1 2 1 1 1
## [10621] 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 1 2 1 2 1 2 2 1 1 1 1 1 2 1 2 2
## [10657] 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2
## [10693] 1 1 2 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1
## [10729] 1 1 1 2 1 2 1 2 1 2 2 1 1 2 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 2 2 1
## [10765] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 2 1
## [10801] 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 1
## [10837] 2 1 2 1 1 2 1 1 1 1 2 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 2 2 2 2 1 1 1 2 1 1
## [10873] 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 2 1 2 2 1 1 1 1 2 2 1 1 2 1 1 1 1 2 2
## [10909] 1 1 2 2 1 2 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 2 1
## [10945] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1
## [10981] 1 2 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1
## [11017] 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1
## [11053] 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1
## [11089] 1 1 1 1 2 1 1 2 1 1 2 2 1 1 1 2 1 1 1 2 1 2 1 1 1 2 2 1 1 2 1 1 1 1 2 2
## [11125] 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2
## [11161] 1 1 1 2 2 1 2 2 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1
## [11197] 2 1 1 1 1 1 1 1 2 2 1 2 1 2 2 1 1 1 2 2 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1
## [11233] 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1
## [11269] 1 1 2 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1
## [11305] 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 2 1 2 1
## [11341] 1 2 2 2 2 1 2 1 2 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 2 1 2 1 2 2 1 1 2 1 1 1
## [11377] 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 2 2 1 1 1 2 1 1 1 1 1 1 2 1
## [11413] 2 1 1 2 1 1 1 1 1 1 1 1 2 1 2 2 1 1 2 1 2 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1
```

```
## [11449] 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 1 1 2 1 2 1 2 1 1 1 2 2 1 2 1 1 1 1
## [11485] 1 1 1 1 1 1 2 1 1 2 2 2 2 2 2 1 2 1 1 1 2 1 1 1 2 2 1 2 1 1 2 1 1 1 2 1
## [11521] 1 1 1 1 1 1 1 1 2 1 2 2 2 2 2 1 1 1 2 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1
## [11557] 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1
## [11593] 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 2 2 1
## [11629] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2
## [11665] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 2 1 2 2 1 1 1
## [11701] 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 2 2 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1 2 1
## [11737] 1 1 2 1 1 1 1 2 2 1 1 1 1 1 2 1 2 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 2 1 1 1
## [11773] 2 1 2 1 2 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 2 1 1 1 1 1 1 2 2 2 1 1 1 1 1
## [11809] 2 1 2 1 1 1 1 1 2 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
## [11845] 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 2 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1
## [11881] 1 1 1 2 2 1 2 2 2 1 1 2 1 1 2 2 1 2 2 1 1 1 1 1 1 2 1 2 1 1 2 2 1 1 2 1
## [11917] 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 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 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
## [11989] 1 1 1 1 2 1 2 2 1 1 1 1 1 1 2 2 2 1 2 1 1 1 2 1 2 1 2 2 1 1 2 1 1 1 1 2
## [12025] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 2 1 2 2 1 1 1 1 2 1 2 1 1 2 1
## [12061] 2 2 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1 2 1 1 2 1 2 1 1 2 2 1 2 1 1
## [12097] 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 2 1 1 1 2 1 1
## [12133] 1 1 2 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1
## [12169] 1 1 1 2 1 2 1 1 1 2 1 1 1 2 1 2 2 1 1 2 1 1 1 1 1 1 2 2 2 1 2
```

```
# The graph shows that we have got 3 clearly distinguishable clusters for Ozone and Solar.R data points
# Let's see how clustering has performed on Wind and Temp attributes.
```

