IP_Week13

2022-07-22

Online cryptography course advertising

Specifying the Data Analytic Question

> As a Data Science Consultant, I have been employed by an online cryptography course entrepreneur to help her identify which individuals are most likely to click on her ads.

Defining the Metric for Success

- Identifying which individuals are most likely to click on her ads

Understanding the context

> A Kenyan entrepreneur has created an online cryptography course and would want to advertise it on her blog. She currently targets audiences originating from various countries. She would like to identify which individuals are most likely to click on her ads.

Recording the Experimental Design

- Reading the Data
- Tidying the Dataset
- Exploratory Analysis
- Implementing the Solution
- Challenging the solution
- Follow-up questions

Data Relevance

> All the variables given are relevant to the entrepreneur and will help know more the how the online course advertisement was received

Reading and checking the data

```
# downloading tidyverse packages and library

install.packages('tidyverse', repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/Lenovo/AppData/Local/R/win-library
/4.2'
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked
##
```

```
## The downloaded binary packages are in
##  C:\Users\Lenovo\AppData\Local\Temp\RtmpID6235\downloaded_packages

library(tidyverse)

## ── Attaching packages
## ─────────────────────────────────────
## tidyverse 1.3.2 ──

## ✔ ggplot2 3.3.6     ✔ purrr   0.3.4
## ✔ tibble  3.1.7     ✔ dplyr   1.0.9
## ✔ tidyr   1.2.0     ✔ stringr 1.4.0
## ✔ readr   2.1.2     ✔ forcats 0.5.1
## ── Conflicts ──────────────────────────────────────── tidyverse_co
nflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()

#Read the dataset

advert <- read_csv("C://Users//Lenovo//Downloads//DB_prep//advertising.
csv")

## Rows: 1000 Columns: 10
## ── Column specification ──────────────────────────────────────────
──────────
## Delimiter: ","
## chr  (3): Ad Topic Line, City, Country
## dbl  (6): Daily Time Spent on Site, Age, Area Income, Daily Internet
 Usage, ...
## dttm (1): Timestamp
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

#Checking the head and tail of the data

head(advert)

## # A tibble: 6 × 10
##   `Daily Time Spent…`   Age `Area Income` `Daily Interne…` `Ad Topic
 Line` City
##                 <dbl> <dbl>         <dbl>            <dbl> <chr>
       <chr>
## 1                69.0    35        61834.             256. Cloned 5t
hgene… Wrig…
## 2                80.2    31        68442.             194. Monitored
 nati… West…
## 3                69.5    26        59786.             236. Organic b
```

```
ottom… Davi…
## 4                74.2    29        54806.             246. Triple-bu
ffere… West…
## 5                68.4    35        73890.             226. Robust lo
gisti… Sout…
## 6                60.0    23        59762.             227. Sharable
clien… Jami…
## # … with 4 more variables: Male <dbl>, Country <chr>, Timestamp <dtt
m>,
## #   `Clicked on Ad` <dbl>

tail(advert)

## # A tibble: 6 × 10
##   `Daily Time Spent…`   Age `Area Income` `Daily Interne…` `Ad Topic
 Line` City
##                 <dbl> <dbl>         <dbl>            <dbl> <chr>
       <chr>
## 1                43.7    28        63127.             173. Front-lin
e bif… Nich…
## 2                73.0    30        71385.             209. Fundament
al mo… Duff…
## 3                51.3    45        67782.             134. Grass-roo
ts co… New …
## 4                51.6    51        42416.             120. Expanded
intan… Sout…
## 5                55.6    19        41921.             188. Proactive
 band… West…
## 6                45.0    26        29876.             178. Virtual 5
thgen… Ronn…
## # … with 4 more variables: Male <dbl>, Country <chr>, Timestamp <dtt
m>,
## #   `Clicked on Ad` <dbl>

# Lists variables in the dataset
names(advert)

##  [1] "Daily Time Spent on Site" "Age"
##  [3] "Area Income"              "Daily Internet Usage"
##  [5] "Ad Topic Line"            "City"
##  [7] "Male"                     "Country"
##  [9] "Timestamp"                "Clicked on Ad"

# Seeing the structure of the dataset
str(advert)

## spec_tbl_df [1,000 × 10] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Daily Time Spent on Site: num [1:1000] 69 80.2 69.5 74.2 68.4 ...
##  $ Age                     : num [1:1000] 35 31 26 29 35 23 33 48 30
 20 ...
##  $ Area Income             : num [1:1000] 61834 68442 59786 54806 73
```

```
890 ...
##  $ Daily Internet Usage    : num [1:1000] 256 194 236 246 226 ...
##  $ Ad Topic Line           : chr [1:1000] "Cloned 5thgeneration orch
estration" "Monitored national standardization" "Organic bottom-line se
rvice-desk" "Triple-buffered reciprocal time-frame" ...
##  $ City                    : chr [1:1000] "Wrightburgh" "West Jodi"
"Davidton" "West Terrifurt" ...
##  $ Male                    : num [1:1000] 0 1 0 1 0 1 0 1 1 1 ...
##  $ Country                 : chr [1:1000] "Tunisia" "Nauru" "San Mar
ino" "Italy" ...
##  $ Timestamp               : POSIXct[1:1000], format: "2016-03-27 00:
53:11" "2016-04-04 01:39:02" ...
##  $ Clicked on Ad           : num [1:1000] 0 0 0 0 0 0 0 1 0 0 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   `Daily Time Spent on Site` = col_double(),
##   ..   Age = col_double(),
##   ..   `Area Income` = col_double(),
##   ..   `Daily Internet Usage` = col_double(),
##   ..   `Ad Topic Line` = col_character(),
##   ..   City = col_character(),
##   ..   Male = col_double(),
##   ..   Country = col_character(),
##   ..   Timestamp = col_datetime(format = ""),
##   ..   `Clicked on Ad` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
#The rows and columns in the data
cat("The dataset has ", dim(advert)[1], "rows and ", dim(advert)[2], "
columns")
```

```
## The dataset has  1000 rows and  10  columns
```

```
#checking the datatypes on the columns
sapply(advert, class)
```

```
## $`Daily Time Spent on Site`
## [1] "numeric"
##
## $Age
## [1] "numeric"
##
## $`Area Income`
## [1] "numeric"
##
## $`Daily Internet Usage`
## [1] "numeric"
##
## $`Ad Topic Line`
## [1] "character"
```

```
##
## $City
## [1] "character"
##
## $Male
## [1] "numeric"
##
## $Country
## [1] "character"
##
## $Timestamp
## [1] "POSIXct" "POSIXt"
##
## $`Clicked on Ad`
## [1] "numeric"
```

```
#summary of the dataset
#Basic descriptive statistics and frequencies.

summary(advert)
```

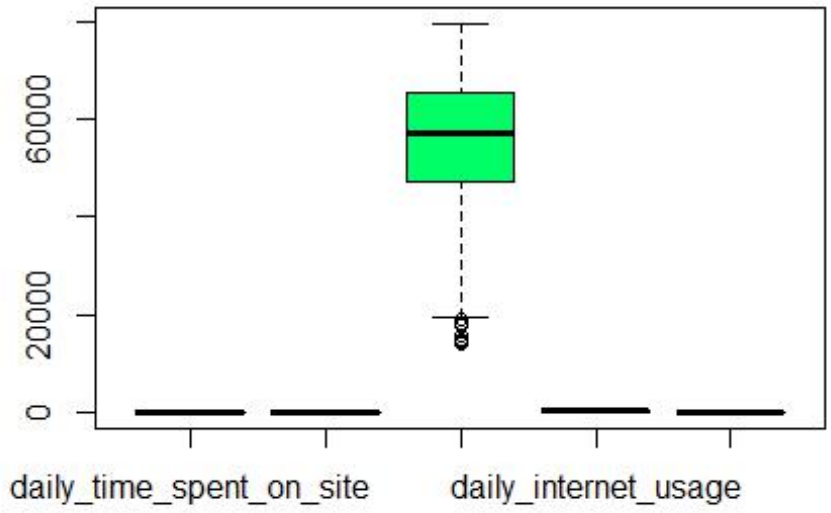
```
##  Daily Time Spent on Site      Age         Area Income    Daily Inte
rnet Usage
##  Min.   :32.60            Min.   :19.00   Min.   :13996   Min.   :10
4.8
##  1st Qu.:51.36            1st Qu.:29.00   1st Qu.:47032   1st Qu.:13
8.8
##  Median :68.22            Median :35.00   Median :57012   Median :18
3.1
##  Mean   :65.00            Mean   :36.01   Mean   :55000   Mean   :18
0.0
##  3rd Qu.:78.55            3rd Qu.:42.00   3rd Qu.:65471   3rd Qu.:21
8.8
##  Max.   :91.43            Max.   :61.00   Max.   :79485   Max.   :27
0.0
##  Ad Topic Line          City                Male         Country

##  Length:1000        Length:1000        Min.   :0.000   Length:1000

##  Class :character   Class :character   1st Qu.:0.000   Class :charac
ter
##  Mode  :character   Mode  :character   Median :0.000   Mode  :charac
ter
##                                        Mean   :0.481

##                                        3rd Qu.:1.000

##                                        Max.   :1.000

##    Timestamp                       Clicked on Ad
```

```
##  Min.   :2016-01-01 02:52:10.00   Min.   :0.0
##  1st Qu.:2016-02-18 02:55:42.00   1st Qu.:0.0
##  Median :2016-04-07 17:27:29.50   Median :0.5
##  Mean   :2016-04-10 10:34:06.64   Mean   :0.5
##  3rd Qu.:2016-05-31 03:18:14.00   3rd Qu.:1.0
##  Max.   :2016-07-24 00:22:16.00   Max.   :1.0

class(advert) #Structure of the columns

## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

Tidying the dataset

```
## Importing packages for plotting

library(tidyr)
library(ggplot2)
library(dplyr)

names(advert)<- tolower(names(advert)) # make the column names to lower
case

names(advert)<- str_replace_all(names(advert), c(" " = "_")) # Replacin
g the white spaces in the column names

head(advert)

## # A tibble: 6 × 10
##   daily_time_spent…   age area_income daily_internet_… ad_topic_line
 city   male
##               <dbl> <dbl>       <dbl>            <dbl> <chr>
 <chr> <dbl>
## 1              69.0    35      61834.             256. Cloned 5thge…
 Wrig…     0
## 2              80.2    31      68442.             194. Monitored na…
 West…     1
## 3              69.5    26      59786.             236. Organic bott…
 Davi…     0
## 4              74.2    29      54806.             246. Triple-buffe…
 West…     1
## 5              68.4    35      73890.             226. Robust logis…
 Sout…     0
## 6              60.0    23      59762.             227. Sharable cli…
 Jami…     1
## # … with 3 more variables: country <chr>, timestamp <dttm>, clicked_
on_ad <dbl>

# Creating a Subset
df <-subset(advert, select = -c(    city,male,  country,ad_topic_line,
timestamp)) # selects Age    ,Area.Income    ,Daily.Internet.Usage,  Cl
icked.on.Ad
print("Modified Data Frame")
```

```
## [1] "Modified Data Frame"

head(df)

## # A tibble: 6 × 5
##   daily_time_spent_on_site   age area_income daily_internet_usage cl
icked_on_ad
##                      <dbl> <dbl>       <dbl>                 <dbl>
      <dbl>
## 1                     69.0    35      61834.                  256.
          0
## 2                     80.2    31      68442.                  194.
          0
## 3                     69.5    26      59786.                  236.
          0
## 4                     74.2    29      54806.                  246.
          0
## 5                     68.4    35      73890.                  226.
          0
## 6                     60.0    23      59762.                  227.
          0

#Checking for outliers

boxplot(df, col = rainbow(ncol(df)))
```



The area income has outliers only

```
# Listing the outliers
OutVals = boxplot(df, plot=FALSE)$out
OutVals
```

```
## [1] 17709.98 18819.34 15598.29 15879.10 14548.06 13996.50 14775.50 1
8368.57
```

```
#Removing outliers
advert1 <-advert

advert1<- advert1[-which(advert1$area_income %in% OutVals),]
#check the difference
print(dim(advert))
```

```
## [1] 1000   10
```

```
print(dim(advert1))
```

```
## [1] 992  10
```

```
#Checking for duplicates in data

duplicated_rows <- advert1[duplicated(advert1),]
duplicated_rows
```

```
## # A tibble: 0 × 10
## # … with 10 variables: daily_time_spent_on_site <dbl>, age <dbl>,
## #   area_income <dbl>, daily_internet_usage <dbl>, ad_topic_line <ch
r>,
## #   city <chr>, male <dbl>, country <chr>, timestamp <dttm>,
## #   clicked_on_ad <dbl>
```

No duplicates

```
# Checking the number of missing per column/variable
colSums(is.na(advert1))
```

```
## daily_time_spent_on_site                      age              area_
income
##                        0                        0
     0
##     daily_internet_usage            ad_topic_line
  city
##                        0                        0
     0
##                     male                  country                tim
estamp
##                        0                        0
     0
##            clicked_on_ad
##                        0
```

No missing data in any column

Exploratory data analysis

Univariate analysis

Measures of Central Tendency

```
#Checking the mean
advert1.dist.mean <- colMeans(subset(advert1, select = c(daily_time_spe
nt_on_site,age,  area_income,    daily_internet_usage)), na.rm = TRUE)
advert1.dist.mean
```

```
## daily_time_spent_on_site                      age              area_
income
##                 65.03979                35.98286              55312.
80720
##     daily_internet_usage
##                179.98504
```

The mean age of those that visited the blog was 35, and on average the daily time spent was 65 minutes, and the average area income of those that visited the blog was 55312 and had average data usage of 179.98

```
# Check for median

advert1.dist.median <- apply(subset(advert1, select = c(daily_time_spen
t_on_site,age,   area_income,    daily_internet_usage)),2,median, na.rm
 = TRUE)
advert1.dist.median
```

```
## daily_time_spent_on_site                      age              area_
income
##                   68.390                  35.000                572
28.185
##     daily_internet_usage
##                  183.425
```

The median age of those that visited the blog was 35, and on median daily time spent was 68 minutes, and the median area income of those that visited the blog was 57228.185 and had a data usage of 183.425

```
#Get the mode

getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

```
attach(advert)
getmode(age);getmode(daily_time_spent_on_site); getmode(area_income);ge
tmode(daily_internet_usage)

## [1] 31

## [1] 62.26

## [1] 61833.9

## [1] 167.22
```

The mode for ages, area income, daily_time_spent on site and daily internet usage was 31 years, 62.26, 61833.9 and 167.22 respectively meaning this were the most frequent characteristics of the users of her blog

Measures of dispersion

```
#check the minimum values of every column
advert.dist.min <- apply(subset(advert1, select = c(daily_time_spent_on
_site,age,   area_income,    daily_internet_usage)),2,min, na.rm = TRUE)
advert.dist.min

## daily_time_spent_on_site                      age              area_
income
##                    32.60                    19.00                 19
345.36
##     daily_internet_usage
##                   104.78
```

The minimum age of those that visit her blog is 19 years and minimum time spent on site is 32 and the minimum area income is 19345.6 and the minimum internet usage is 104.78

```
#check the maximum values of every column
advert.dist.max <- apply(subset(advert1, select = c(daily_time_spent_on
_site,age,   area_income,    daily_internet_usage)),2,max, na.rm = TRUE)
advert.dist.max

## daily_time_spent_on_site                      age              area_
income
##                    91.43                    61.00                 79
484.80
##     daily_internet_usage
##                   269.96
```

