HW3 Part2

```
if (!require(dplyr)) install.packages("dplyr")
```

```
## Loading required package: dplyr
```

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

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

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

```
library(dplyr)
if (!require(DataExplorer)) install.packages("DataExplorer")
```

```
## Loading required package: DataExplorer
```

```
## Warning: package 'DataExplorer' was built under R version 3.6.3
```

```
library(DataExplorer)
if (!require(ggplot2)) install.packages("ggplot2")
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
setwd("~/6203")
data <- read.csv("KAG.csv", stringsAsFactors = FALSE)
```

```
data <- data %>% mutate(CTR = round(((Clicks / Impressions) * 100),4),
                        CPC = ifelse(Clicks != 0, round(Spent / Clicks,4), Spent),
                        CostPerConv_Total = ifelse(Total_Conversion !=0,round(Spent/Total_Conver
sion,4),Spent),
                        CostPerConv_Approved = ifelse(Approved_Conversion!=0,round(Spent/Approve
d_Conversion,4),Spent),
                        CPM = round((Spent / Impressions) * 1000, 2) )
```

#Q.1 Which ad (provide ad_id as the answer) among the ads that have the least CPC led to the most impressions?

```
data %>% filter(CPC == min(CPC)) %>% filter(Impressions == max(Impressions)) %>% select(ad_id)
```

```
##     ad_id
## 1 1121094
```

It is 1121094.

#Q.2 What campaign (provide compaign_id as the answer) had spent least efficiently on brand awareness on an average(i.e. most Cost per mille or CPM: use total cost for the campaign / total impressions in thousands)?

```
data %>% group_by(campaign_id) %>% summarise(n_ads = length(ad_id),campaign_CPM = mean(CPM)) %>%
  arrange(desc(campaign_CPM))
```

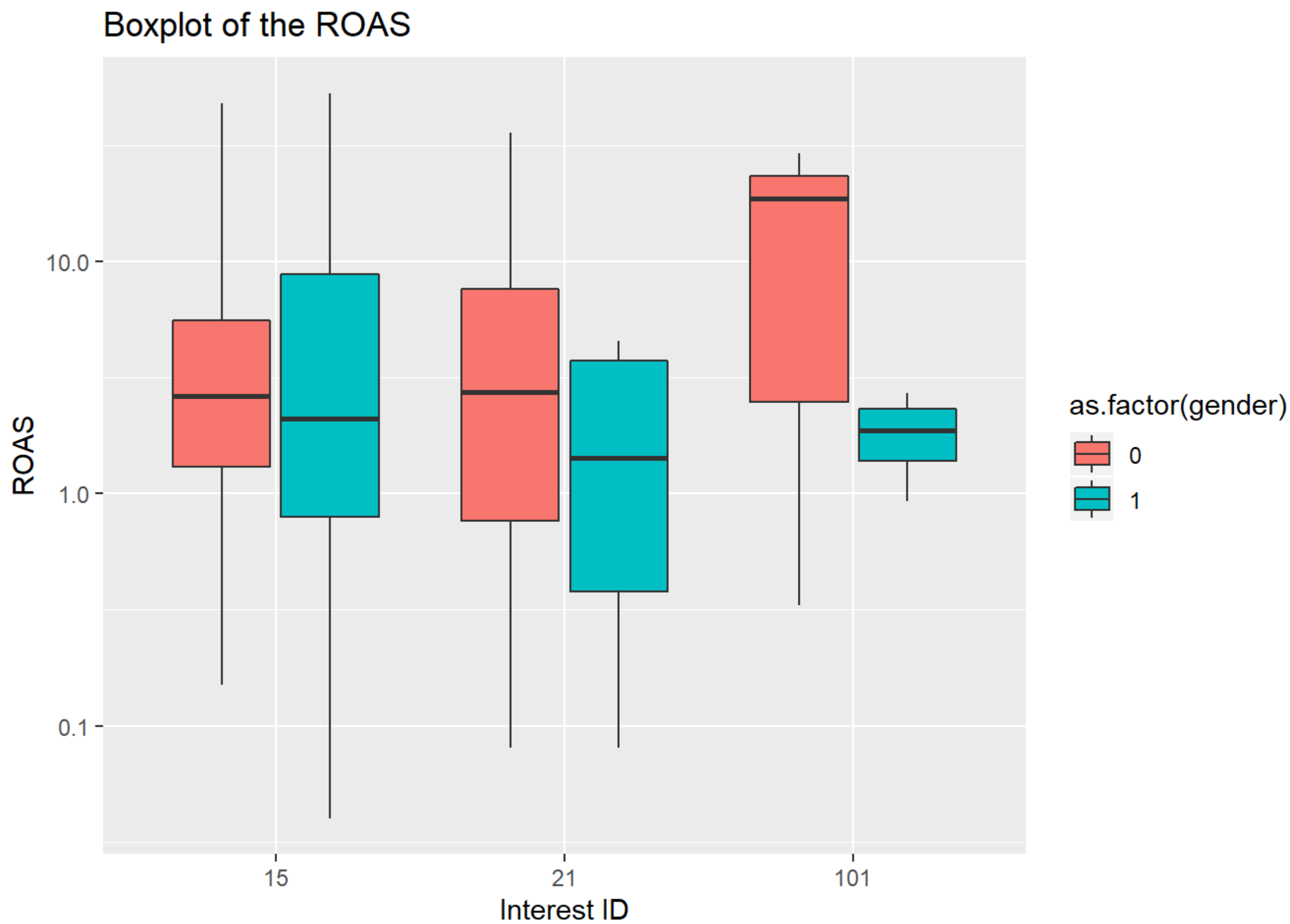
```
## # A tibble: 3 x 3
##   campaign_id n_ads campaign_CPM
##         <int> <int>        <dbl>
## 1        1178   625        0.250
## 2         916    54        0.241
## 3         936   464        0.224
```

1178 campaign_CPM 0.2503360

#Q.3 Assume each conversion (‘Total_Conversion’) is worth $5, each approved conversion (‘Approved_Conversion’) is worth $50. ROAS (return on advertising spent) is revenue as a percentage of the advertising spent . Calculate ROAS and round it to two decimals.