```
fviz_cluster(km.out, data = df.model)
```



```
table(km.out$cluster,df$revenue)
```

```
##
##     FALSE TRUE
##   1  7934 1409
##   2  2357  499
```

The clustering was unable to correctly cluster majority of the data which was False but a better job in classifying True.

Challenging the solution

Hierarchical clustering

Scaling the data frame.

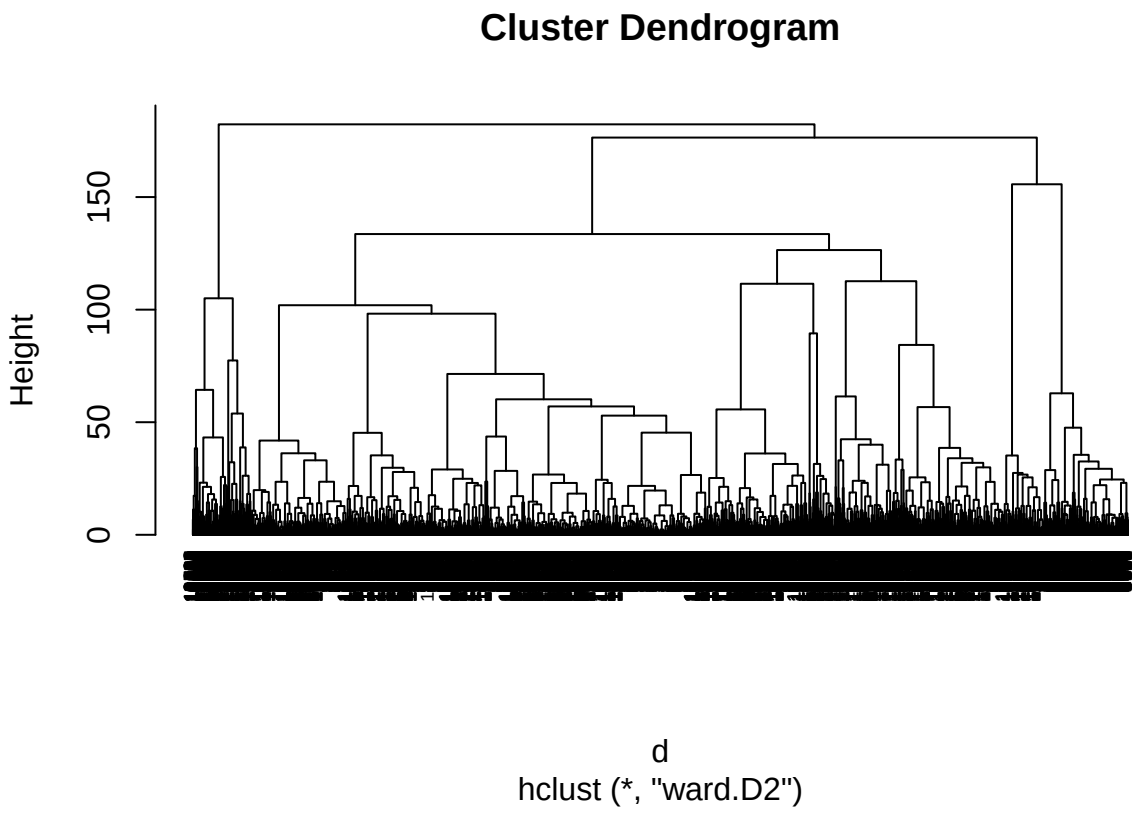
```
dfscaled <- as.data.frame(scale(df.model))
head(dfscaled)
```