The maximum age of those that visit her blog is 61 years and maximum time spent on site is 91.43 and the maximum area income is 79484.8 and the maximum internet usage is 269.69

```
#check the range of  values of every column
advert.dist.range <- apply(subset(advert1, select = c(daily_time_spent_
```

```
on_site, age,    area_income,    daily_internet_usage)),2,range, na.rm
= TRUE)
advert.dist.range
```

```
##      daily_time_spent_on_site age area_income daily_internet_usage
## [1,]                    32.60  19    19345.36                104.78
## [2,]                    91.43  61    79484.80                269.96
```

```
#check the quantiles  values of every column
advert.dist.quantiles <- apply(subset(advert1, select = c(daily_time_sp
ent_on_site,age, area_income,    daily_internet_usage)),2,quantile, na.
rm = TRUE)
advert.dist.quantiles
```

```
##      daily_time_spent_on_site age area_income daily_internet_usage
## 0%                     32.600  19    19345.36             104.7800
## 25%                    51.285  29    47332.82             138.6475
## 50%                    68.390  35    57228.18             183.4250
## 75%                    78.585  42    65518.96             218.8425
## 100%                   91.430  61    79484.80             269.9600
```

```
#check the variation of  values of every column
#The variance is a numerical measure of how the data values is disperse
d around the mean.

advert.dist.variance <- apply(subset(advert1, select = c(daily_time_spe
nt_on_site, age, area_income,    daily_internet_usage)),2,var, na.rm =
TRUE)
advert.dist.variance
```

```
## daily_time_spent_on_site                      age              area_
income
##             2.528609e+02             7.745379e+01             1.6913
76e+08
##     daily_internet_usage
##             1.938785e+03
```

```
#check the standard deviation of  values of every column
#Standard deviation tells you how spread out the data is. It is a measu
re of how far each observed value is from the mean.

advert.dist.sd <- apply(subset(advert1, select = c(daily_time_spent_on_
site,age,    area_income,    daily_internet_usage)),2,sd, na.rm = TRUE)
advert.dist.sd
```

```
## daily_time_spent_on_site                      age              area_
income
##                15.901600                 8.800784             13005.
290554
##     daily_internet_usage
##                44.031632
```

Area income values are highly spread out from the mean

```
#check the skewness every column
install.packages("moments", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/Lenovo/AppData/Local/R/win-library
/4.2'
## (as 'lib' is unspecified)
```

```
## package 'moments' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Lenovo\AppData\Local\Temp\RtmpID6235\downloaded_packages
```

```
library(moments)
advert.dist.skewness <- apply(subset(advert1, select = c(daily_time_spe
nt_on_site,age,  area_income,    daily_internet_usage)),2,skewness, na.
rm = TRUE)
advert.dist.skewness
```

```
## daily_time_spent_on_site                      age              area_
income
##              -0.37679250               0.48509707              -0.57
508362
##     daily_internet_usage
##              -0.03390524
```

if the peak is toward the right and the left tail is longer, we say that the distribution is skewed left or negatively skewed.

Area income is moderately negatively skewed

Daily time spent, age and daily internet usage is approximately symmetric

```
#check the kurtosis every column
#install.packages("moments", repos = "http://cran.us.r-project.org")
#library(moments)
advert.dist.kurtosis <- apply(subset(advert1, select = c(daily_time_spe
nt_on_site,age,  area_income,    daily_internet_usage)),2,kurtosis, na.
rm = TRUE)
advert.dist.kurtosis
```

```
## daily_time_spent_on_site                      age              area_
income
##                 1.898712                 2.599489                 2.
708115
##     daily_internet_usage
##                 1.719177
```

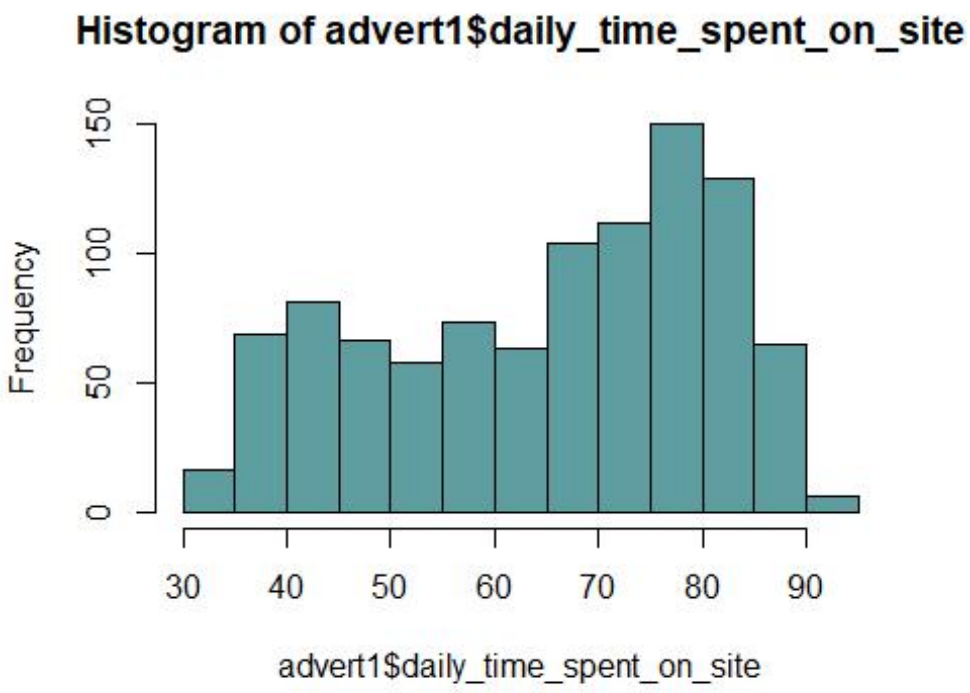
A distribution with kurtosis <3 like for all our variables is called platykurtic. Compared to a normal distribution, its tails are shorter and thinner, and often its central peak is lower and broader.

The reason for this is because the extreme values are less than that of the normal distribution.

Univariate graphical

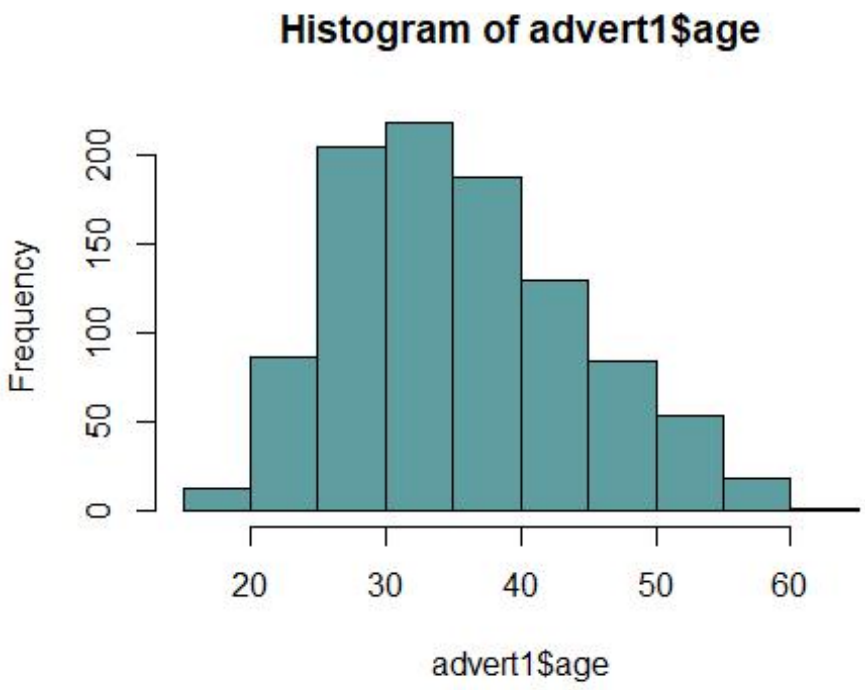
A histogram shows the frequency distribution of a quantitative variable. The area of each bar is equal to the frequency of items found in each class.

```
#see the daily_time_spent_on_site distribution
hist(advert1$daily_time_spent_on_site, col='cadetblue')
```



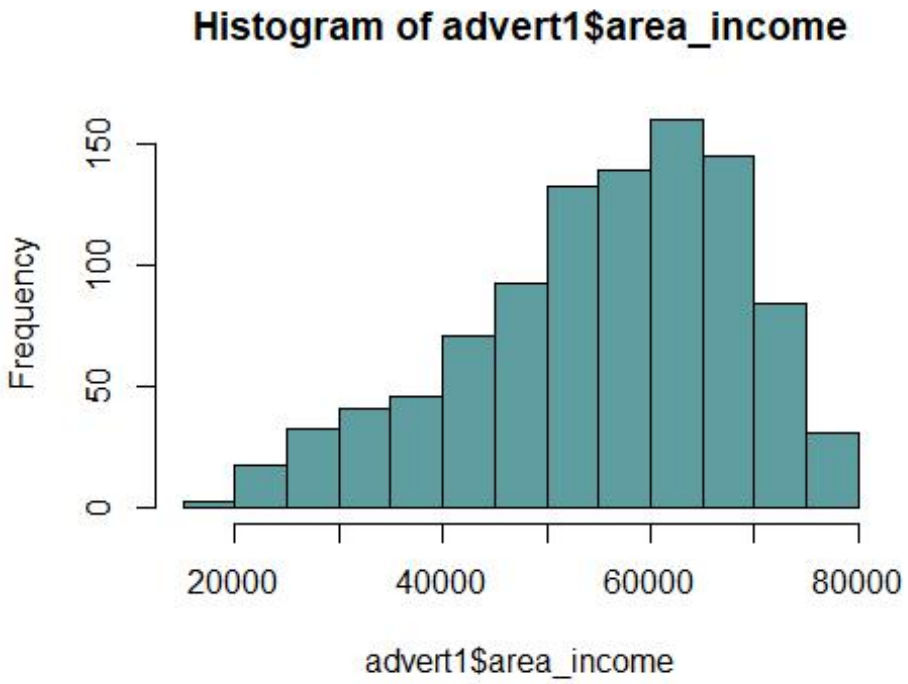
Most people spend around 70-85 daily time on the blog

```
#See the age distribution
hist(advert1$age, col='cadetblue')
```



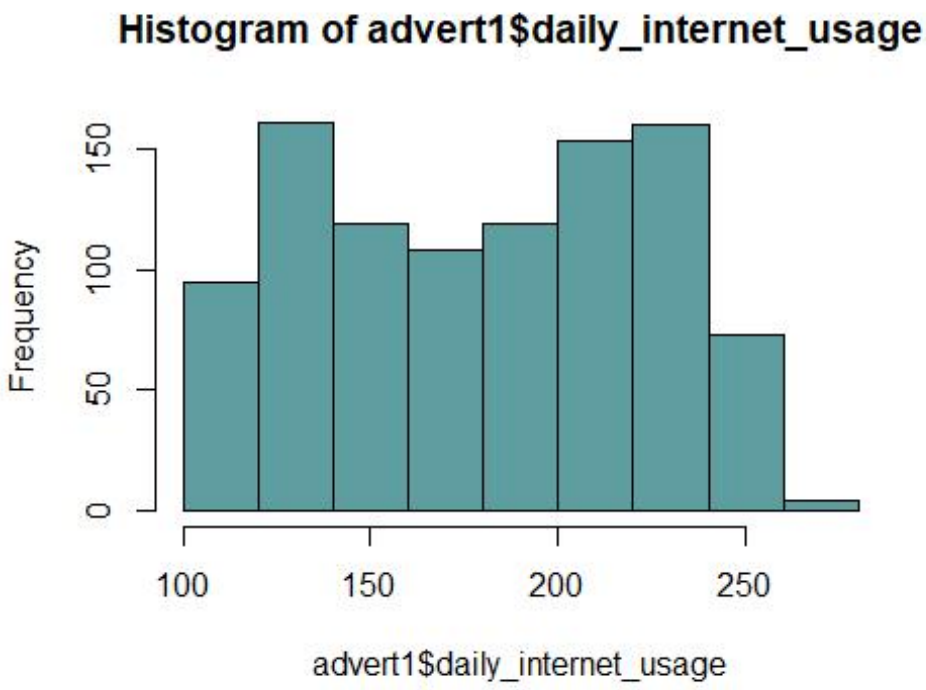
Most people that spend time on the blog are between 25-35 years

```
#See the area_income distribution
hist(advert1$area_income, col='cadetblue')
```



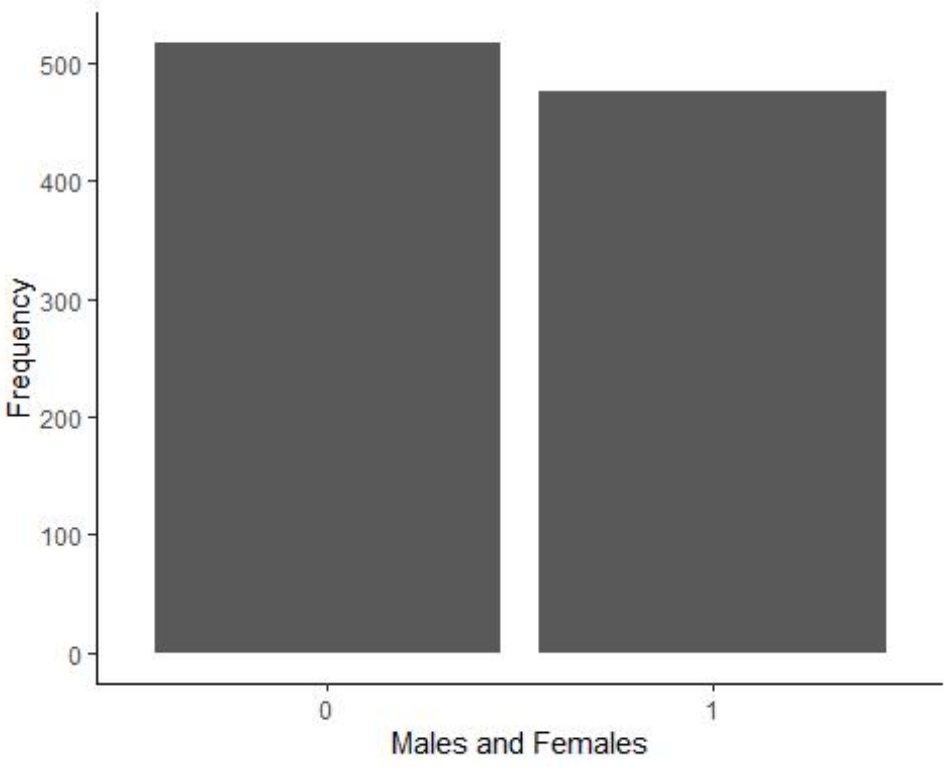
Most people that spend time on the blog have an area income of 5000-7000

```
#See the daily_internet_usage distribution
hist(advert1$daily_internet_usage, col='cadetblue')
```



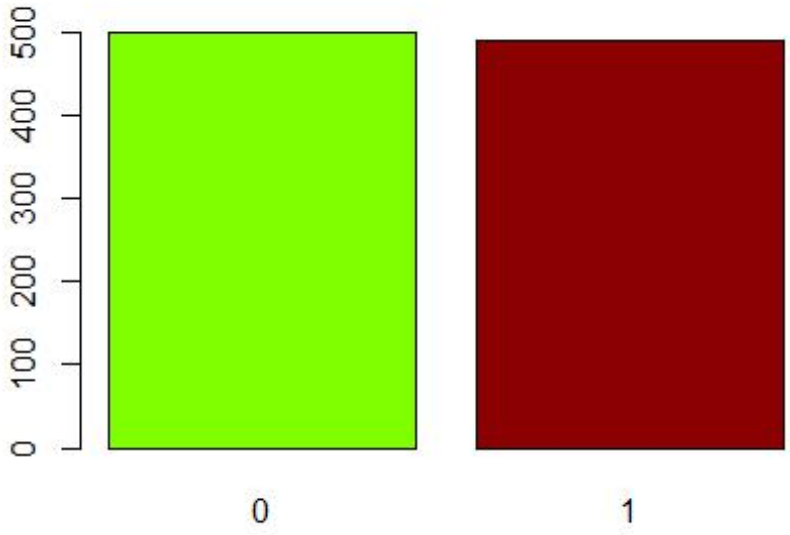
Most daily internet usage in the blog is around 120-140 and 220-240

```
ggplot(advert1,aes(x=toupper(male)))+geom_bar()+xlab(label = "Males and
 Females")+ylab(label = "Frequency")+theme_classic()
```



female is 0 hence most people on her blog are female

```
# Getting specific columns which is clicked on ad
clickad <- advert1$clicked_on_ad

# Applying the table() function will compute the frequency distribution
 of the male variable
# ---
#
clicked_ad_frequency <- table(clickad)


# Then applying the barplot function to produce its bar graph
# ---
#
barplot(clicked_ad_frequency, col=c("chartreuse", "red4"))
```