Make a boxplot of the ROAS grouped by gender for interest = 15, 21, 101 (or interest_id = 15, 21, 101) in one graph. Also try to use the function ‘+ scale_y_log10()’ in ggplot to make the visualization look better (to do so, you just need to add ‘+ scale_y_log10()’ after your ggplot function). The x-axis label should be ‘Interest ID’ while the y-axis label should be ROAS.

```
data <- data %>% mutate(ROAS = round(((5*Total_Conversion) +(50*Approved_Conversion))/Spent,2))

data2= filter(data,interest==15|interest==21|interest==101)
data2=filter(data2, Spent!= 0)

ggplot(data2, aes(x = as.factor(interest), y = ROAS, fill = as.factor(gender)))+ xlab("Interest
 ID") +scale_y_log10()+ geom_boxplot()+ ggtitle("Boxplot of the ROAS")
```



#Q.4 Summarize the median and mean of ROAS by genders when campign_id == 1178.

```
data %>%
  filter(campaign_id==1178)%>%
  filter(Spent!= 0)%>%
  group_by(gender) %>%
  summarise(median_ROAS = median(ROAS),mean_ROAS = mean(ROAS))
```

```
## # A tibble: 2 x 3
##   gender median_ROAS mean_ROAS
##    <int>       <dbl>     <dbl>
## 1      0        1.19      2.55
## 2      1        0.73      1.58
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 3.6.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.2
```

```
## -- Attaching packages ------------------------------ tidyverse 1.3.0 --
```

```
## v tibble  2.1.3     v stringr 1.4.0
## v tidyr   1.0.0     v forcats 0.4.0
## v purrr   0.3.3
```

```
## Warning: package 'tibble' was built under R version 3.6.2
```

```
## Warning: package 'forcats' was built under R version 3.6.2
```

```
## -- Conflicts ---------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
if (!require(correlationfunnel)) install.packages("correlationfunnel")
```

```
## Loading required package: correlationfunnel
```

```
## Warning: package 'correlationfunnel' was built under R version 3.6.3
```

```
## == Using correlationfunnel? ============================================
## You might also be interested in applied data science training for business.
## </> Learn more at - www.business-science.io </>
```

```
library(correlationfunnel)
library(DataExplorer)
if (!require(WVPlots)) install.packages("WVPlots")
```

```
## Loading required package: WVPlots
```

```
## Warning: package 'WVPlots' was built under R version 3.6.3
```

```
library(WVPlots)
if (!require(ggthemes)) install.packages("ggthemes")
```

```
## Loading required package: ggthemes
```

```
## Warning: package 'ggthemes' was built under R version 3.6.3
```

```
library(ggthemes)
library(ROCR)
```

```
## Warning: package 'ROCR' was built under R version 3.6.3
```

```
## Loading required package: gplots
```

```
## Warning: package 'gplots' was built under R version 3.6.3
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(caret)
```

```
## Loading required package: lattice
```

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

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

```
if (!require(corrplot)) install.packages("corrplot")
```

```
## Loading required package: corrplot
```

```
## Warning: package 'corrplot' was built under R version 3.6.3
```

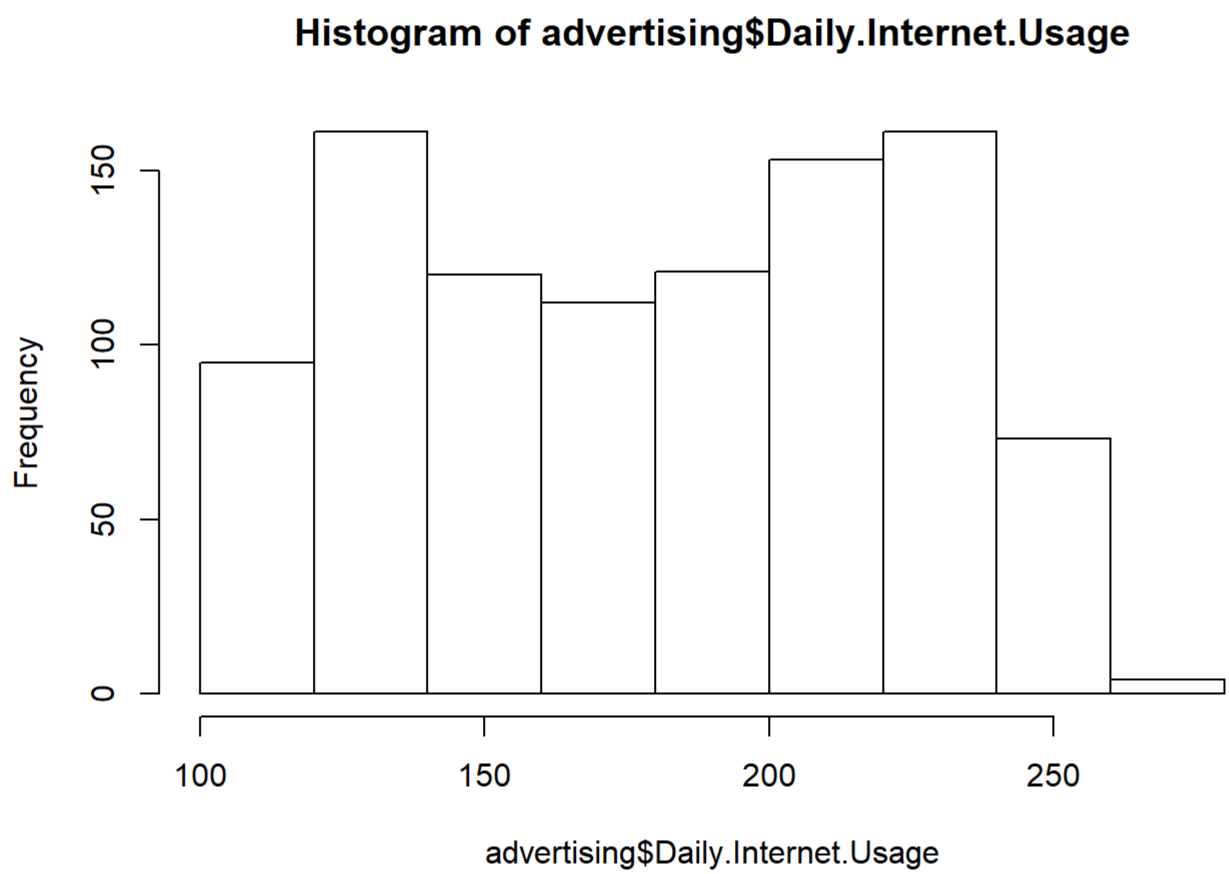
```
## corrplot 0.84 loaded
```

```
library(corrplot)

setwd("~/6203")
advertising <- read.csv("advertising1.csv", stringsAsFactors = FALSE)
as.factor(advertising$Clicked.on.Ad)
```