```
##   administrative administrative_duration informational informational_duration
## 1     -0.7025315              -0.4601081    -0.3988128             -0.2462725
## 2     -0.7025315              -0.4601081    -0.3988128             -0.2462725
## 3     -0.7025315              -0.4657410    -0.3988128             -0.2533417
## 4     -0.7025315              -0.4601081    -0.3988128             -0.2462725
## 5     -0.7025315              -0.4601081    -0.3988128             -0.2462725
## 6     -0.7025315              -0.4601081    -0.3988128             -0.2462725
##   productrelated productrelated_duration  bouncerates  exitrates pagevalues
## 1     -0.6963635              -0.6289343  3.954699721  3.4273070 -0.3190356
## 2     -0.6739424              -0.5955997 -0.450343788  1.2650121 -0.3190356
## 3     -0.6963635              -0.6294551  3.954699721  3.4273070 -0.3190356
## 4     -0.6739424              -0.6275453  0.650917089  2.1299300 -0.3190356
## 5     -0.4945739              -0.3020990 -0.009839437  0.1838646 -0.3190356
## 6     -0.2927843              -0.5486101 -0.102577188 -0.3661929 -0.3190356
##   specialday     month operatingsystems    browser     region traffictype
## 1 -0.3103105 -1.538019       -1.2396607 -0.7939682 -0.8962939 -0.76562243
## 2 -0.3103105 -1.538019       -0.1371074 -0.2093703 -0.8962939 -0.51660683
## 3 -0.3103105 -1.538019        2.0679992 -0.7939682  2.4336556 -0.26759123
## 4 -0.3103105 -1.538019        0.9654459 -0.2093703 -0.4800502 -0.01857564
## 5 -0.3103105 -1.538019        0.9654459  0.3752276 -0.8962939 -0.01857564
## 6 -0.3103105 -1.538019       -0.1371074 -0.2093703 -0.8962939 -0.26759123
##   visitortype    weekend
## 1  -0.4032028 -0.5528638
## 2  -0.4032028 -0.5528638
## 3  -0.4032028 -0.5528638
## 4  -0.4032028 -0.5528638
## 5  -0.4032028  1.8086156
## 6  -0.4032028 -0.5528638
```

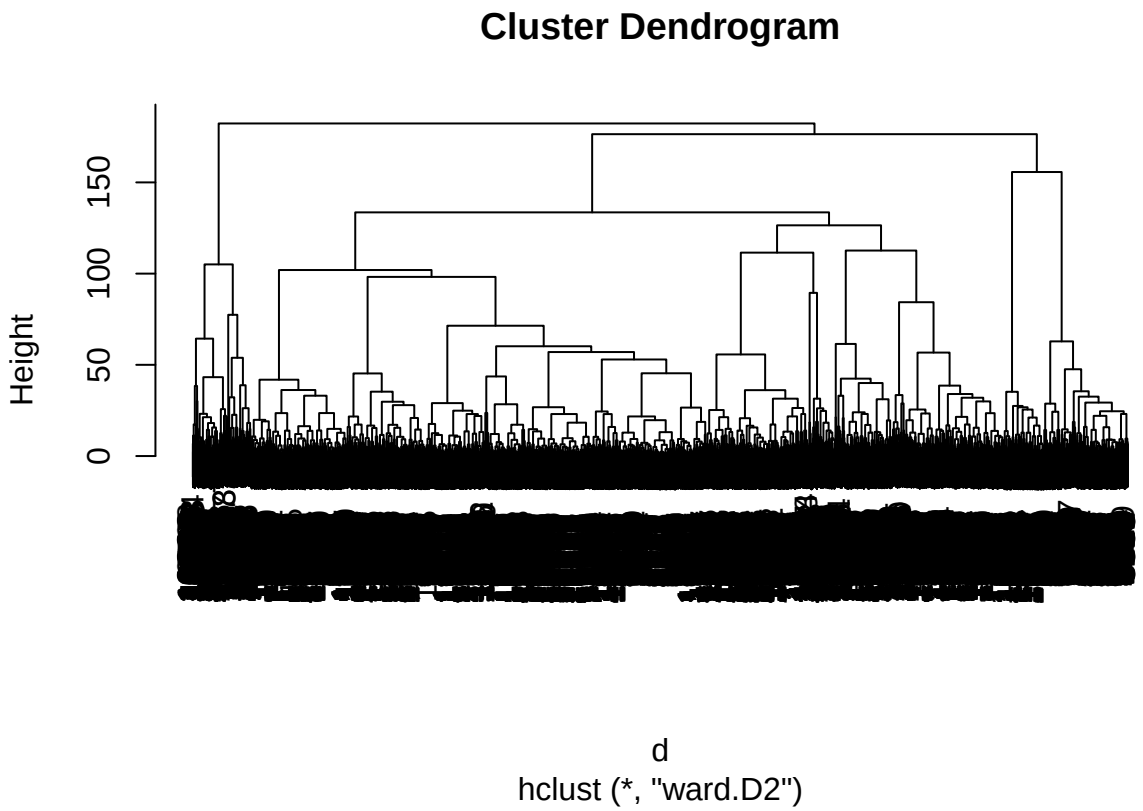
```r
# Making the clustering model
# hclust.out <- hclust (dist(df_norm), method = "complete")
# summary(hclust.out)


d <- dist(dfscaled, method = "euclidean")

# We then hierarchical clustering using the Ward's method
# ---
#
set.seed(240)
res.hc <- hclust(d, method = "ward.D2" )
res.hc
```

```
##
## Call:
## hclust(d = d, method = "ward.D2")
##
## Cluster method   : ward.D2
## Distance         : euclidean
## Number of objects: 12199
```