The number of people that click the ad on the blog are almost equal

```
#Distribution of the countries
table(advert1$country)
```

```
##
##                                        Afghanistan
##                                                  8
##                                            Albania
##                                                  7
##                                            Algeria
##                                                  5
##                                     American Samoa
##                                                  5
##                                            Andorra
##                                                  2
##                                             Angola
##                                                  4
##                                           Anguilla
##                                                  6
##        Antarctica (the territory South of 60 deg S)
##                                                  3
##                                  Antigua and Barbuda
##                                                  5
##                                          Argentina
##                                                  2
##                                            Armenia
```

```
##                                                 3
##                                             Aruba
##                                                 1
##                                         Australia
##                                                 8
##                                           Austria
##                                                 5
##                                        Azerbaijan
##                                                 2
##                                           Bahamas
##                                                 7
##                                           Bahrain
##                                                 5
##                                        Bangladesh
##                                                 4
##                                          Barbados
##                                                 5
##                                           Belarus
##                                                 6
##                                           Belgium
##                                                 5
##                                            Belize
##                                                 4
##                                             Benin
##                                                 2
##                                           Bermuda
##                                                 1
##                                            Bhutan
##                                                 2
##                                           Bolivia
##                                                 6
##                             Bosnia and Herzegovina
##                                                 7
##                          Bouvet Island (Bouvetoya)
##                                                 5
##                                            Brazil
##                                                 5
## British Indian Ocean Territory (Chagos Archipelago)
##                                                 1
##                             British Virgin Islands
##                                                 3
##                                  Brunei Darussalam
##                                                 5
##                                           Bulgaria
##                                                 6
##                                       Burkina Faso
##                                                 4
##                                            Burundi
##                                                 7
##                                           Cambodia
```

```
##                                                 7
##                                          Cameroon
##                                                 5
##                                            Canada
##                                                 5
##                                        Cape Verde
##                                                 1
##                                     Cayman Islands
##                                                 5
##                             Central African Republic
##                                                 2
##                                              Chad
##                                                 4
##                                             Chile
##                                                 4
##                                             China
##                                                 6
##                                   Christmas Island
##                                                 6
##                                          Colombia
##                                                 2
##                                           Comoros
##                                                 2
##                                             Congo
##                                                 4
##                                       Cook Islands
##                                                 3
##                                        Costa Rica
##                                                 6
##                                     Cote d'Ivoire
##                                                 4
##                                           Croatia
##                                                 6
##                                              Cuba
##                                                 5
##                                            Cyprus
##                                                 8
##                                     Czech Republic
##                                                 9
##                                            Denmark
##                                                 3
##                                          Djibouti
##                                                 2
##                                          Dominica
##                                                 5
##                                 Dominican Republic
##                                                 4
##                                            Ecuador
##                                                 5
##                                              Egypt
```

```
##                                                 5
##                                       El Salvador
##                                                 5
##                                 Equatorial Guinea
##                                                 4
##                                           Eritrea
##                                                 7
##                                           Estonia
##                                                 3
##                                          Ethiopia
##                                                 7
##                           Falkland Islands (Malvinas)
##                                                 4
##                                      Faroe Islands
##                                                 3
##                                              Fiji
##                                                 7
##                                           Finland
##                                                 5
##                                            France
##                                                 9
##                                      French Guiana
##                                                 4
##                                   French Polynesia
##                                                 5
##                           French Southern Territories
##                                                 5
##                                             Gabon
##                                                 6
##                                            Gambia
##                                                 2
##                                           Georgia
##                                                 4
##                                           Germany
##                                                 1
##                                             Ghana
##                                                 4
##                                         Gibraltar
##                                                 3
##                                            Greece
##                                                 8
##                                         Greenland
##                                                 5
##                                           Grenada
##                                                 4
##                                        Guadeloupe
##                                                 2
##                                              Guam
##                                                 4
##                                         Guatemala
```

```
##                                                 4
##                                          Guernsey
##                                                 3
##                                            Guinea
##                                                 3
##                                     Guinea-Bissau
##                                                 2
##                                            Guyana
##                                                 5
##                                             Haiti
##                                                 2
##                   Heard Island and McDonald Islands
##                                                 3
##                       Holy See (Vatican City State)
##                                                 3
##                                          Honduras
##                                                 5
##                                         Hong Kong
##                                                 6
##                                           Hungary
##                                                 6
##                                           Iceland
##                                                 3
##                                             India
##                                                 2
##                                         Indonesia
##                                                 6
##                                              Iran
##                                                 5
##                                           Ireland
##                                                 3
##                                         Isle of Man
##                                                 3
##                                            Israel
##                                                 4
##                                             Italy
##                                                 5
##                                           Jamaica
##                                                 5
##                                             Japan
##                                                 4
##                                            Jersey
##                                                 5
##                                            Jordan
##                                                 1
##                                        Kazakhstan
##                                                 4
##                                             Kenya
##                                                 4
##                                          Kiribati
```

```
##                                                 1
##                                             Korea
##                                                 5
##                                            Kuwait
##                                                 2
##                                    Kyrgyz Republic
##                                                 6
##                     Lao People's Democratic Republic
##                                                 4
##                                             Latvia
##                                                 4
##                                            Lebanon
##                                                 5
##                                            Lesotho
##                                                 1
##                                            Liberia
##                                                 8
##                             Libyan Arab Jamahiriya
##                                                 4
##                                      Liechtenstein
##                                                 6
##                                          Lithuania
##                                                 3
##                                         Luxembourg
##                                                 6
##                                              Macao
##                                                 3
##                                           Macedonia
##                                                 2
##                                         Madagascar
##                                                 6
##                                             Malawi
##                                                 4
##                                           Malaysia
##                                                 3
##                                           Maldives
##                                                 4
##                                               Mali
##                                                 4
##                                              Malta
##                                                 6
##                                    Marshall Islands
##                                                 1
##                                         Martinique
##                                                 4
##                                         Mauritania
##                                                 2
##                                          Mauritius
##                                                 4
##                                            Mayotte
```

```
##                                                 6
##                                            Mexico
##                                                 6
##                                        Micronesia
##                                                 8
##                                           Moldova
##                                                 6
##                                            Monaco
##                                                 3
##                                          Mongolia
##                                                 6
##                                        Montenegro
##                                                 2
##                                        Montserrat
##                                                 1
##                                           Morocco
##                                                 3
##                                        Mozambique
##                                                 1
##                                           Myanmar
##                                                 5
##                                           Namibia
##                                                 2
##                                             Nauru
##                                                 3
##                                             Nepal
##                                                 3
##                                       Netherlands
##                                                 4
##                            Netherlands Antilles
##                                                 6
##                                   New Caledonia
##                                                 2
##                                     New Zealand
##                                                 4
##                                       Nicaragua
##                                                 3
##                                           Niger
##                                                 3
##                                            Niue
##                                                 3
##                                  Norfolk Island
##                                                 5
##                       Northern Mariana Islands
##                                                 3
##                                          Norway
##                                                 2
##                                        Pakistan
##                                                 5
##                                           Palau
```

```
##                                          4
##                            Palestinian Territory
##                                          3
##                                     Panama
##                                          2
##                                 Papua New Guinea
##                                          5
##                                   Paraguay
##                                          3
##                                      Peru
##                                          8
##                                Philippines
##                                          6
##                             Pitcairn Islands
##                                          2
##                                    Poland
##                                          6
##                                   Portugal
##                                          3
##                                Puerto Rico
##                                          6
##                                     Qatar
##                                          6
##                                  Reunion
##                                          2
##                                  Romania
##                                          1
##                             Russian Federation
##                                          3
##                                   Rwanda
##                                          5
##                               Saint Barthelemy
##                                          2
##                                   Saint Helena
##                                          5
##                              Saint Kitts and Nevis
##                                          1
##                                    Saint Lucia
##                                          2
##                                   Saint Martin
##                                          4
##                          Saint Pierre and Miquelon
##                                          5
##                  Saint Vincent and the Grenadines
##                                          6
##                                       Samoa
##                                          6
##                                    San Marino
##                                          3
##                               Sao Tome and Principe
```

```
##                                                 2
##                                      Saudi Arabia
##                                                 4
##                                           Senegal
##                                                 8
##                                            Serbia
##                                                 5
##                                        Seychelles
##                                                 3
##                                      Sierra Leone
##                                                 2
##                                         Singapore
##                                                 6
##                           Slovakia (Slovak Republic)
##                                                 2
##                                          Slovenia
##                                                 1
##                                           Somalia
##                                                 5
##                                      South Africa
##                                                 8
##        South Georgia and the South Sandwich Islands
##                                                 2
##                                             Spain
##                                                 3
##                                         Sri Lanka
##                                                 4
##                                             Sudan
##                                                 2
##                                          Suriname
##                                                 2
##                         Svalbard & Jan Mayen Islands
##                                                 6
##                                         Swaziland
##                                                 2
##                                            Sweden
##                                                 4
##                                       Switzerland
##                                                 4
##                                 Syrian Arab Republic
##                                                 3
##                                             Taiwan
##                                                 7
##                                        Tajikistan
##                                                 2
##                                          Tanzania
##                                                 3
##                                          Thailand
##                                                 4
##                                       Timor-Leste
```

```
##                                                 5
##                                              Togo
##                                                 3
##                                           Tokelau
##                                                 4
##                                             Tonga
##                                                 5
##                                  Trinidad and Tobago
##                                                 3
##                                           Tunisia
##                                                 4
##                                            Turkey
##                                                 8
##                                       Turkmenistan
##                                                 6
##                           Turks and Caicos Islands
##                                                 5
##                                            Tuvalu
##                                                 4
##                                            Uganda
##                                                 4
##                                           Ukraine
##                                                 5
##                               United Arab Emirates
##                                                 6
##                                     United Kingdom
##                                                 3
##                  United States Minor Outlying Islands
##                                                 4
##                              United States of America
##                                                 5
##                         United States Virgin Islands
##                                                 4
##                                           Uruguay
##                                                 5
##                                         Uzbekistan
##                                                 2
##                                            Vanuatu
##                                                 6
##                                          Venezuela
##                                                 7
##                                            Vietnam
##                                                 3
##                                    Wallis and Futuna
##                                                 4
##                                       Western Sahara
##                                                 7
##                                              Yemen
##                                                 3
##                                             Zambia
```

```
##                                                 4
##                                          Zimbabwe
##                                                 6

tt = table(advert1$country)
max(tt)

## [1] 9
```

France have the most people that visited the blog

```
#Distribution of cities
table(advert1$city)

##
##               Adamsbury                Adamside                Adamss
tad
##                       1                       1
  1
##                Alanview           Alexanderfurt           Alexanderv
iew
##                       1                       1
  1
##           Alexandrafort             Alexisland             Aliciat
own
##                       1                       1
  1
##            Alvaradoport            Alvarezland             Amandaf
ort
##                       1                       1
  1
##             Amandahaven             Amandaland                Amyf
urt
##                       1                       1
  1
##                Amyhaven        Andersonchester             Andersonf
urt
##                       1                       1
  1
##             Andersonton          Andrewborough             Andrewmo
uth
##                       1                       1
  1
##              Angelhaven            Anthonyfurt           Ashleyches
ter
##                       1                       1
  1
##             Ashleymouth          Austinborough             Austinl
and
##                       1                       1
  1
```

```
##                       1                       1
  1
##              West James               West Jane         West Jeremys
ide
##                       1                       1
  1
##       West Jessicahaven               West Jodi             West Jos
eph
##                       1                       1
  1
##              West Julia             West Justin           West Katief
urt
##                       1                       1
  1
##          West Kevinfurt              West Lacey            West Leah
ton
##                       1                       1
  1
##        West Lindseybury               West Lisa              West Lu
cas
##                       1                       1
  1
##          West Mariafort        West Melaniefurt       West Melissash
ire
##                       1                       1
  1
##       West Michaelhaven        West Michaelport       West Michaelsh
ire
##                       1                       1
  1
##        West Michaelstad             West Pamela              West Ra
ndy
##                       1                       1
  1
##       West Raymondmouth        West Rhondamouth            West Rica
rdo
##                       1                       1
  1
##            West Richard         West Robertside            West Royt
own
##                       1                       1
  1
##            West Russell               West Ryan           West Saman
tha
##                       1                       1
  1
##            West Shannon             West Sharon              West Sh
aun
##                       2                       1
```

```
  1
##             West Steven             West Sydney             West Tan
ner
##                       2                       1
  1
##              West Tanya          West Terrifurt             West Tho
mas
##                       1                       1
  1
##          West Tinashire        West Travismouth          West Wendyl
and
##                       1                       1
  1
##            West William     West Zacharyborough              Westsh
ire
##                       1                       1
  1
##               Whiteport             Whitneyfort              Wilcoxp
ort
##                       1                       1
  1
##            Williammouth             Williamport         Williamsboro
ugh
##                       1                       1
  1
##            Williamsfort           Williamsmouth            Williamsp
ort
##                       1                       1
  3
##            Williamsside             Williamstad             Wilsonbu
rgh
##                       1                       1
  1
##             Wintersfort               Wongland             Wrightbu
rgh
##                       1                       1
  2
##              Wrightview               Yangside              Youngbu
rgh
##                       1                       1
  1
##               Youngfort                  Yuton             Zacharys
tad
##                       1                       1
  1
##              Zacharyton
##                       1
```

Very few cities had more that one person visiting the site

```
#Distribution of cities
table(advert1$ad_topic_line)
```

```
##
##                        Adaptive 24hour Graphic Interface
##                                                        1
##                           Adaptive asynchronous attitude
##                                                        1
##                    Adaptive context-sensitive application
##                                                        1
##                   Adaptive contextually-based methodology
##                                                        1
##                      Adaptive demand-driven knowledgebase
##                                                        1
##                               Adaptive uniform capability
##                                                        1
##                                Advanced 24/7 productivity
##                                                        1
##                           Advanced 5thgeneration capability
##                                                        1
##                            Advanced didactic conglomeration
##                                                        1
##                   Advanced disintermediate data-warehouse
##                                                        1
##                             Advanced exuding conglomeration
##                                                        1
##                               Advanced full-range migration
##                                                        1
##                               Advanced heuristic firmware
##                                                        1
##                                 Advanced local task-force
##                                                        1
##                       Advanced modular Local Area Network
##                                                        1
##                            Advanced systemic productivity
##                                                        1
##                      Advanced web-enabled standardization
##                                                        1
##                            Ameliorated actuating workforce
##                                                        1
##               Ameliorated bandwidth-monitored contingency
##                                                        1
##                         Ameliorated client-driven forecast
##                                                        1
##                      Ameliorated coherent open architecture
##                                                        1
##              Ameliorated contextually-based collaboration
##                                                        1
##                               Ameliorated discrete extranet
##                                                        1
```

```
##          Vision-oriented asynchronous Internet solution
##                                                        1
##     Vision-oriented attitude-oriented Internet solution
##                                                        1
##                  Vision-oriented bifurcated contingency
##                                                        1
##             Vision-oriented contextually-based extranet
##                                                        1
##                  Vision-oriented human-resource synergy
##                                                        1
##                      Vision-oriented methodical support
##                                                        1
##                   Vision-oriented multi-tasking success
##                                                        1
##                Vision-oriented next generation solution
##                                                        1
##                   Vision-oriented optimizing middleware
##                                                        1
##                     Vision-oriented real-time framework
##                                                        1
##                   Vision-oriented system-worthy forecast
##                                                        1
##                   Vision-oriented uniform knowledgebase
##                                                        1
##                            Visionary analyzing structure
##                                                        1
##                         Visionary asymmetric encryption
##                                                        1
##                     Visionary client-driven installation
##                                                        1
##                  Visionary maximized process improvement
##                                                        1
##                   Visionary mission-critical application
##                                                        1
##                         Visionary multi-tasking alliance
##                                                        1
##                             Visionary reciprocal circuit
##                                                        1
```

Bivariate analysis

```
names(advert1)
```

```
##  [1] "daily_time_spent_on_site" "age"
##  [3] "area_income"              "daily_internet_usage"
##  [5] "ad_topic_line"            "city"
##  [7] "male"                     "country"
##  [9] "timestamp"                "clicked_on_ad"
```

```
#Assigning the each column to the  their variable for easier manipulati
on

age <- advert1$age

daily_time_spent_on_site <- advert1$daily_time_spent_on_site

area_income <- advert1$area_income

daily_internet_usage <- advert1$daily_internet_usage
```

Covariance of various variables

Age and other variables

```
cov(age, daily_time_spent_on_site)

## [1] -46.5009

cov(age, area_income)

## [1] -20614.92

cov(age, daily_internet_usage)

## [1] -142.5798
```

Area income and other variables

```
cov( area_income, daily_time_spent_on_site)

## [1] 65151.28

cov( area_income, daily_internet_usage)

## [1] 200896.3
```

Daily internet usage and daily time spent

```
cov( daily_time_spent_on_site, daily_internet_usage)

## [1] 363.8961
```

Covariance indicates the relationship of two variables whenever one variable changes.