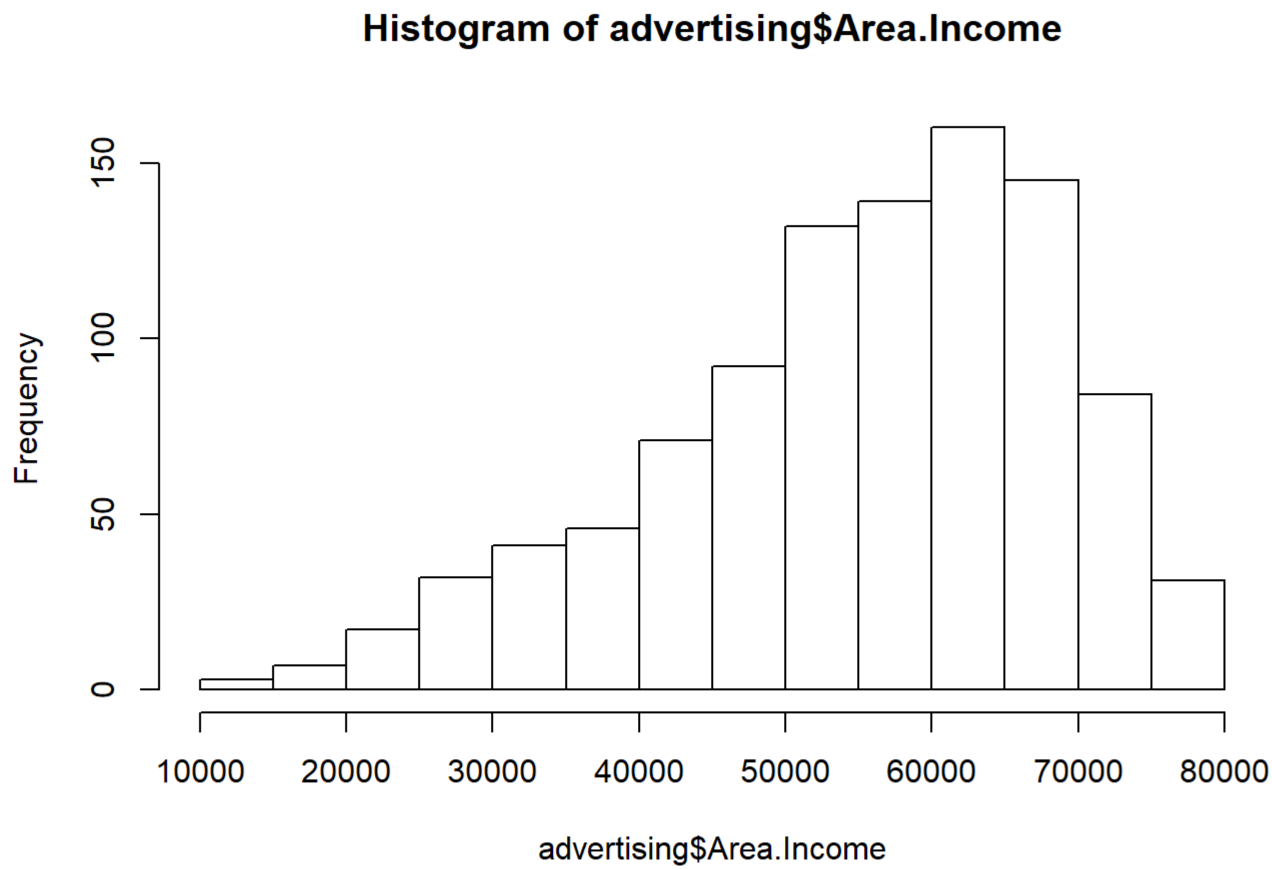
```
##    [1] 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 0 0 1 1
##   [35] 1 0 1 0 1 1 0 0 0 0 0 1 0 0 1 1 0 0 1 1 1 0 1 1 0 1 0 0 0 0 1 0 1 1
##   [69] 0 1 1 0 1 1 1 0 1 0 1 1 0 0 1 1 0 1 0 1 1 1 1 1 0 1 1 0 1 1 1 0 1 0
##  [103] 0 0 0 0 0 1 1 0 1 1 0 1 0 0 1 1 1 1 0 0 0 1 1 0 1 0 0 0 1 1 1 0 1 1
##  [137] 1 1 0 0 0 1 1 0 0 1 1 1 1 1 0 0 1 0 0 0 1 1 0 1 0 0 0 0 1 1 1 0 1 0
##  [171] 1 0 0 0 1 0 1 0 1 0 1 1 1 0 0 1 1 0 1 1 1 1 1 1 0 1 1 0 0 0 0 0 1 0
##  [205] 0 1 0 0 1 1 0 1 0 1 0 1 1 1 1 1 0 0 1 1 0 1 1 1 0 0 0 1 1 1 1 1 1 0
##  [239] 1 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 1 0 0 1 0 1 0 1 1 1 0 1 1 0 1 0 1 0
##  [273] 0 0 0 1 0 0 0 0 1 1 1 0 1 0 1 0 1 1 1 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1
##  [307] 0 0 0 1 0 0 1 0 0 1 0 0 0 1 1 0 0 0 0 1 1 0 0 1 0 0 1 0 0 1 0 0 0 0
##  [341] 1 1 0 0 1 0 0 1 0 1 0 0 0 0 1 0 1 1 1 0 1 1 0 1 0 1 0 0 0 0 1 1 0 1
##  [375] 0 0 0 1 1 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 1 0 1 1 0 0 1 0 1 0 1 0 1 1
##  [409] 1 1 1 0 0 1 0 1 1 0 0 0 1 0 1 1 1 1 1 0 1 0 0 0 1 0 0 1 0 0 1 0 1 0
##  [443] 1 1 1 0 1 0 1 0 1 1 0 0 1 0 1 0 1 0 1 1 0 1 0 1 1 1 1 0 1 0 0 0 1 0
##  [477] 0 1 1 1 0 0 0 1 1 1 0 0 1 0 1 1 0 1 1 0 0 1 0 1 1 0 0 1 1 0 0 1 1 0
##  [511] 1 0 0 1 0 1 0 1 1 1 1 1 0 1 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1
##  [545] 0 1 0 0 0 0 0 0 1 1 1 0 1 0 0 0 1 1 0 0 1 0 1 0 0 0 1 0 0 1 1 1 1 0
##  [579] 0 0 1 1 1 1 1 0 0 1 0 1 1 1 0 0 1 1 0 0 0 1 1 1 1 0 1 1 0 0 1 1 1 1
##  [613] 0 0 0 1 1 0 1 0 0 0 1 0 0 1 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 0 0 0 1
##  [647] 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 1 0 1 0 0 1 1 0 1 0 1 0 0 1 1 0 1
##  [681] 0 1 1 0 1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 1 1 0 0 0 1 0 1 1 1 0 0 1
##  [715] 0 1 1 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 1 0 0 1 1 1 1 1
##  [749] 1 0 1 0 0 0 0 0 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 0 1 0 1 1
##  [783] 0 0 1 1 0 1 0 1 1 1 0 1 1 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0
##  [817] 1 1 0 0 1 0 1 0 0 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 0 1 1 0 0 1
##  [851] 0 1 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 0 0 0 1 0 0 1
##  [885] 0 1 1 1 0 1 0 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 1 0 1 0 1 1 1 0 1 1 1 0
##  [919] 0 0 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1 1
##  [953] 1 1 0 1 1 0 0 0 1 0 0 0 0 1 1 0 1 1 1 1 1 0 1 1 1 1 0 0 1 0 1 0 0 1
##  [987] 0 1 0 0 1 1 1 0 1 1 1 1 0 1
## Levels: 0 1
```