```r
# Lastly, we plot the obtained dendrogram
# ---
#
plot(res.hc, cex = 0.6, hang = -1)
```



```
# Drawing the output of the Hierachial cluster
plot (res.hc)
```



As we can see this is a very big mess. Therefore, the number of clusters can be used as a metric to cut down the tree.

Now we have to cut the tree using the predetermined optimal cluster number

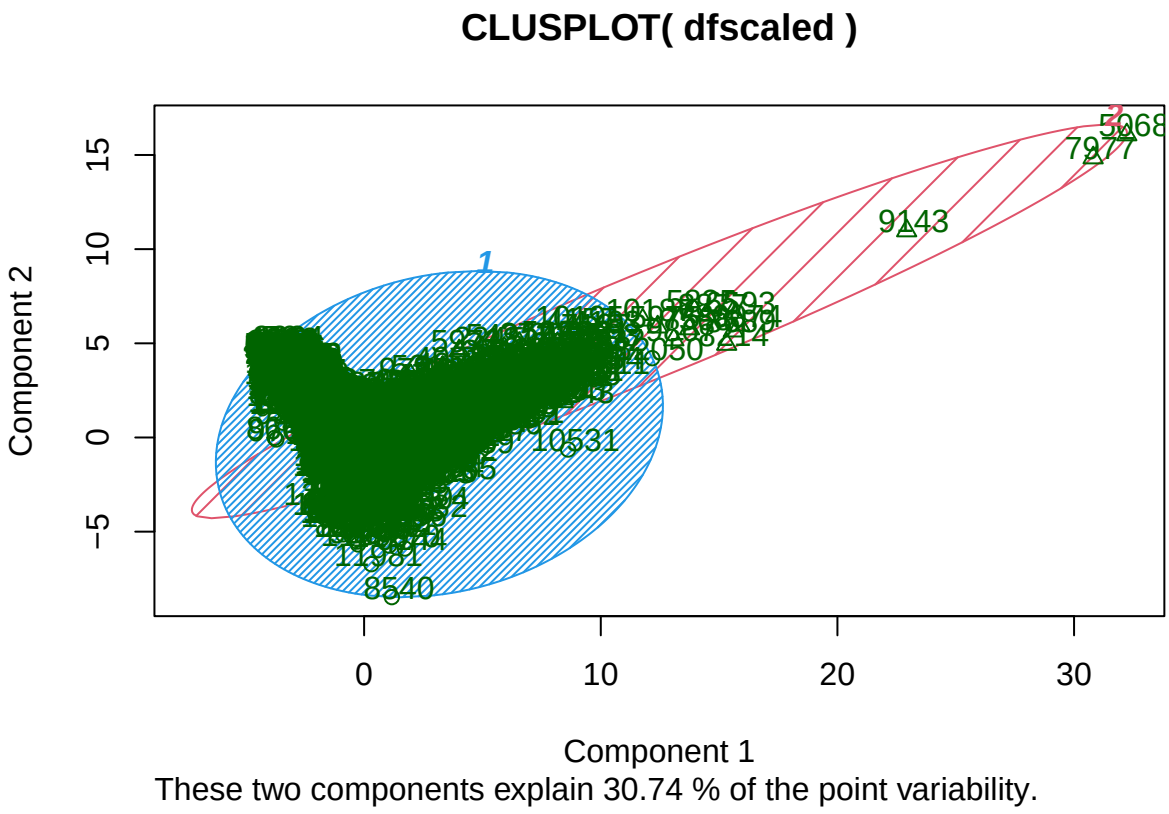
```
# cut the tree
cut.df <- cutree(res.hc, k = 2)
```

```
table(cut.df,df$revenue)
```

```
##
## cut.df FALSE TRUE
##      1  9778 1659
##      2   513  249
```

Hierarchical clustering did a better job than k means clustering, For the False but not much for the True

```
library(cluster)
clusplot(dfscaled, cut.df, color=TRUE, shade=TRUE,
    labels=2, lines=0)
```



Conclusion

Both Hierarchical and kmeans were unable to predict correctly with hierarchical having a problem with True and Kmeans with False.

Follow up Questions

Was the data provided enough to answer the question? Yes, the data was sufficient.

What could be done to improve the quality of data?

The data had many columns. However, applying dimension reduction techniques to the data would make it more understandable and easy to work with.