If an increase in one variable results in an increase in the other variable, both variables are said to have a positive covariance.

Area income and daily time spent, area income and daily internet usage and daily internet usage and daily internet usage have positive covariance

Area income and internet usage have the strongest positive relationship

Decreases in one variable also cause a decrease in the other. Both variables move together in the same direction when they change.

Age and the other variables have negative covariance

age and area income have the highest negative relationship

Correlation Coefficient

```
cor(age, daily_time_spent_on_site)

## [1] -0.3322762

cor(age, area_income)

## [1] -0.180111

cor(age, daily_internet_usage)

## [1] -0.3679358

cor( area_income, daily_time_spent_on_site)

## [1] 0.3150374

cor( area_income, daily_internet_usage)

## [1] 0.3508222

cor( daily_time_spent_on_site, daily_internet_usage)

## [1] 0.5197228
```

Age and other variables are weakly negatively linearly related

While daily time spent on site and daily internet usage has the highest positive linearly relationship

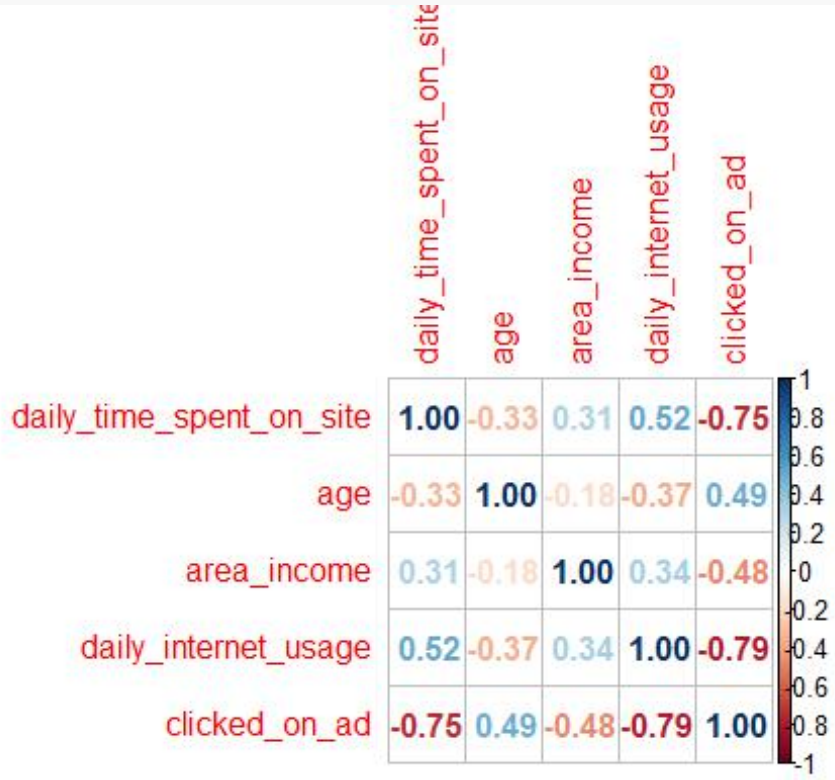
Correlation matrix

```
install.packages("corrplot", repos = "http://cran.us.r-project.org") # 
used to draw correlation matrix

## Installing package into 'C:/Users/Lenovo/AppData/Local/R/win-library
/4.2'
## (as 'lib' is unspecified)

## Warning in download.file(url, destfile, method, mode = "wb", ...): d
ownloaded
## length 3323896 != reported length 3844728

## Warning in download.file(url, destfile, method, mode = "wb", ...): U
RL 'http://
## lib.stat.cmu.edu/R/CRAN/bin/windows/contrib/4.2/corrplot_0.92.zip':
```

```
Timeout of
## 60 seconds was reached

## Error in download.file(url, destfile, method, mode = "wb", ...) :
##   download from 'http://cran.us.r-project.org/bin/windows/contrib/4.
2/corrplot_0.92.zip' failed

## Warning in download.packages(pkgs, destdir = tmpd, available = avail
able, :
## download of package 'corrplot' failed
```

Plotting a correlation matrix

```
M<-cor(df) #find the correlation
library(corrplot)

## corrplot 0.92 loaded

corrplot(M, method="number") #Compute and visualize the correlation coe
fficients
```

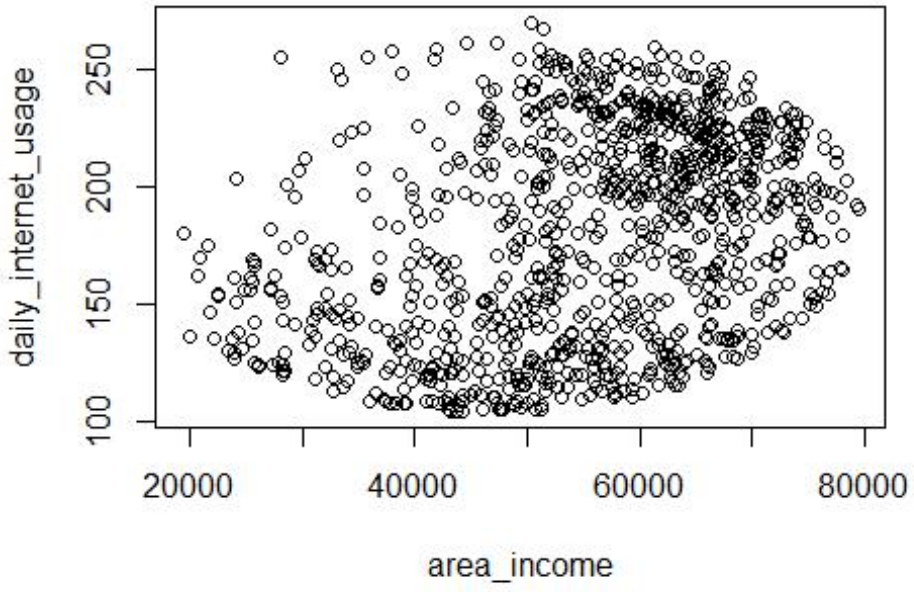


Those with blue have positive correlation coefficient while those in red have negative correlation coefficient

Scatter plot

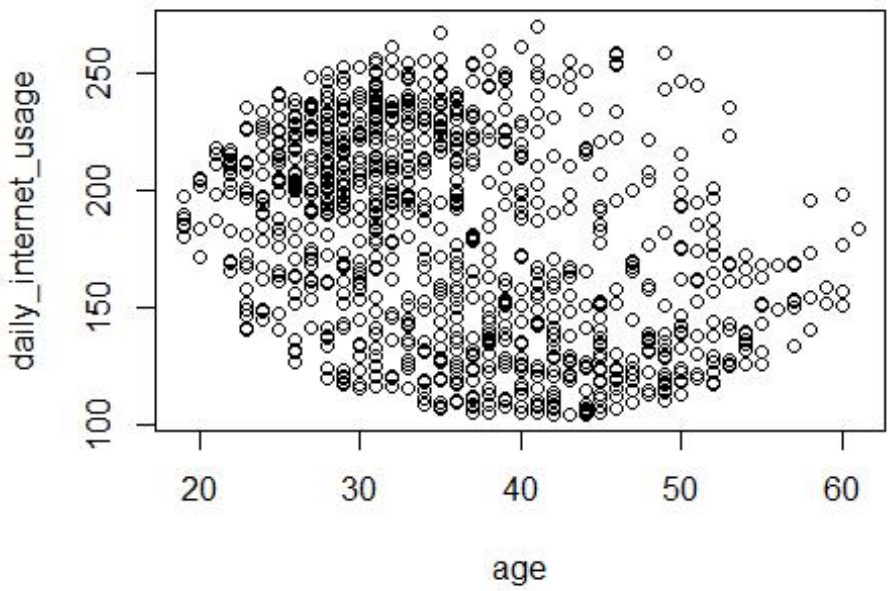
Area income and internet usage

```
plot(area_income,daily_internet_usage,  xlab="area_income", ylab="daily
_internet_usage")
```



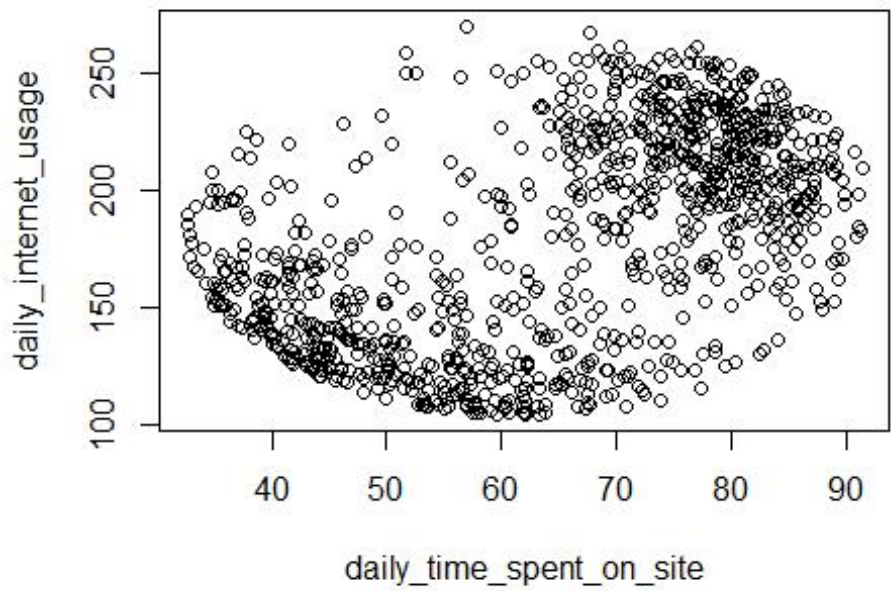
Age and Internet usage

```
plot(age,daily_internet_usage,  xlab="age", ylab="daily_internet_usage")
```



The scatter plots above reveals a weak relationships between area_internet and internet_usage and age and internet_usage

```
plot(daily_time_spent_on_site,daily_internet_usage,  xlab="daily_time_s
pent_on_site", ylab="daily_internet_usage")
```



The scatter plot above reveals a moderate positive relationship between daily_time-spent and internet_usage and age and internet_usage

Modeling

```
head(advert) #overview of dataset
```

```
## # A tibble: 6 × 10
##   daily_time_spent…   age area_income daily_internet_… ad_topic_line
 city   male
##               <dbl> <dbl>       <dbl>            <dbl> <chr>
 <chr> <dbl>
## 1              69.0    35      61834.             256. Cloned 5thge…
 Wrig…     0
## 2              80.2    31      68442.             194. Monitored na…
 West…     1
## 3              69.5    26      59786.             236. Organic bott…
 Davi…     0
## 4              74.2    29      54806.             246. Triple-buffe…
 West…     1
## 5              68.4    35      73890.             226. Robust logis…
 Sout…     0
## 6              60.0    23      59762.             227. Sharable cli…
 Jami…     1
## # … with 3 more variables: country <chr>, timestamp <dttm>, clicked_
on_ad <dbl>
```

```r
#dropping the year, country, city and ad topic line columns
advert$ad_topic_line <- NULL
advert$city <- NULL
advert$country <- NULL
advert$year <- NULL
advert$timestamp <- NULL
head(advert)
```

```
## # A tibble: 6 × 6
##   daily_time_spent_on_si…   age area_income daily_internet_…  male c
licked_on_ad
##                     <dbl> <dbl>       <dbl>            <dbl> <dbl>
       <dbl>
## 1                    69.0    35      61834.             256.     0
          0
## 2                    80.2    31      68442.             194.     1
          0
## 3                    69.5    26      59786.             236.     0
          0
## 4                    74.2    29      54806.             246.     1
          0
## 5                    68.4    35      73890.             226.     0
          0
## 6                    60.0    23      59762.             227.     1
          0
```

```r
advert$clicked_on_ad =as.factor(advert$clicked_on_ad)
head(advert)
```

```
## # A tibble: 6 × 6
##   daily_time_spent_on_si…   age area_income daily_internet_…  male c
licked_on_ad
##                     <dbl> <dbl>       <dbl>            <dbl> <dbl> <
fct>
## 1                    69.0    35      61834.             256.     0 0
## 2                    80.2    31      68442.             194.     1 0
## 3                    69.5    26      59786.             236.     0 0
## 4                    74.2    29      54806.             246.     1 0
## 5                    68.4    35      73890.             226.     0 0
## 6                    60.0    23      59762.             227.     1 0
```

```r
advert$male <- as.numeric(as.character(advert$male))
head(advert)
```

```
## # A tibble: 6 × 6
##   daily_time_spent_on_si…   age area_income daily_internet_…  male c
licked_on_ad
##                     <dbl> <dbl>       <dbl>            <dbl> <dbl> <
fct>
## 1                    69.0    35      61834.             256.     0 0

## 2                    80.2    31      68442.             194.     1 0

## 3                    69.5    26      59786.             236.     0 0

## 4                    74.2    29      54806.             246.     1 0

## 5                    68.4    35      73890.             226.     0 0

## 6                    60.0    23      59762.             227.     1 0
```

```
# Normalizing the dataset so that no particular attribute
# has more impact on modeling algorithm than others.
normalize <- function(x){
  return ((x-min(x)) / (max(x)-min(x)))
}
#data$Age<- normalize(data$Age)
advert$area_income<- normalize(advert$area_income)
advert$daily_internet_usage<- normalize(advert$daily_internet_usage)
advert$daily_time_spent_on_site<- normalize(advert$daily_time_spent_on_
site)
advert$male<- normalize(advert$male)
advert$age<- normalize(advert$age)
head(advert)
```

```
## # A tibble: 6 × 6
##   daily_time_spent_on_s…    age area_income daily_internet_…  male c
licked_on_ad
##                    <dbl>  <dbl>       <dbl>            <dbl> <dbl> <
fct>
## 1                  0.618 0.381        0.730            0.916     0 0

## 2                  0.810 0.286        0.831            0.539     1 0

## 3                  0.627 0.167        0.699            0.797     0 0

## 4                  0.706 0.238        0.623            0.854     1 0

## 5                  0.608 0.381        0.915            0.731     0 0

## 6                  0.466 0.0952       0.699            0.738     1 0
```

```
advert$male <- NULL
head(advert)
```