#Q.5 a) We aim to explore the dataset so that we can better choose a model to implement. Plot histograms for at least 2 of the continuous variables in the dataset. Note it is acceptable to plot more than 2.

```
hist(advertising$Daily.Internet.Usage)
```

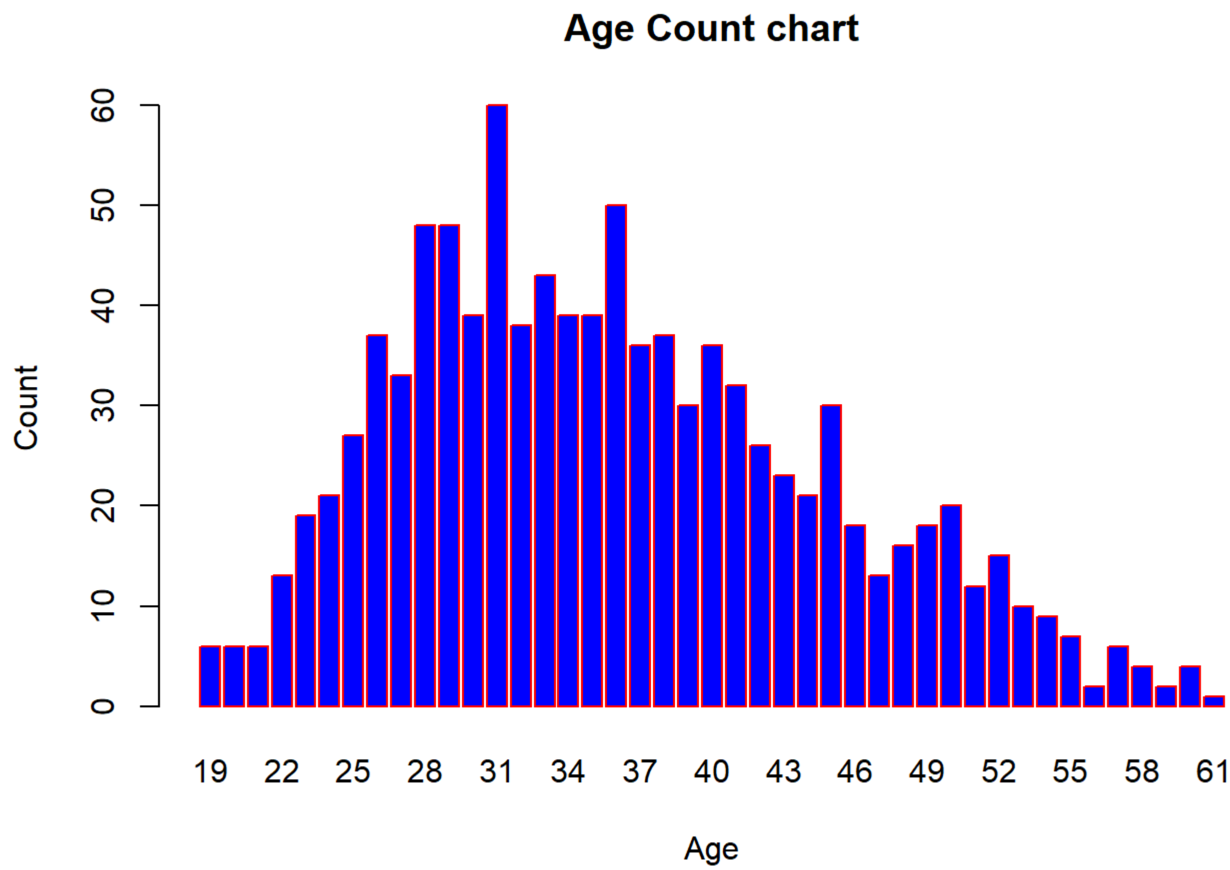



```
hist(advertising$Area.Income)
```

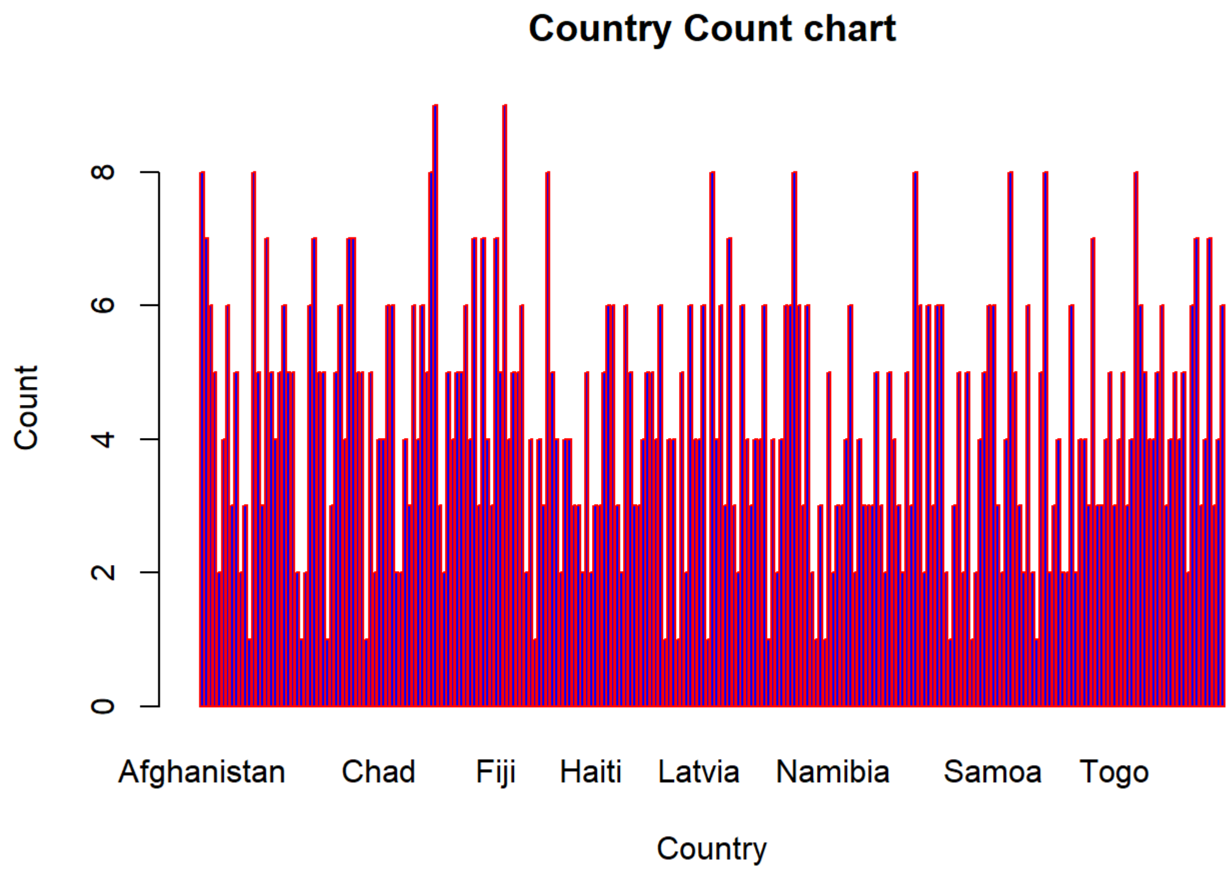


#Q.5b) Again on the track of exploring the dataset, plot at least 2 bar charts reflecting the counts of different values for different variables. Note it is acceptable to plot more than 2.

```
age <-advertising%>%
      group_by(Age) %>%
      summarise(n_age = length(Daily.Internet.Usage))


country<-advertising%>%
      group_by(Country) %>%
      summarise(n_country = length(Daily.Internet.Usage))


# Plot the bar chart
barplot(age$n_age,names.arg=age$Age,xlab="Age",ylab="Count",col="blue",
main="Age Count chart",border="red")
```

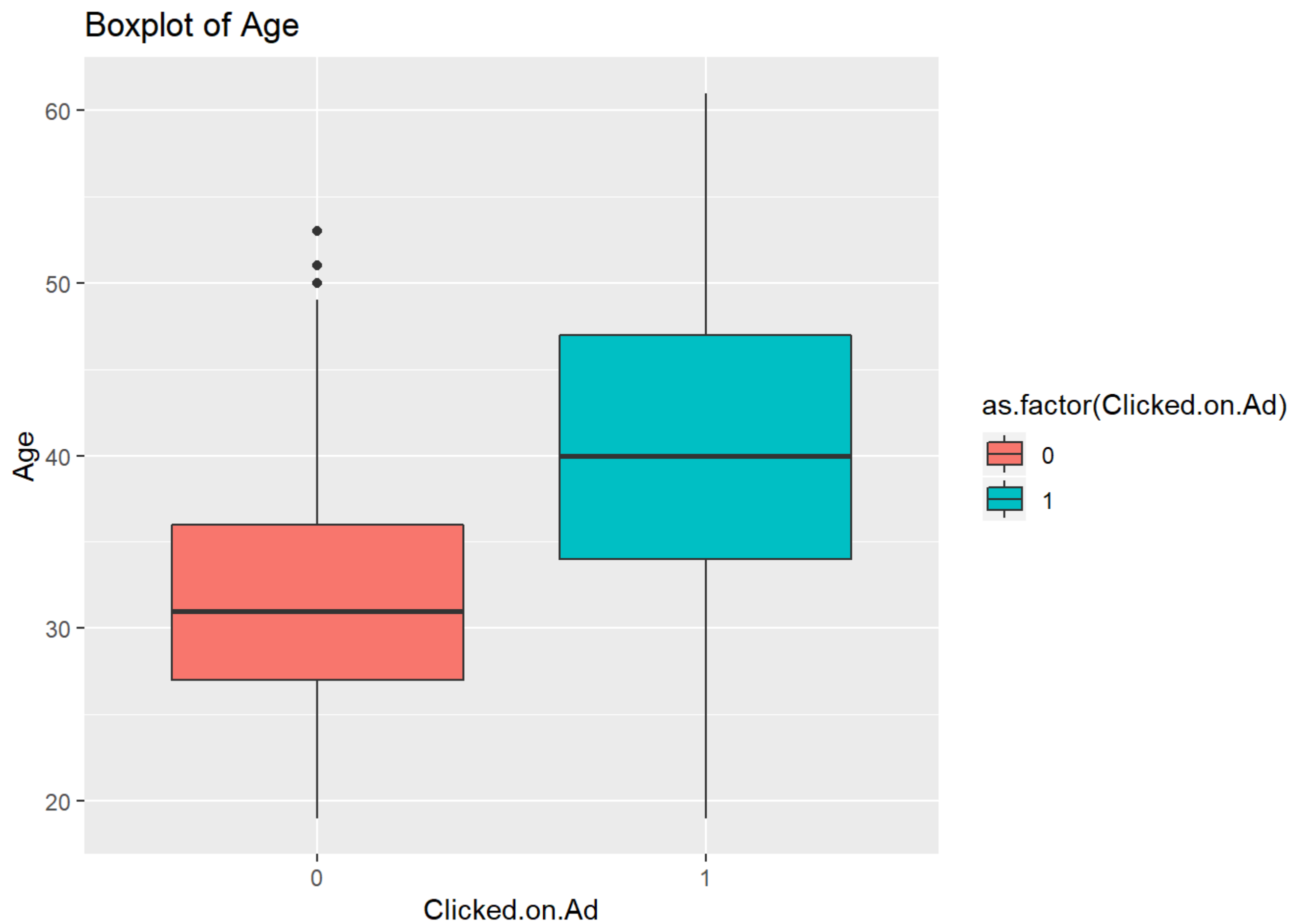



```
barplot(country$n_country,names.arg=country$Country,xlab="Country",ylab="Count",col="blue",
main="Country Count chart",border="red")
```