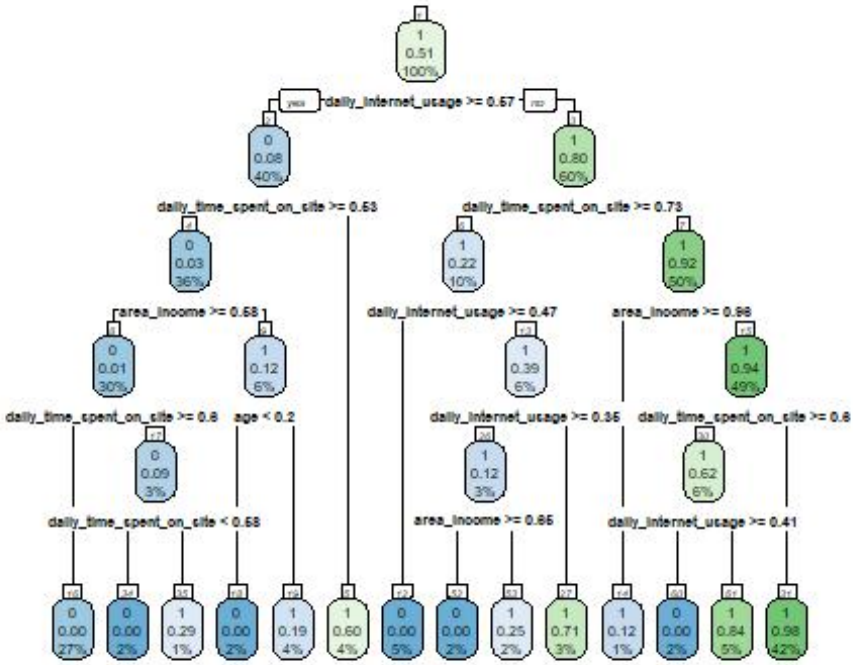
```
## # A tibble: 6 × 5
##   daily_time_spent_on_site    age area_income daily_internet_usage c
licked_on_ad
##                      <dbl>  <dbl>       <dbl>                 <dbl> <
fct>
## 1                    0.618 0.381        0.730                 0.916 0

## 2                    0.810 0.286        0.831                 0.539 0

## 3                    0.627 0.167        0.699                 0.797 0

## 4                    0.706 0.238        0.623                 0.854 0

## 5                    0.608 0.381        0.915                 0.731 0

## 6                    0.466 0.0952       0.699                 0.738 0
```

Decision Trees

```
install.packages("rattle", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/Lenovo/AppData/Local/R/win-library
/4.2'
## (as 'lib' is unspecified)

## Warning in download.file(url, destfile, method, mode = "wb", ...): d
ownloaded
## length 4750996 != reported length 6369685

## Warning in download.file(url, destfile, method, mode = "wb", ...): U
RL 'http://
## lib.stat.cmu.edu/R/CRAN/bin/windows/contrib/4.2/rattle_5.5.1.zip': T
imeout of 60
## seconds was reached

## Error in download.file(url, destfile, method, mode = "wb", ...) :
##   download from 'http://cran.us.r-project.org/bin/windows/contrib/4.
2/rattle_5.5.1.zip' failed

## Warning in download.packages(pkgs, destdir = tmpd, available = avail
able, :
## download of package 'rattle' failed

#Loading libraries
library(rpart,quietly = TRUE)
library(caret,quietly = TRUE)

##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift

library(rpart.plot,quietly = TRUE)
library(rattle)

## Loading required package: bitops

## Rattle: A free graphical interface for data science with R.
## Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.

#data splicing
set.seed(123)
train <- sample(1:nrow(advert),size = ceiling(0.80*nrow(advert)),replac
e = FALSE)
# training set
ad_train <- advert[train,]
# test set
ad_test <- advert[-train,]

#Penalty matrix
penalty.matrix <- matrix(c(0, 1, 10,0), byrow = TRUE, nrow = 2)
#Building our model
tree <- rpart(clicked_on_ad ~., data = ad_train, parms=list(loss=penalt
y.matrix), method = 'class')
tree

## n= 800
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 800 389 1 (0.486250000 0.513750000)
##    2) daily_internet_usage>=0.5662308 319 270 0 (0.915360502 0.08463
9498)
##      4) daily_time_spent_on_site>=0.5281319 289  90 0 (0.968858131 0.
031141869)
##        8) area_income>=0.5787783 238  30 0 (0.987394958 0.012605042)

##         16) daily_time_spent_on_site>=0.6013089 215  10 0 (0.9953488
37 0.004651163) *
##         17) daily_time_spent_on_site< 0.6013089 23  20 0 (0.91304347
8 0.086956522)
##           34) daily_time_spent_on_site< 0.5802312 16   0 0 (1.000000
000 0.000000000) *
##           35) daily_time_spent_on_site>=0.5802312 7   5 1 (0.7142857
14 0.285714286) *
##        9) area_income< 0.5787783 51  45 1 (0.882352941 0.117647059)
```

```
##         18) age< 0.202381 19   0 0 (1.000000000 0.000000000) *
##         19) age>=0.202381 32  26 1 (0.812500000 0.187500000) *
##      5) daily_time_spent_on_site< 0.5281319 30  12 1 (0.400000000 0.
600000000) *
##    3) daily_internet_usage< 0.5662308 481  97 1 (0.201663202 0.79833
6798)
##      6) daily_time_spent_on_site>=0.7324494 83  65 1 (0.783132530 0.
216867470)
##       12) daily_internet_usage>=0.4720002 37   0 0 (1.000000000 0.00
0000000) *
##       13) daily_internet_usage< 0.4720002 46  28 1 (0.608695652 0.39
1304348)
##         26) daily_internet_usage>=0.3478932 25  22 1 (0.880000000 0.
120000000)
##           52) area_income>=0.6463641 13   0 0 (1.000000000 0.0000000
00) *
##           53) area_income< 0.6463641 12   9 1 (0.750000000 0.2500000
00) *
##         27) daily_internet_usage< 0.3478932 21   6 1 (0.285714286 0.
714285714) *
##      7) daily_time_spent_on_site< 0.7324494 398  32 1 (0.080402010 0.
919597990)
##       14) area_income>=0.9611263 8   7 1 (0.875000000 0.125000000) *
##       15) area_income< 0.9611263 390  25 1 (0.064102564 0.935897436)

##         30) daily_time_spent_on_site>=0.6013089 50  19 1 (0.38000000
0 0.620000000)
##           60) daily_internet_usage>=0.4080094 13   0 0 (1.000000000
0.000000000) *
##           61) daily_internet_usage< 0.4080094 37   6 1 (0.162162162
0.837837838) *
##         31) daily_time_spent_on_site< 0.6013089 340   6 1 (0.0176470
59 0.982352941) *

#visualizing the tree
rpart.plot(tree, nn=TRUE)
```




```
#making predictions with our model
pred <- predict(object = tree, ad_test[,-5], type = 'class')
#calculating accuracy
t <- table(ad_test$clicked_on_ad, pred)
confusionMatrix(t)
```

```
## Confusion Matrix and Statistics
##
##    pred
##      0  1
##   0 88 23
##   1  1 88
##
##                Accuracy : 0.88
##                  95% CI : (0.8267, 0.9216)
##     No Information Rate : 0.555
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7629
##
##  Mcnemar's Test P-Value : 1.814e-05
##
##             Sensitivity : 0.9888
##             Specificity : 0.7928
##          Pos Pred Value : 0.7928
##          Neg Pred Value : 0.9888
##              Prevalence : 0.4450
```

```
##          Detection Rate : 0.4400
##    Detection Prevalence : 0.5550
##       Balanced Accuracy : 0.8908
##
##        'Positive' Class : 0
##
```

Challenging the solution

SVM

```
library('caret')
intrain <- createDataPartition(y = advert$clicked_on_ad, p= 0.7, list =
 FALSE)
training <- advert[intrain,]
testing <- advert[-intrain,]
dim(training)
```

```
## [1] 700   5
```

```
dim(testing)
```

```
## [1] 300   5
```

```
#building our model
#
trctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
svm_Linear <- train(clicked_on_ad ~., data = training, method = "svmLin
ear",
trControl=trctrl,
preProcess = c("center", "scale"),
tuneLength = 10)
svm_Linear
```

```
## Support Vector Machines with Linear Kernel
##
## 700 samples
##   4 predictor
##   2 classes: '0', '1'
##
## Pre-processing: centered (4), scaled (4)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 630, 630, 630, 630, 630, 630, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.9671429  0.9342857
##
## Tuning parameter 'C' was held constant at a value of 1
```

```r
#making predictions
test_pred <- predict(svm_Linear, newdata = testing)
test_pred
```

```
##   [1] 0 0 0 0 0 1 0 1 0 0 1 0 1 0 0 0 0 1 0 1 1 1 0 1 0 0 1 1 1 1 1
1 0 0 0 1 0
##  [38] 0 0 1 1 0 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 0
1 1 1 1 1 1
##  [75] 0 1 0 1 0 1 1 1 0 0 1 0 1 0 0 0 0 1 0 1 0 0 1 1 1 0 0 0 0 1 0
0 1 0 0 0 0
## [112] 0 0 1 0 0 1 1 1 0 1 0 0 1 1 1 0 1 1 1 1 0 0 1 0 1 0 0 0 0 1 1
0 1 1 0 0 1
## [149] 0 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0 1 0 1 0 1 1 1 1 1 1 1 1 0 1 0
1 0 0 1 0 0
## [186] 1 1 1 0 1 0 0 0 1 0 0 0 1 1 0 1 1 1 0 0 0 0 1 0 1 1 1 1 0 0 0
0 1 0 0 0 1
## [223] 1 1 1 1 0 1 1 0 0 1 1 0 1 1 1 1 0 0 1 0 0 1 1 1 0 1 1 0 0 0 0
1 0 0 0 0 0
## [260] 0 1 0 1 0 1 1 1 0 1 0 0 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 1 0 0
1 1 1 0 1 0
## [297] 1 1 1 1
## Levels: 0 1
```

```r
#checking accuracy of model
confusionMatrix(table(test_pred, testing$clicked_on_ad))
```

```
## Confusion Matrix and Statistics
##
##
## test_pred   0   1
##         0 146   8
##         1   4 142
##
##                Accuracy : 0.96
##                  95% CI : (0.9312, 0.9792)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.92
##
##  Mcnemar's Test P-Value : 0.3865
##
##             Sensitivity : 0.9733
##             Specificity : 0.9467
##          Pos Pred Value : 0.9481
##          Neg Pred Value : 0.9726
##              Prevalence : 0.5000
##          Detection Rate : 0.4867
##    Detection Prevalence : 0.5133
##       Balanced Accuracy : 0.9600
```

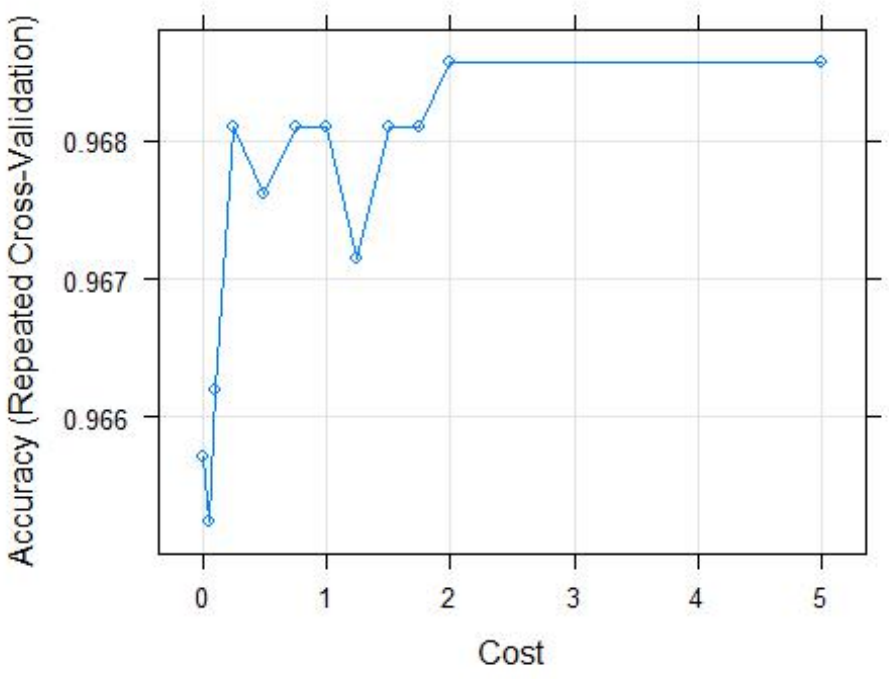
```
##
##        'Positive' Class : 0
##

#Hyper parameter tuning
grid <- expand.grid(C = c(0,0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25,
1.5, 1.75, 2,5))
svm_Linear_Grid <- train(clicked_on_ad ~., data = training, method = "s
vmLinear",
trControl=trctrl,
preProcess = c("center", "scale"),
tuneGrid = grid,
tuneLength = 10)

## Warning: model fit failed for Fold01.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold02.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold03.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold04.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold05.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold06.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold07.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold08.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold09.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters
```

```
## Warning: model fit failed for Fold10.Rep1: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold01.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold02.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold03.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold04.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold05.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold06.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold07.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold08.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold09.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold10.Rep2: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold01.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold02.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters
```

```
## Warning: model fit failed for Fold03.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold04.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold05.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold06.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold07.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold08.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold09.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning: model fit failed for Fold10.Rep3: C=0.00 Error in .local(x,
 ...) :
##   No Support Vectors found. You may want to change your parameters

## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
trainInfo, :
## There were missing values in resampled performance measures.

## Warning in train.default(x, y, weights = w, ...): missing values fou
nd in
## aggregated results

svm_Linear_Grid

## Support Vector Machines with Linear Kernel
##
## 700 samples
##   4 predictor
##   2 classes: '0', '1'
##
## Pre-processing: centered (4), scaled (4)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 630, 630, 630, 630, 630, 630, ...
## Resampling results across tuning parameters:
```

```
##
##   C     Accuracy   Kappa
##   0.00       NaN        NaN
##   0.01  0.9657143  0.9314286
##   0.05  0.9652381  0.9304762
##   0.10  0.9661905  0.9323810
##   0.25  0.9680952  0.9361905
##   0.50  0.9676190  0.9352381
##   0.75  0.9680952  0.9361905
##   1.00  0.9680952  0.9361905
##   1.25  0.9671429  0.9342857
##   1.50  0.9680952  0.9361905
##   1.75  0.9680952  0.9361905
##   2.00  0.9685714  0.9371429
##   5.00  0.9685714  0.9371429
##
## Accuracy was used to select the optimal model using the largest valu
e.
## The final value used for the model was C = 2.
```

```
plot(svm_Linear_Grid)
```




```
#Making predictions with the model after tuning.
test_pred_grid <- predict(svm_Linear_Grid, newdata = testing)
test_pred_grid
```

```
##   [1] 0 0 0 0 0 1 0 1 0 0 1 0 1 0 0 0 0 1 0 1 1 1 0 1 0 0 1 1 1 1 1
1 0 0 0 1 0
##  [38] 0 0 1 1 0 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 0
1 1 1 1 1 1
##  [75] 0 1 0 1 0 1 1 1 0 0 1 0 1 0 0 0 0 1 0 1 0 0 1 1 1 0 0 0 0 1 0
0 1 0 0 0 0
## [112] 0 0 1 0 0 1 1 1 0 1 0 0 1 1 1 0 1 1 1 1 0 0 1 0 1 0 0 0 0 1 1
0 1 1 0 0 1
## [149] 0 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0 1 0 1 0 1 1 1 1 1 1 1 1 0 1 0
1 0 0 1 0 0
## [186] 1 1 1 0 1 0 0 0 1 0 0 0 1 1 0 1 1 1 0 0 0 0 1 0 1 1 1 1 0 0 0
0 1 0 0 0 1
## [223] 1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 0 0 1 0 0 1 1 1 0 1 1 0 0 0 0
1 0 0 0 0 0
## [260] 0 1 0 1 0 1 1 1 0 1 0 0 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 1 0 0
1 1 1 0 1 0
## [297] 1 1 1 1
## Levels: 0 1
```

```
#checking the accuracy
confusionMatrix(table(test_pred_grid, testing$clicked_on_ad))
```

```
## Confusion Matrix and Statistics
##
##
## test_pred_grid   0   1
##              0 145   8
##              1   5 142
##
##                Accuracy : 0.9567
##                  95% CI : (0.927, 0.9767)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9133
##
##  Mcnemar's Test P-Value : 0.5791
##
##             Sensitivity : 0.9667
##             Specificity : 0.9467
##          Pos Pred Value : 0.9477
##          Neg Pred Value : 0.9660
##              Prevalence : 0.5000
##          Detection Rate : 0.4833
##    Detection Prevalence : 0.5100
##       Balanced Accuracy : 0.9567
##
##        'Positive' Class : 0
##
```

Conclusion

- The age and gender do not determine whether an individual clicks on an ad. This is probably because their interests on the internet are different from what the ad is about.
- Daily time spent on a site has a negative correlation on whether an individual clicks on an ad probably because they are already on the site and are aware of what the ad is about.
- The model created using SVM performs better with an accuracy of 95.6% than the one created using decision trees which has an accuracy of 88.5%.
- Hyper parameter tuning doesn't do much in improving the svm model performance.
- We achieved our metric of success since both our models achieved an accuracy score of above 85%.

Implementing the solution

```
#Create a dataframe that selects those that clicked an ad
yes <-  advert1 %>% filter(advert1$clicked_on_ad  == 1);   # Select tho
se  clicked on ad

#summary of those that clicked the ad
summary(yes)
```

```
##  daily_time_spent_on_site      age         area_income    daily_inte
rnet_usage
##  Min.   :32.60            Min.   :19.00   Min.   :19345   Min.   :10
4.8
##  1st Qu.:42.58            1st Qu.:34.00   1st Qu.:39697   1st Qu.:12
3.3
##  Median :51.27            Median :40.00   Median :49867   Median :13
8.5
##  Mean   :53.03            Mean   :40.35   Mean   :49141   Mean   :14
4.9
##  3rd Qu.:61.92            3rd Qu.:47.00   3rd Qu.:59403   3rd Qu.:16
0.4
##  Max.   :91.37            Max.   :61.00   Max.   :78521   Max.   :27
0.0
##  ad_topic_line          city                male          country

##  Length:492         Length:492         Min.   :0.0000   Length:492

##  Class :character   Class :character   1st Qu.:0.0000   Class :chara
cter
##  Mode  :character   Mode  :character   Median :0.0000   Mode  :chara
cter
##                                        Mean   :0.4573
```

```
##                                          3rd Qu.:1.0000

##                                          Max.   :1.0000

##    timestamp                       clicked_on_ad
##  Min.   :2016-01-01 15:14:24.00   Min.   :1
##  1st Qu.:2016-02-17 23:19:07.25   1st Qu.:1
##  Median :2016-04-07 20:36:22.00   Median :1
##  Mean   :2016-04-10 17:57:40.06   Mean   :1
##  3rd Qu.:2016-05-31 03:18:14.00   3rd Qu.:1
##  Max.   :2016-07-24 00:22:16.00   Max.   :1
```

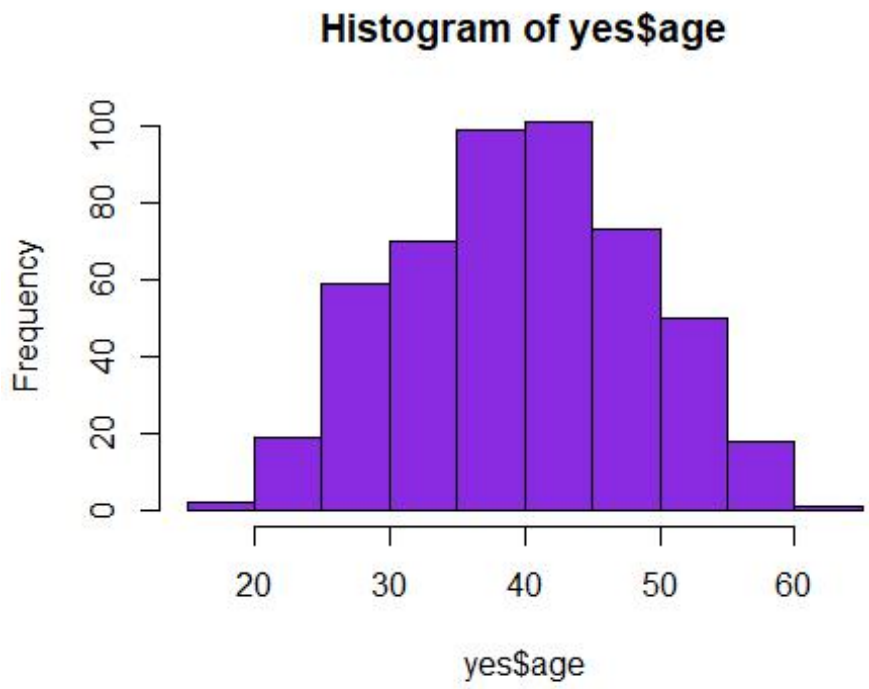
The mean years of those that clicked the ad was 40years

The mean daily time of those that clicked the ad spent was 53

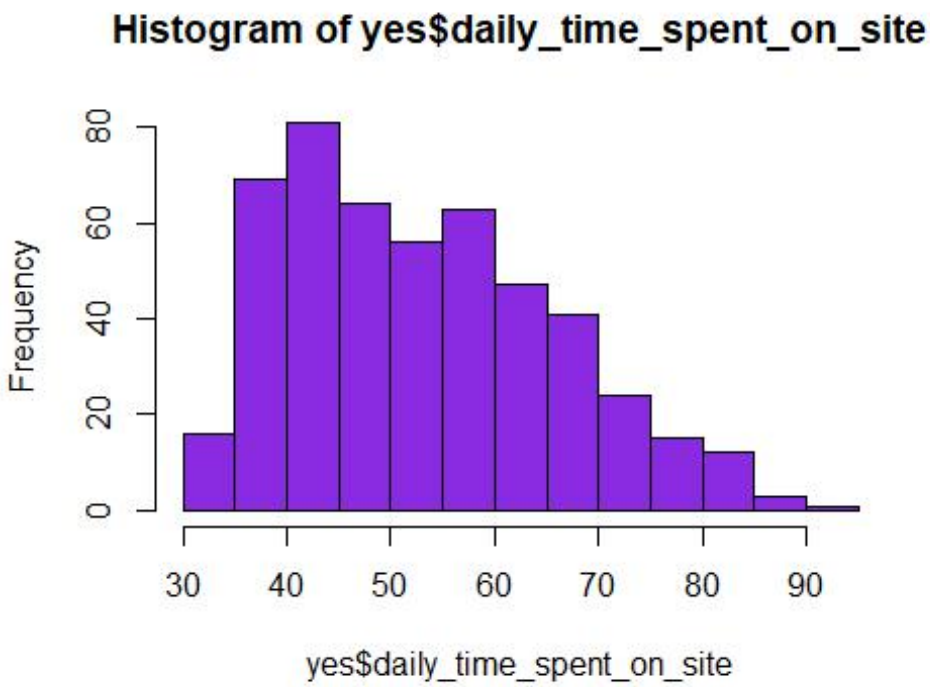
The area income mean of those that clicked the ad was 49141

The daily internet usage mean was 144.9

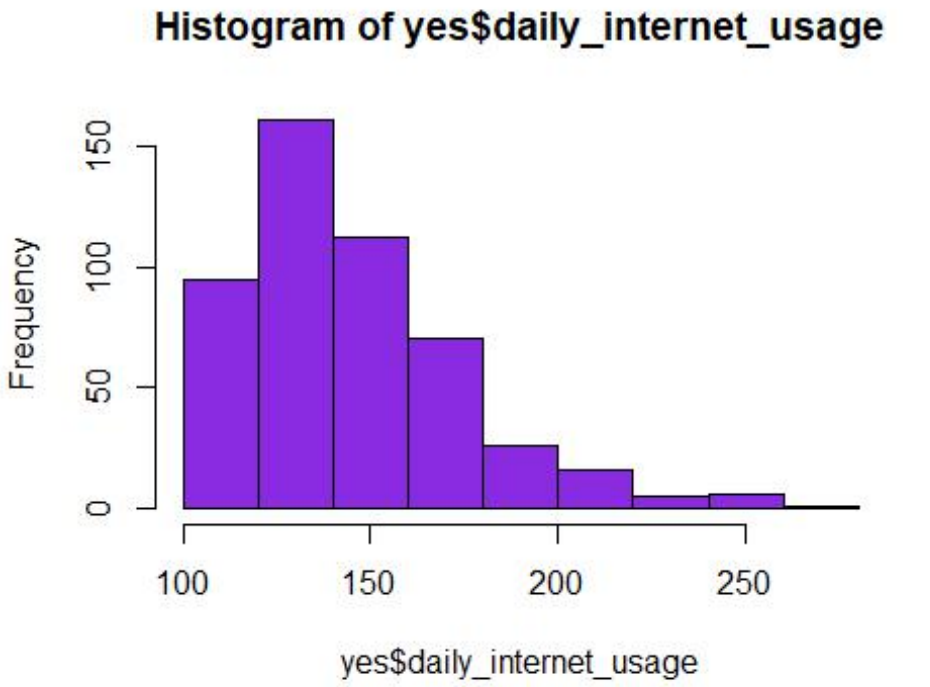
```
#See the age distribution
hist(yes$age, col='blueviolet')
```



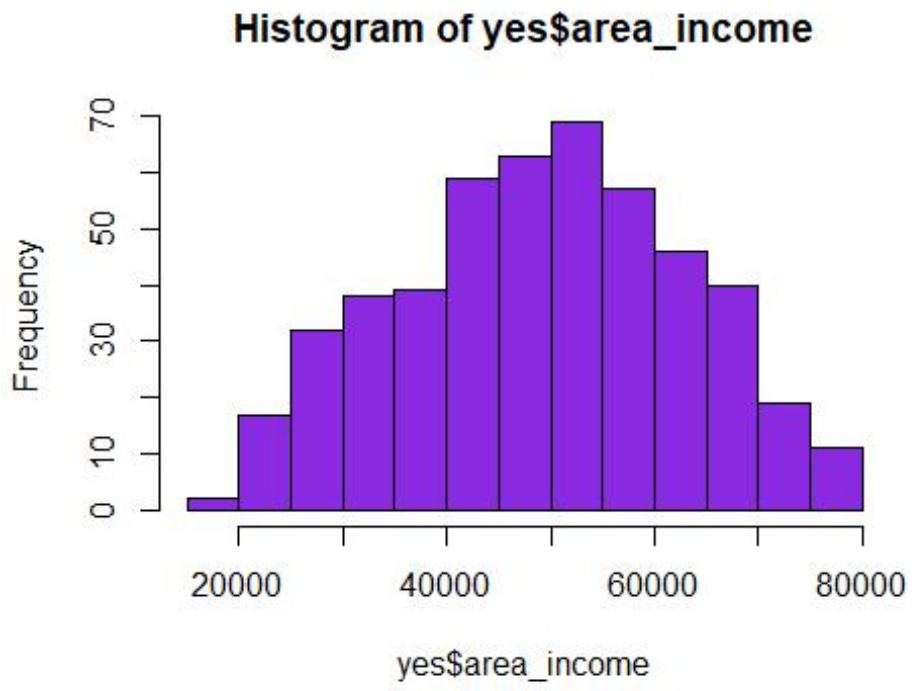

```
#See the daily_time_spent_on_site distribution
hist(yes$daily_time_spent_on_site, col='blueviolet')
```




```r
#See the daily_internet_usage distribution
hist(yes$daily_internet_usage, col='blueviolet')
```



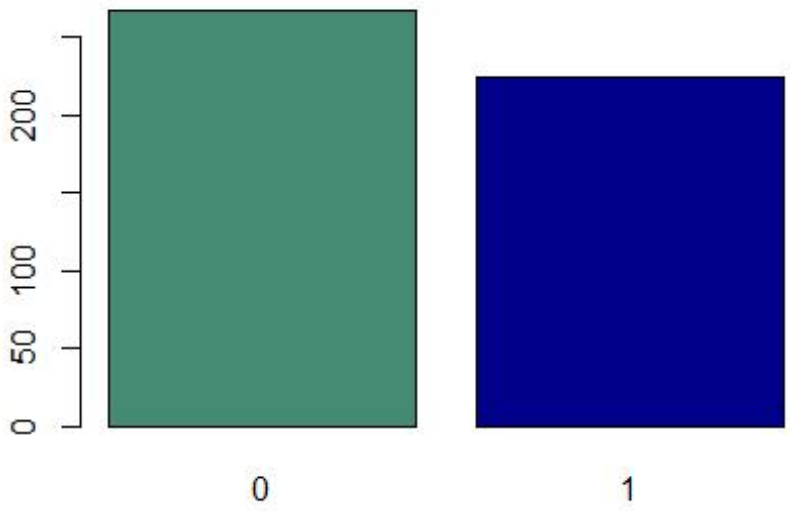
```r
#See the area_income distribution
hist(yes$area_income, col='blueviolet')
```



Those that clicked the ad most

- were around the ages of 30-40
- daily spent time on site was 35-50,
- have a daily internet usage of 100-150
- had an area income 40000-55000

```r
# Getting specific column - male
male1 <- yes$male

# Applying the table() function will compute the frequency distribution
 of the male variable
# ---
#
males_frequency1 <- table(male1)


# Then applying the barplot function to produce its bar graph
# ---
#
barplot(males_frequency1, col=c("aquamarine4", "blue4"))
```