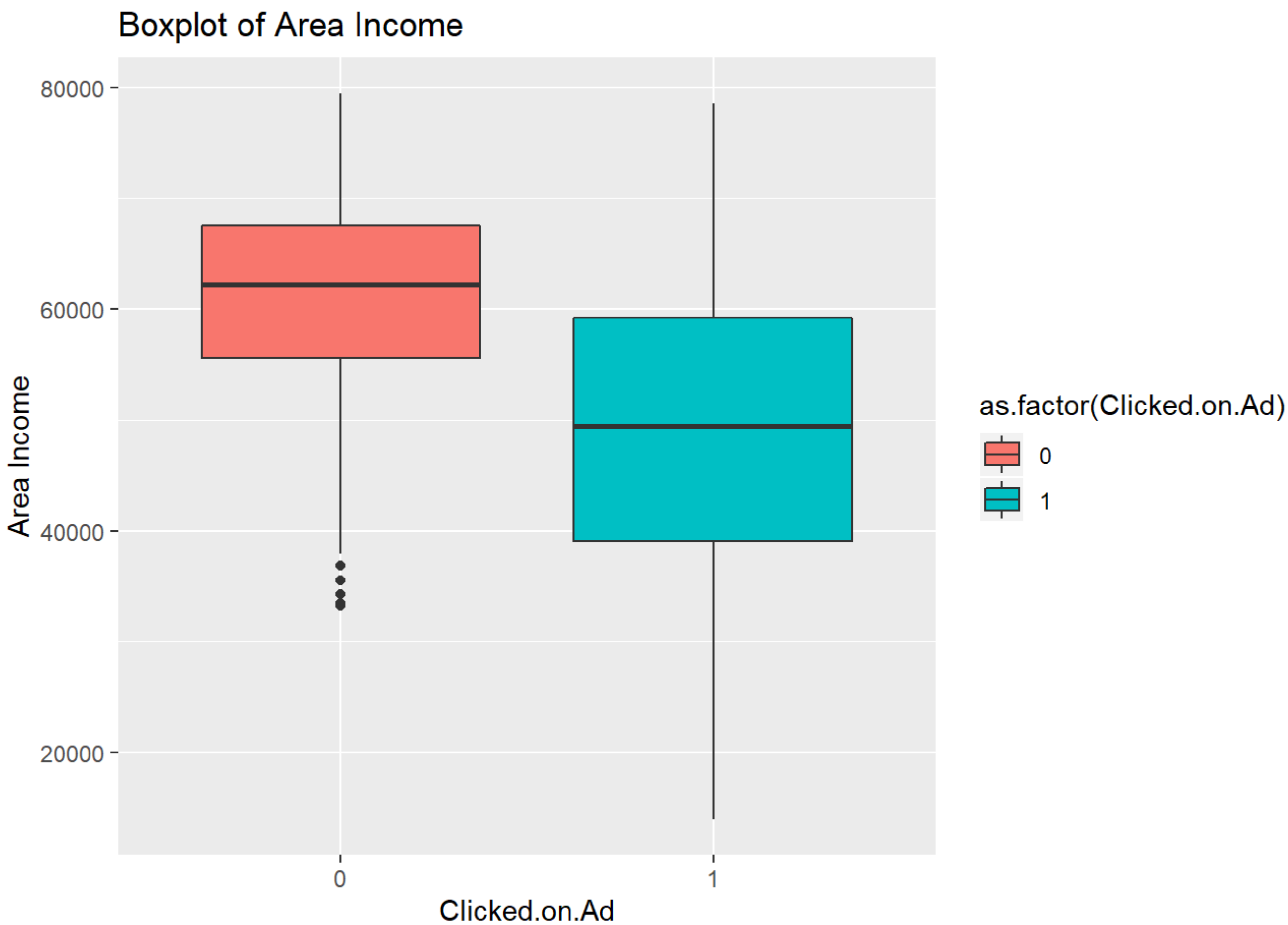


#Q.5 c) Plot boxplots for Age, Area.Income, Daily.Internet.Usage and Daily.Time.Spent.on.Site separated by the variable Clicked.on.Ad. To clarify, we want to create 4 plots, each of which has 2 boxplots: 1 for people who clicked on the ad, one for those who didn't

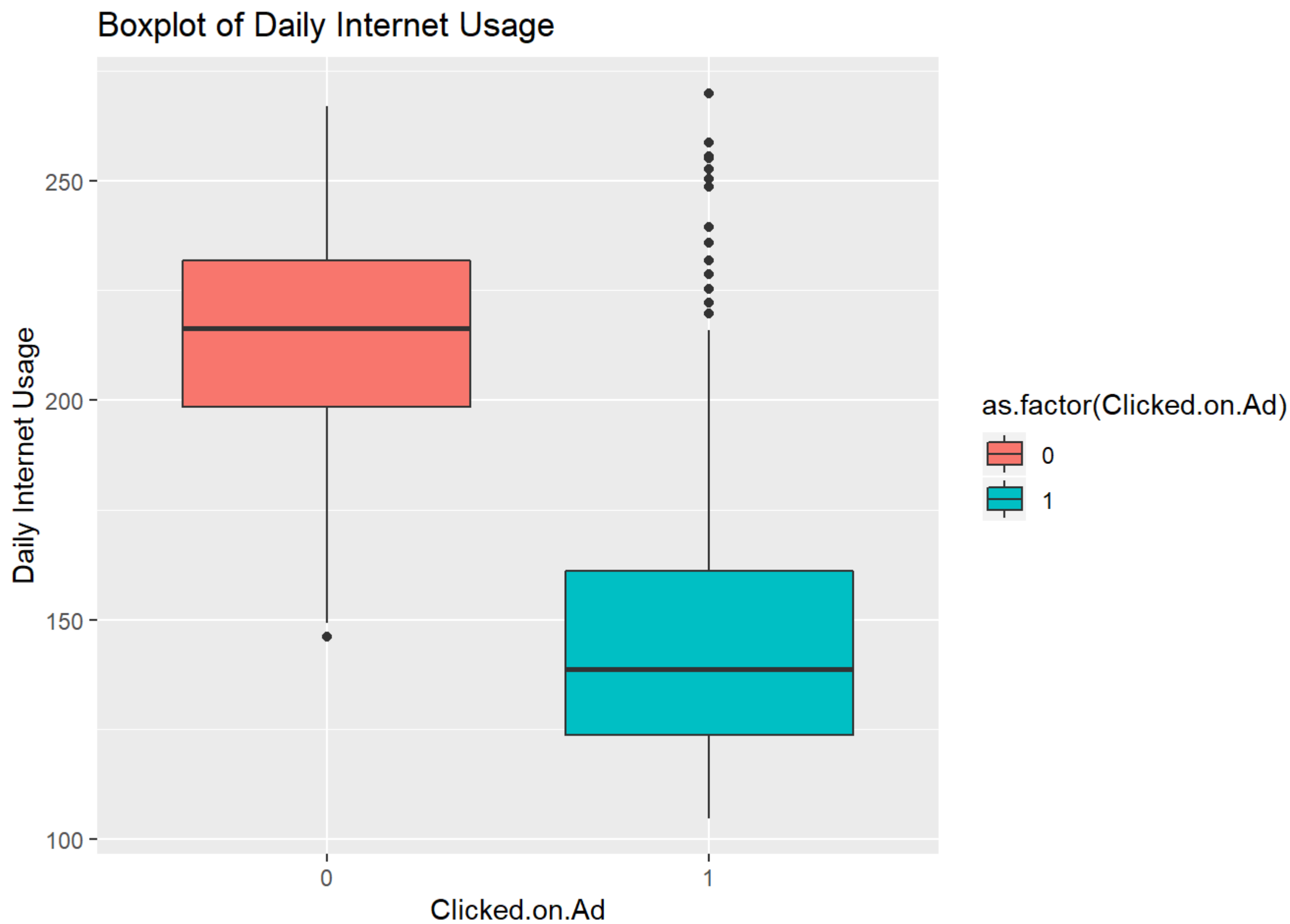
```
ggplot(advertising, aes(x = as.factor(advertising$Clicked.on.Ad),y = advertising$Age, fill = as.
factor(Clicked.on.Ad)))+  geom_boxplot()+ ggtitle("Boxplot of Age")+ ylab("Age")+xlab("Clicked.o
n.Ad")
```