More females clicked on the ad than males Female=0

```
#Distribution of the countries

yy= table(yes$country)

print(max(yy))

## [1] 7

yy

##
##                                          Afghanistan
##                                                    5
##                                              Albania
##                                                    4
##                                              Algeria
##                                                    2
##                                       American Samoa
##                                                    3
##                                              Andorra
##                                                    2
##                                               Angola
##                                                    1
##                                             Anguilla
##                                                    3
```

```
##        Antarctica (the territory South of 60 deg S)
##                                                   2
##                                  Antigua and Barbuda
##                                                   4
##                                           Argentina
##                                                   1
##                                             Armenia
##                                                   1
##                                           Australia
##                                                   7
##                                             Austria
##                                                   1
##                                             Bahamas
##                                                   4
##                                             Bahrain
##                                                   2
##                                          Bangladesh
##                                                   2
##                                            Barbados
##                                                   2
##                                             Belarus
##                                                   3
##                                             Belgium
##                                                   2
##                                              Belize
##                                                   2
##                                               Benin
##                                                   1
##                                              Bhutan
##                                                   1
##                                 Bosnia and Herzegovina
##                                                   3
##                             Bouvet Island (Bouvetoya)
##                                                   2
##                                              Brazil
##                                                   3
## British Indian Ocean Territory (Chagos Archipelago)
##                                                   1
##                                 British Virgin Islands
##                                                   1
##                                      Brunei Darussalam
##                                                   2
##                                             Bulgaria
##                                                   4
##                                          Burkina Faso
##                                                   1
##                                               Burundi
##                                                   2
##                                              Cambodia
##                                                   2
```

```
##                                                 Canada
##                                                      3
##                                           Cayman Islands
##                                                      3
##                                  Central African Republic
##                                                      1
##                                                   Chad
##                                                      2
##                                                  Chile
##                                                      3
##                                                  China
##                                                      4
##                                         Christmas Island
##                                                      4
##                                                Colombia
##                                                      1
##                                                 Comoros
##                                                      1
##                                                   Congo
##                                                      3
##                                              Cook Islands
##                                                      1
##                                                Costa Rica
##                                                      2
##                                             Cote d'Ivoire
##                                                      3
##                                                    Cuba
##                                                      4
##                                                  Cyprus
##                                                      4
##                                             Czech Republic
##                                                      4
##                                                 Denmark
##                                                      2
##                                                Djibouti
##                                                      1
##                                                Dominica
##                                                      2
##                                      Dominican Republic
##                                                      2
##                                                 Ecuador
##                                                      2
##                                                   Egypt
##                                                      3
##                                              El Salvador
##                                                      3
##                                       Equatorial Guinea
##                                                      3
##                                                 Eritrea
##                                                      3
```

```
##                                             Estonia
##                                                   1
##                                            Ethiopia
##                                                   7
##                           Falkland Islands (Malvinas)
##                                                   2
##                                       Faroe Islands
##                                                   2
##                                                Fiji
##                                                   3
##                                             Finland
##                                                   1
##                                              France
##                                                   5
##                                       French Guiana
##                                                   3
##                                    French Polynesia
##                                                   1
##                           French Southern Territories
##                                                   1
##                                              Gambia
##                                                   1
##                                             Georgia
##                                                   2
##                                             Germany
##                                                   1
##                                               Ghana
##                                                   2
##                                              Greece
##                                                   3
##                                           Greenland
##                                                   1
##                                             Grenada
##                                                   2
##                                          Guadeloupe
##                                                   1
##                                                Guam
##                                                   2
##                                           Guatemala
##                                                   3
##                                            Guernsey
##                                                   2
##                                              Guinea
##                                                   2
##                                       Guinea-Bissau
##                                                   1
##                                              Guyana
##                                                   3
##                                               Haiti
##                                                   1
```

```
##                     Heard Island and McDonald Islands
##                                                     2
##                         Holy See (Vatican City State)
##                                                     1
##                                               Honduras
##                                                     2
##                                              Hong Kong
##                                                     4
##                                                Hungary
##                                                     5
##                                                Iceland
##                                                     1
##                                              Indonesia
##                                                     4
##                                                   Iran
##                                                     3
##                                                Ireland
##                                                     1
##                                              Isle of Man
##                                                     1
##                                                 Israel
##                                                     2
##                                                  Italy
##                                                     1
##                                                Jamaica
##                                                     2
##                                                  Japan
##                                                     2
##                                                 Jersey
##                                                     3
##                                             Kazakhstan
##                                                     2
##                                                  Kenya
##                                                     4
##                                               Kiribati
##                                                     1
##                                                  Korea
##                                                     3
##                                                 Kuwait
##                                                     1
##                                          Kyrgyz Republic
##                                                     1
##                       Lao People's Democratic Republic
##                                                     2
##                                                 Latvia
##                                                     4
##                                                Lebanon
##                                                     3
##                                                Liberia
##                                                     6
```

```
##                                Libyan Arab Jamahiriya
##                                                     2
##                                         Liechtenstein
##                                                     6
##                                             Lithuania
##                                                     3
##                                            Luxembourg
##                                                     2
##                                                 Macao
##                                                     3
##                                              Macedonia
##                                                     1
##                                            Madagascar
##                                                     2
##                                                Malawi
##                                                     2
##                                              Maldives
##                                                     2
##                                                  Mali
##                                                     1
##                                                 Malta
##                                                     3
##                                       Marshall Islands
##                                                     1
##                                             Martinique
##                                                     3
##                                            Mauritania
##                                                     1
##                                             Mauritius
##                                                     1
##                                               Mayotte
##                                                     5
##                                                Mexico
##                                                     4
##                                            Micronesia
##                                                     4
##                                               Moldova
##                                                     2
##                                                Monaco
##                                                     1
##                                              Mongolia
##                                                     4
##                                            Montenegro
##                                                     2
##                                            Montserrat
##                                                     1
##                                               Morocco
##                                                     1
##                                               Myanmar
##                                                     1
```

```
##                                          Namibia
##                                                1
##                                            Nauru
##                                                1
##                                      Netherlands
##                                                3
##                           Netherlands Antilles
##                                                2
##                                  New Caledonia
##                                                2
##                                    New Zealand
##                                                2
##                                          Niger
##                                                2
##                                   Norfolk Island
##                                                2
##                          Northern Mariana Islands
##                                                2
##                                         Norway
##                                                1
##                                       Pakistan
##                                                1
##                                          Palau
##                                                2
##                             Palestinian Territory
##                                                2
##                                  Papua New Guinea
##                                                3
##                                       Paraguay
##                                                1
##                                           Peru
##                                                5
##                                    Philippines
##                                                3
##                                  Pitcairn Islands
##                                                1
##                                         Poland
##                                                3
##                                       Portugal
##                                                1
##                                    Puerto Rico
##                                                3
##                                          Qatar
##                                                2
##                                       Romania
##                                                1
##                                  Russian Federation
##                                                1
##                                         Rwanda
##                                                2
```

```
##                                     Saint Barthelemy
##                                                    2
##                                         Saint Helena
##                                                    2
##                                  Saint Kitts and Nevis
##                                                    1
##                                           Saint Lucia
##                                                    1
##                                         Saint Martin
##                                                    2
##                              Saint Pierre and Miquelon
##                                                    3
##                     Saint Vincent and the Grenadines
##                                                    3
##                                               Samoa
##                                                    4
##                                           San Marino
##                                                    1
##                                Sao Tome and Principe
##                                                    2
##                                        Saudi Arabia
##                                                    3
##                                             Senegal
##                                                    5
##                                              Serbia
##                                                    3
##                                            Seychelles
##                                                    1
##                                        Sierra Leone
##                                                    2
##                                           Singapore
##                                                    1
##                                            Slovenia
##                                                    1
##                                             Somalia
##                                                    2
##                                        South Africa
##                                                    6
##        South Georgia and the South Sandwich Islands
##                                                    1
##                                               Spain
##                                                    3
##                                            Suriname
##                                                    1
##                         Svalbard & Jan Mayen Islands
##                                                    4
##                                              Sweden
##                                                    1
##                                         Switzerland
##                                                    3
```

```
##                                 Syrian Arab Republic
##                                                    1
##                                               Taiwan
##                                                    4
##                                           Tajikistan
##                                                    1
##                                             Tanzania
##                                                    1
##                                             Thailand
##                                                    2
##                                           Timor-Leste
##                                                    1
##                                                 Togo
##                                                    1
##                                              Tokelau
##                                                    3
##                                                Tonga
##                                                    2
##                                  Trinidad and Tobago
##                                                    2
##                                              Tunisia
##                                                    1
##                                               Turkey
##                                                    7
##                                           Turkmenistan
##                                                    2
##                              Turks and Caicos Islands
##                                                    3
##                                               Tuvalu
##                                                    3
##                                               Uganda
##                                                    4
##                                              Ukraine
##                                                    1
##                                 United Arab Emirates
##                                                    3
##                                       United Kingdom
##                                                    2
##                  United States Minor Outlying Islands
##                                                    2
##                              United States of America
##                                                    3
##                         United States Virgin Islands
##                                                    2
##                                              Uruguay
##                                                    1
##                                           Uzbekistan
##                                                    1
##                                              Vanuatu
##                                                    1
```

```
##                                          Venezuela
##                                                  3
##                                            Vietnam
##                                                  2
##                                    Wallis and Futuna
##                                                  1
##                                       Western Sahara
##                                                  4
##                                              Yemen
##                                                  2
##                                             Zambia
##                                                  3
##                                           Zimbabwe
##                                                  4
```

Turkey has the highest number of those people that clicked the ad

```
ss= table(yes$city)

print(max(ss))

## [1] 2

ss

##
##               Adamsbury                Adamside           Alexanderf
urt
##                       1                       1
  1
##           Alexanderview             Aliciatown            Alvaradop
ort
##                       1                       1
  1
##             Alvarezland             Amandafort             Amandaha
ven
##                       1                       1
  1
##         Andersonchester            Andersonfurt           Andrewboro
ugh
##                       1                       1
  1
##              Bernardton              Bethburgh            Blairboro
ugh
##                       1                       1
  1
##             Blevinstown             Boyerberg            Bradleybu
rgh
##                       1                       1
  1
##             Bradleyside              Bradyfurt               Brandil
```

```
and
##                         1                         1
  1
##              Brandonbury              Brandymouth              Brendaches
ter
##                         1                         1
  1
##                Brianfurt                Brianland           Brittanyboro
ugh
##                         1                         1
  1
##                Brownport                Brownton                Brownt
own
##                         1                         1
  1
##              Burgessside              Butlerfort              Cameronb
erg
##                         1                         1
  1
##                Carterland           Catherinefort           Cervantessh
ire
##                         1                         1
  1
##              Chapmanmouth             Charlenetown              Charlesp
ort
##                         1                         1
  1
##             Charlottefort               Chaseshire              Chrismo
uth
##                         1                         1
  1
##           Christinehaven           Christinetown           Christopherp
ort
##                         1                         1
  1
##              Clarkborough               Codyburgh                Coleb
ury
##                         1                         1
  1
##                 Combsstad               Costaburgh              Courtneyf
ort
##                         1                         1
  1
##              Crawfordfurt          Cunninghamhaven              Curtisp
ort
##                         1                         1
  1
##                Danielview               Davidmouth                Davids
ide
##                         1                         1
```

```
  1
##               Davidstad               Davidview           Davilaches
ter
##                       1                       1
  1
##               Davisfurt              Debraburgh             Destinyf
urt
##                       1                       1
  1
##              Dianaville               Duffystad             Dustinmo
uth
##                       1                       1
  1
##            East Anthony            East Barbara        East Breannaf
urt
##                       1                       1
  1
##           East Brettton      East Brittanyville              East D
ana
##                       1                       1
  1
##              East Donna              East Eric           East Ericp
ort
##                       1                       1
  1
##         East Georgeside        East Heatherside             East He
idi
##                       1                       1
  1
##              East Jason          East Jessefort              East J
ohn
##                       1                       1
  1
##          East Kevinbury            East Lindsey            East Maur
een
##                       1                       1
  1
##       East Michaelmouth       East Michelleberg              East M
ike
##                       1                       1
  1
##        East Rachaelfurt         East Rachelview             East Ron
ald
##                       1                       1
  1
##      East Samanthashire            East Sharon             East Sh
awn
##                       1                       1
  1
##            East Stephen            East Tammie         East Tylersh
```

```
ire
##                         1                         1
  1
##        East Vincentstad      East Yvonnechester              Edwardsp
ort
##                         1                         1
  1
##           Elizabethbury           Elizabethport           Elizabeths
tad
##                         1                         1
  1
##           Ericksonmouth                 Erinton            Estradash
ire
##                         1                         1
  1
##              Evansville              Florestown              Fosters
ide
##                         1                         1
  1
##                Frankbury                Frankport             Fraziersh
ire
##                         1                         1
  1
##              Garciamouth              Garciaside              Garciav
iew
##                         1                         1
  1
##                Grahamberg                Greerport              Guzmanl
and
##                         1                         1
  1
##             Hamiltonfort              Hannaport              Hansenl
and
##                         1                         1
  1
##              Hansenmouth           Harperborough          Hartmanches
ter
##                         1                         1
  1
##              Harveyport             Hawkinsbury             Heatherb
erg
##                         1                         1
  1
##             Helenborough            Hendrixmouth               Henryl
and
##                         1                         1
  1
##           Hernandezside          Hernandezville             Hintonp
ort
##                         1                         1
```

```
  1
##               Hobbsbury             Holderville             Hubbardmo
uth
##                       1                       1
  1
##          Huffmanchester              Hughesport             Jacksonbu
rgh
##                       1                       1
  1
##             Jacksonstad              Jacobstad         Jacquelinesh
ire
##                       1                       1
  1
##               Jamesfurt             Jamesmouth             Jeffreybu
rgh
##                       1                       1
  1
##            Jeffreymouth           Jenniferhaven           Jensenboro
ugh
##                       1                       1
  1
##               Jensenton             Jeremybury             Jessicaha
ven
##                       1                       1
  1
##            Jessicashire             Jessicastad               Joeches
ter
##                       1                       1
  1
##             Johnsontown             Johnsonview           Johnstonmo
uth
##                       1                       1
  1
##            Jonathanland            Jonathantown               Jonessh
ire
##                       1                       1
  1
##                Joneston             Jordanshire               Jordant
own
##                       1                       1
  1
##              Josephberg             Josephmouth               Josephs
tad
##                       1                       1
  1
##              Julietown               Katieport                Keitht
own
##                       1                       1
  1
##               Kentmouth               Kevinberg           Kimberlymo
```

```
uth
##                       1                       1
  1
##             Kingchester               Klineside             Kristinf
urt
##                       1                       1
  1
##             Kristintown             Kyleborough         Lake Allenvi
lle
##                       1                       1
  1
##             Lake Amanda         Lake Beckyburgh        Lake Brandonv
iew
##                       1                       1
  1
##      Lake Cassandraport      Lake Charlottestad    Lake Christopherf
urt
##                       1                       1
  1
##         Lake Conniefurt          Lake Craigview              Lake Da
vid
##                       1                       1
  2
##             Lake Dustin             Lake Edward           Lake Evant
own
##                       1                       1
  1
##              Lake Faith             Lake Gerald               Lake
Ian
##                       1                       1
  1
##              Lake James           Lake Jennifer            Lake Jess
ica
##                       2                       1
  1
##       Lake Jessicaville             Lake Jesus               Lake J
ohn
##                       1                       1
  1
##           Lake Johnbury              Lake Jose         Lake Joshuaf
urt
##                       1                       1
  1
##            Lake Matthew           Lake Michelle       Lake Michelleb
ury
##                       1                       1
  1
##        Lake Rhondaburgh     Lake Stephenborough             Lake Su
san
##                       1                       1
```

```
  1
##              Lake Tracy            Lake Vanessa             Lawsonsh
ire
##                       1                       1
  1
##                Leahside              Lesliebury              Lewismo
uth
##                       1                       1
  1
##                Lisamouth              Lopezberg              Lorivi
lle
##                       2                       1
  1
##                Lovemouth             Mariahview              Markha
ven
##                       1                       1
  1
##             Masseyshire             Mauricefurt           Meghanches
ter
##                       1                       1
  1
##              Melanieton          Melissachester             Meyerss
tad
##                       1                       1
  1
##                 Mezaton            Michaelshire             Michelet
own
##                       1                       1
  1
##             Michellefort            Michelleside             Millerb
ury
##                       1                       2
  2
##            Millerchester              Millerside             Millert
own
##                       1                       1
  1
##              Monicaview             Morganport             Morrismo
uth
##                       1                       1
  1
##              Nelsonfurt              New Amanda           New Angelv
iew
##                       1                       1
  1
##           New Brendafurt      New Charleschester       New Christinat
own
##                       1                       1
  1
##         New Daniellefort             New Darlene          New Debbies
```

```
tad
##                         1                         1
  1
##          New Denisebury               New Henry                 New
Jay
##                         1                         1
  1
##          New Joshuaport          New Julianberg           New Karenb
erg
##                         1                         1
  1
##          New Keithburgh           New Lindaberg           New Lucasbu
rgh
##                         1                         1
  1
##          New Marcusbury             New Matthew             New Mich
ael
##                         1                         1
  1
##               New Nancy             New Patrick              New Rac
hel
##                         1                         1
  1
##             New Rebecca             New Sabrina              New Ta
mmy
##                         1                         1
  1
##              New Teresa             New Theresa              New Tho
mas
##                         1                         1
  1
##             New Timothy               New Tina          New Travist
own
##                         1                         1
  1
##        New Williammouth           Nicholasland        North Aaronbu
rgh
##                         1                         1
  1
##      North Aaronchester           North Andrew        North Andrews
tad
##                         1                         1
  1
##        North Angelastad        North Angelatown             North Ap
ril
##                         1                         1
  1
##     North Brittanyburgh           North Cassie       North Charlesb
ury
##                         1                         1
```

```
  1
##             North Daniel        North Debrashire        North Derekvi
lle
##                        1                       1
  1
##            North Destiny     North Jenniferburgh      North Jessicavi
lle
##                        1                       1
  1
##          North Johntown          North Jonathan             North Ka
tie
##                        1                       1
  1
##         North Kevinside          North Kimberly         North Laural
and
##                        1                       1
  1
##              North Mark          North Maryland          North Merce
des
##                        1                       1
  1
##            North Michael            North Regina       North Ricardot
own
##                        1                       1
  1
##      North Richardburgh    North Russellborough          North Saman
tha
##                        1                       1
  1
##         North Sarashire          North Virginia               Olsons
ide
##                        1                       1
  1
##               Olsonstad             Palmerside             Pattymo
uth
##                        1                       1
  1
##                Penatown             Perryburgh            Petersonf
urt
##                        1                       1
  1
##              Philipberg            Phillipsbury        Port Angelamo
uth
##                        1                       1
  1
##         Port Aprilville              Port Beth             Port Bl
ake
##                        1                       1
  1
##              Port Brenda             Port Brian         Port Brookel
```

```
and
##                         1                         1
  1
##             Port Cassie          Port Christina        Port Christop
her
##                         1                         1
  1
## Port Christopherborough            Port Crystal             Port Den
nis
##                         1                         1
  1
##          Port Derekberg     Port Douglasborough               Port E
ric
##                         1                         1
  1
##          Port Erikhaven           Port Erinberg         Port Georgeb
ury
##                         1                         1
  1
##              Port Jason        Port Jefferybury            Port Jenni
fer
##                         1                         1
  1
##            Port Jessica         Port Joshuafort               Port J
uan
##                         1                         1
  1
##              Port Julie        Port Katelynview       Port Kathleenf
ort
##                         1                         1
  1
##           Port Lawrence        Port Melissaberg       Port Michaelmo
uth
##                         1                         1
  1
##       Port Michealburgh           Port Mitchell         Port Patrick
ton
##                         1                         1
  1
##           Port Paultown             Port Rachel         Port Sarahha
ven
##                         1                         1
  1
##         Port Sarahshire         Port Sherrystad             Pottermo
uth
##                         1                         1
  1
##            Ramirezhaven            Ramirezside            Rebeccamo
uth
##                         1                         1
```

```
  1
##             Reginamouth            Reneechester             Richardsh
ire
##                       1                       1
  1
##          Richardsonland         Richardsonshire               Rivasl
and
##                       1                       1
  1
##              Robertfurt              Robertside          Robertsonbu
rgh
##                       2                       1
  1
##             Robertstown             Roberttown             Robinsont
own
##                       1                       1
  1
##               Rochabury             Rogerburgh              Ronaldp
ort
##                       1                       1
  1
##             Ronniemouth            Russellville              Ryanha
ven
##                       1                       1
  1
##             Salazarbury            Samanthaland             Sandrash
ire
##                       1                       1
  1
##               Sarahland             Shelbyport               Silva
ton
##                       1                       1
  1
##               Smithside            South Aaron             South A
dam
##                       1                       1
  1
##     South Alexisborough         South Cathyfurt      South Cynthiash
ire
##                       1                       1
  1
##            South Daniel        South Davidhaven        South Davidmo
uth
##                       1                       1
  1
##            South George             South Henry        South Jackieb
erg
##                       1                       1
  1
##              South Jade       South Jasminebury        South Jeannep
```

```
ort
##                         1                         1
  1
##            South Jessica              South John       South Johnnymo
uth
##                         1                         1
  1
##               South Kyle          South Lauraton          South Laurat
own
##                         1                         1
  1
##               South Lisa          South Margaret              South M
ark
##                         2                         1
  1
##             South Meghan             South Peter            South Rebe
cca
##                         1                         1
  1
##         South Tiffanyton    South Vincentchester             South Wal
ter
##                         1                         1
  1
##               Staceyfort             Suzannetown              Tammymo
uth
##                         1                         1
  1
##               Tammyshire             Taylorhaven             Taylormo
uth
##                         1                         1
  1
##              Timothyfurt            Timothymouth             Timothyp
ort
##                         1                         1
  1
##                  Tinaton              Tracyhaven           Turnerches
ter
##                         1                         1
  1
##                Tylerport             Vanessastad             Vanessav
iew
##                         1                         1
  1
##               Waltertown              Welchshire              Wendyvi
lle
##                         1                         1
  1
##              West Amanda           West Annefort          West Aprilp
ort
##                         2                         1
```

```
  1
##               West Brad         West Brandonton              West Ca
sey
##                       1                       1
  1
##       West Chloeborough        West Christopher             West Con
nor
##                       1                       1
  1
##              West David         West Derekmouth          West Dylanb
erg
##                       1                       1
  1
##        West Eduardotown          West Ericaport           West Ericf
urt
##                       1                       1
  1
##     West Gabriellamouth             West James              West J
ane
##                       1                       1
  1
##         West Jeremyside       West Jessicahaven             West Ju
lia
##                       1                       1
  1
##             West Justin         West Katiefurt          West Kevinf
urt
##                       1                       1
  1
##            West Leahton        West Lindseybury        West Melanief
urt
##                       1                       1
  1
##       West Melissashire            West Pamela             West Ra
ndy
##                       1                       1
  1
##       West Raymondmouth           West Ricardo            West Rich
ard
##                       1                       1
  1
##               West Ryan           West Shannon             West Ta
nya
##                       1                       2
  1
##          West Tinashire        West Travismouth            West Will
iam
##                       1                       1
  1
##               Westshire              Whiteport             Wilcoxp
```

```
ort
##                         1                         1
  1
##            Williammouth         Williamsborough            Williamsf
ort
##                         1                         1
  1
##           Williamsmouth            Williamsport             Williams
tad
##                         1                         2
  1
##              Wrightview               Yangside               Youngf
ort
##                         1                         1
  1
##                   Yuton             Zacharyton
##                         1                         1
```