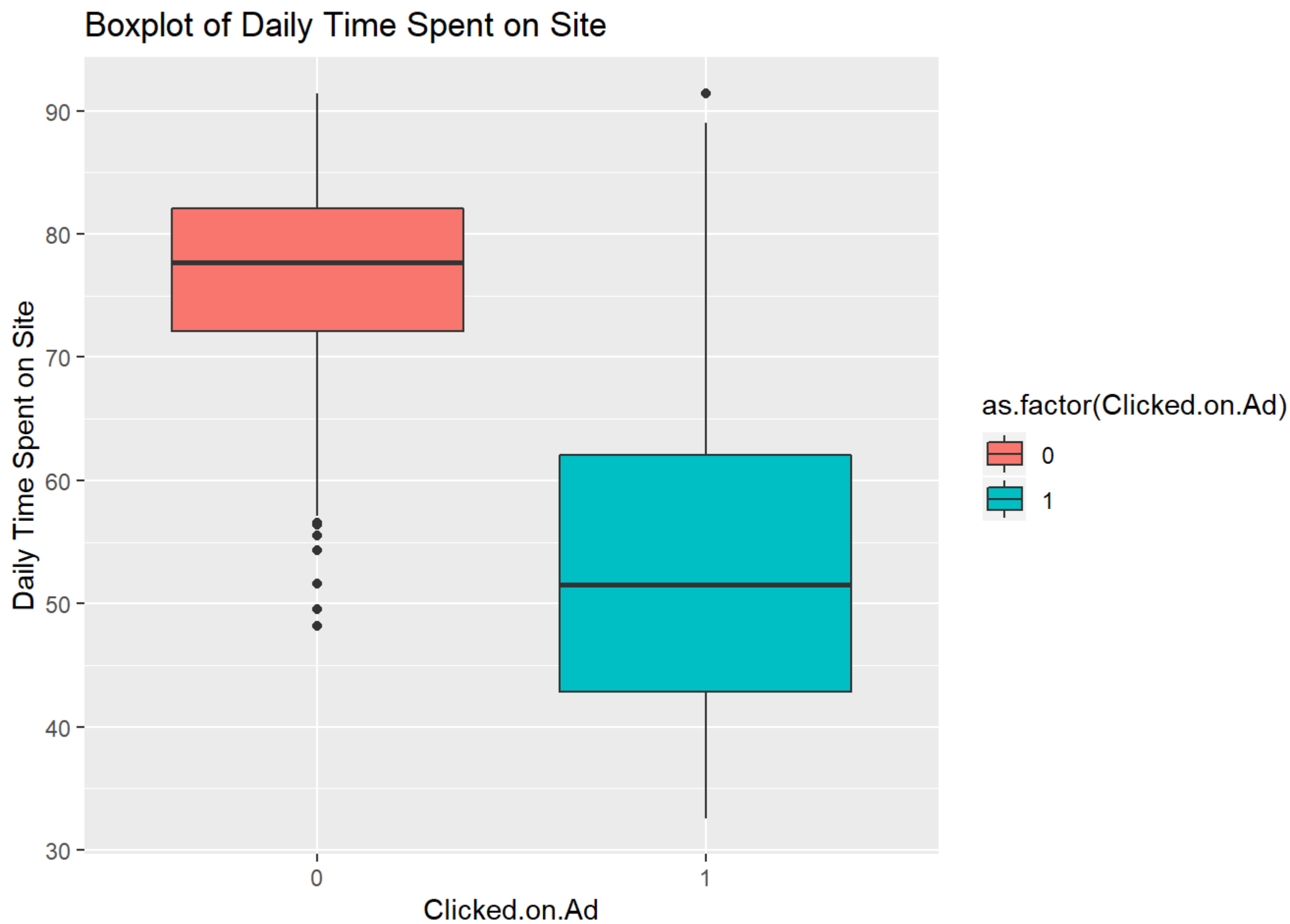

```
ggplot(advertising, aes(x = as.factor(advertising$Clicked.on.Ad),y = advertising$Area.Income, fi
ll = as.factor(Clicked.on.Ad)))+  geom_boxplot()+ ggtitle("Boxplot of Area Income")+ ylab("Area
 Income")+xlab("Clicked.on.Ad")
```




```
ggplot(advertising, aes(x = as.factor(advertising$Clicked.on.Ad),y = advertising$Daily.Internet.
Usage, fill = as.factor(Clicked.on.Ad)))+  geom_boxplot()+ ggtitle("Boxplot of Daily Internet Us
age")+ ylab("Daily Internet Usage")+xlab("Clicked.on.Ad")
```




```
ggplot(advertising, aes(x = as.factor(advertising$Clicked.on.Ad),y = advertising$Daily.Time.Spen
t.on.Site, fill = as.factor(Clicked.on.Ad)))+  geom_boxplot()+ ggtitle("Boxplot of Daily Time Sp
ent on Site")+ ylab("Daily Time Spent on Site")+xlab("Clicked.on.Ad")
```



#Q.5 d) Based on our preliminary boxplots, would you expect an older person to be more likely to click on the ad than someone younger?

Yes, the median to click on Ad is older compare to the median of not clicking the ad.

#Q.6 (a)1. Make a scatter plot for Area.Income against Age. Separate the datapoints by different shapes based on if the datapoint has clicked on the ad or not.

```
if (!require(car)) install.packages("car")
```

```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 3.6.2
```

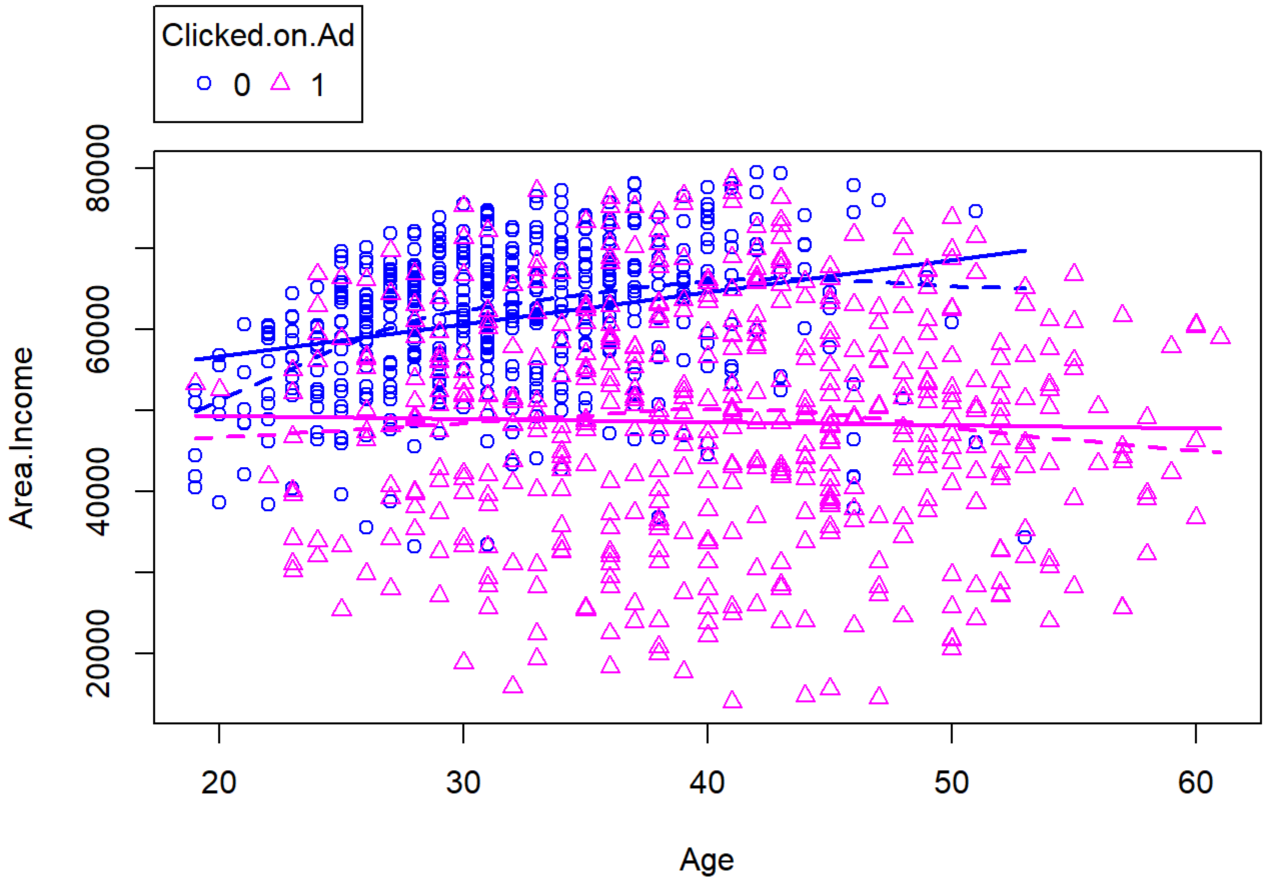
```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

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

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