Lisamouth, Michelleside, Millerbury, Robertfurt, South Lisa, West Shannon and Williamsport cities had the most people that viewed the ad

Pearson correlation and coefficient tests

```
install.packages("ggpubr", repos = "http://cran.us.r-project.org")
```

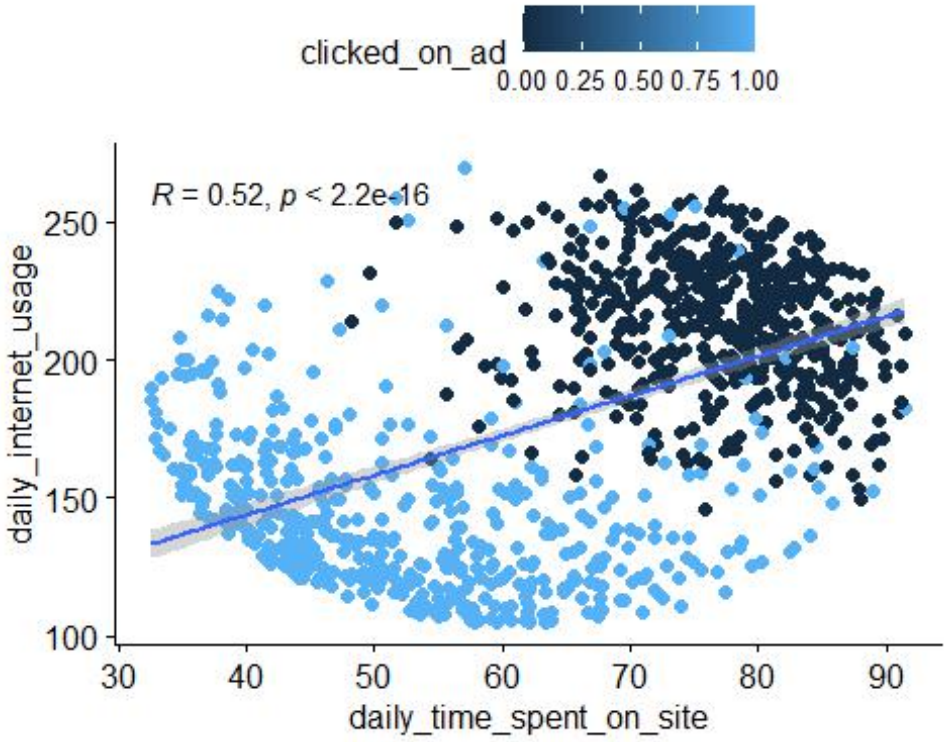
```
## Installing package into 'C:/Users/Lenovo/AppData/Local/R/win-library
/4.2'
## (as 'lib' is unspecified)
```

```
## package 'ggpubr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Lenovo\AppData\Local\Temp\RtmpID6235\downloaded_packages
```

```
library("ggpubr")
```

```
ggscatter(advert1, x = "daily_time_spent_on_site", y = "daily_internet_
usage",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "daily_time_spent_on_site", ylab = "daily_internet_usa
ge",  color = "clicked_on_ad")
```

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




```
res <- cor.test(advert1$daily_internet_usage, advert1$daily_time_spent_
on_site,
                    method = "pearson")
res # Testing the significance of the Pearson correlation coefficients
of daily internet usage and time_spent_on_site

##
##  Pearson's product-moment correlation
##
## data:  advert1$daily_internet_usage and advert1$daily_time_spent_on_
site
## t = 19.141, df = 990, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4727739 0.5637294
## sample estimates:
##       cor
## 0.5197228
```

The p-value of the test is 2.2e-16, which is less than the significance level alpha = 0.05. We can conclude that daily_internet_usage and daily_time_spent_on_site are significantly correlated with a correlation coefficient of 0.5197228 and p-value of 2.2e-16.

Individuals most likely to click on her ads

- are around the ages of 30-40

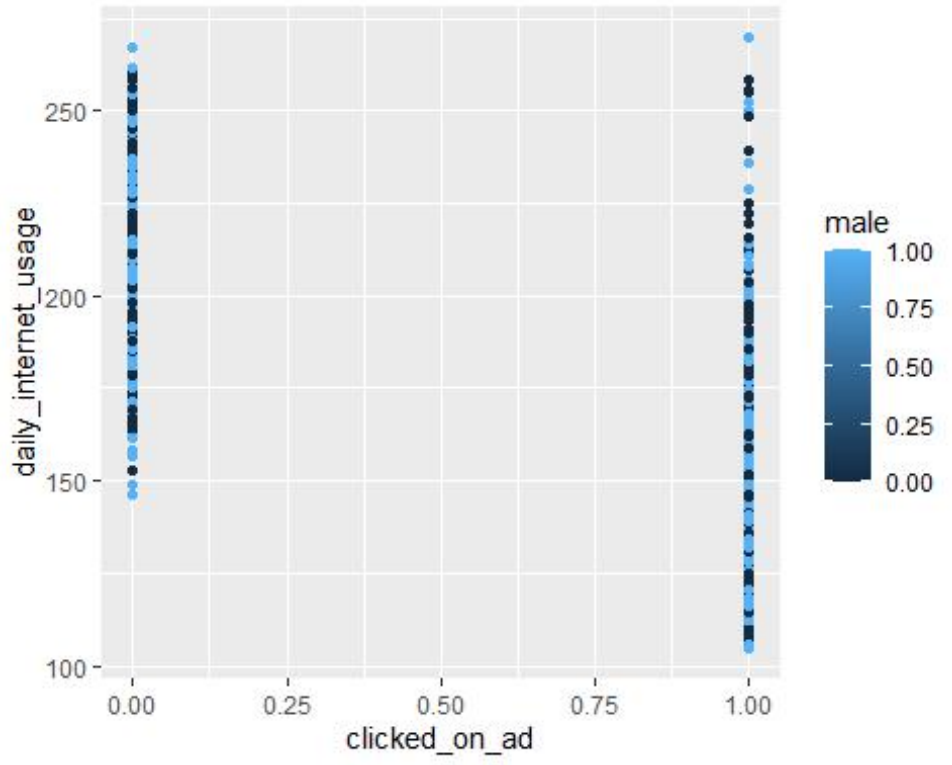
- daily time on site is 35-50,
- has a daily internet usage of 100-150
- has an area income of 40000-55000
- has an average daily internet time spent of 53
- has an average daily internet usage of 144.9

There is a positive relationship between daily_internet_usage and daily_spent_time

Most people who spend the lowest to moderate time on site and have the highest probability of clicking an ad

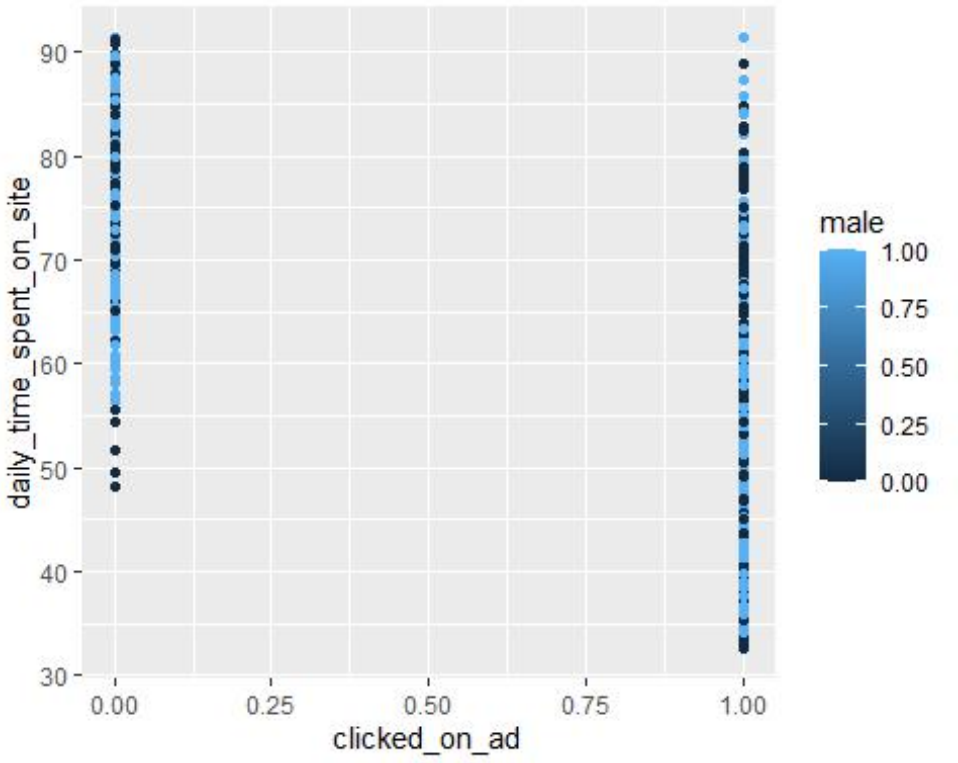
Challenging the Solution

```
#Checking how people clicked the ad depending on their daily internet u
sage while showing their gender
ggplot(data = advert1) +
  geom_point(mapping = aes(x = clicked_on_ad, y = daily_internet_usage,
 color = male))
```




```
#Checking how people clicked the ad depending on their daily_time_spent
_on_site while showing their gender
ggplot(data = advert1) +
```

```
  geom_point(mapping = aes(x = clicked_on_ad, y = daily_time_spent_on_s
ite, color = male))
```



Use classifier models such as Decision Tree Classifier to best predict those that clicked on the ad.

Follow up Questions

Did we have the right question?

yes, the research question was clear and specific.

Did we have the right data?

Yes, the data was relevant to the project. The dataset was also large enough for us to draw some answers.

Do we need other data to answer the research question?

Not necessarily.