```
library(car)
scatterplot(Area.Income ~ Age | Clicked.on.Ad, data = advertising,
            grid = FALSE, frame = TRUE)
```

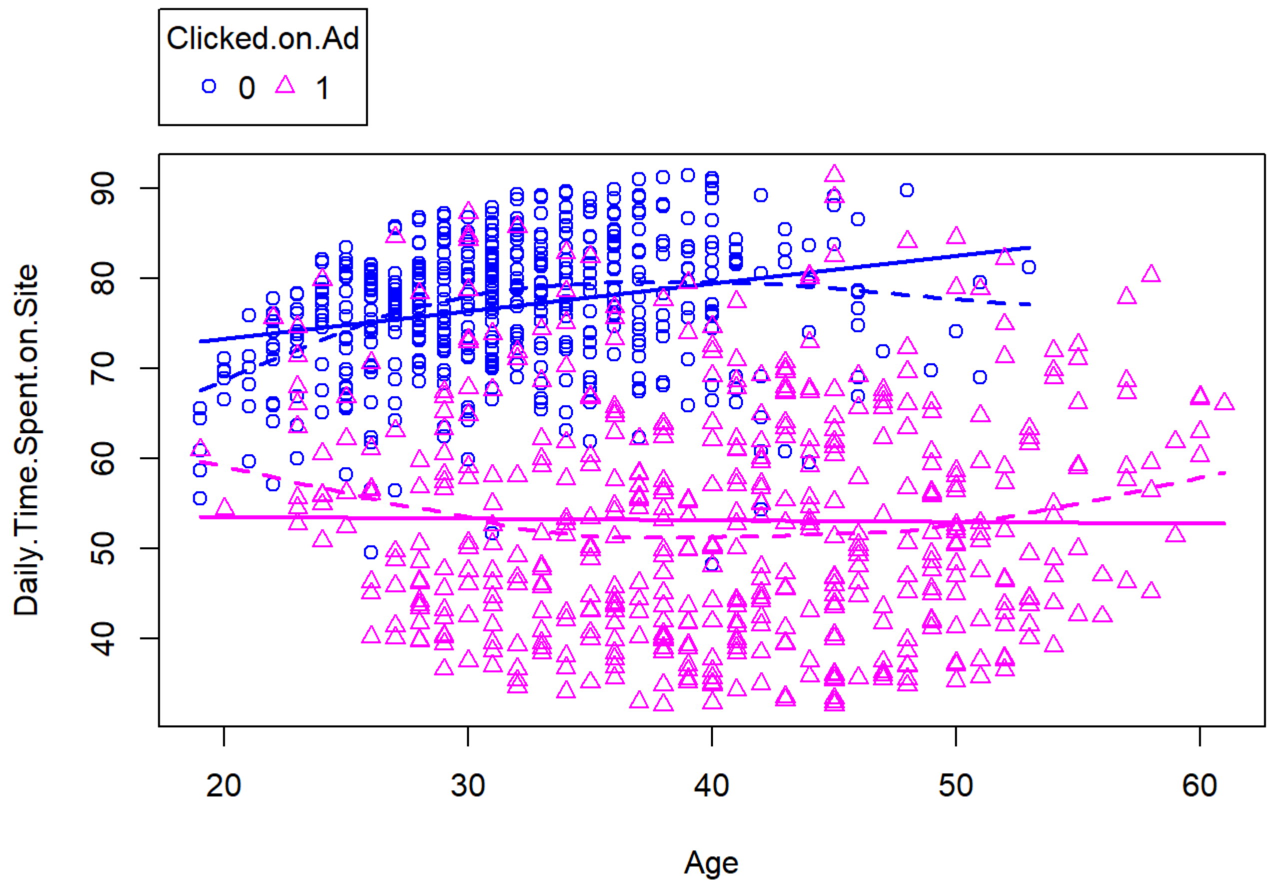


#Q.6 (a)2. Based on this plot, would you expect a 31-year-old person with an Area income of $62,000 to click on the ad or not?

Based on this plot, i expect a 31-year-old person with an Area income of $62,000 not clicking the ad.

#Q.6 (b)1. Similar to part a), create a scatter plot for Daily.Time.Spent.on.Site against Age. Separate the datapoints by different shapes based on if the datapoint has clicked on the ad or not.

```
scatterplot(Daily.Time.Spent.on.Site ~ Age | Clicked.on.Ad, data = advertising,
            grid = FALSE, frame = TRUE)
```



#Q.6 (b)2. Based on this plot, would you expect a 50-year-old person who spends 60 minutes daily on the site to click on the ad or not?

Based on this plot, I expect a 50-year-old person who spends 60 minutes daily on the site to click on the ad.

#Q.7 (a) 1. Now that we have done some exploratory data analysis to get a better understanding of our raw data, we can begin to move towards designing a model to predict advert clicks.

2. Generate a correlation funnel (using the correlation funnel package) to see which of the variable in the dataset have the most correlation with having clicked the advert.

NOTE: Here we are creating the correlation funnel in regards to HAVING clicked the advert, rather than not. This will lead to a minor distinction in your code between the 2 cases. However, it will not affect your results and subsequent variable selection.

```
advertising$Age <-as.double(advertising$Age)
advertising$Male<-as.double(advertising$Male)
advertising$Clicked.on.Ad<-as.double(advertising$Clicked.on.Ad)
advertising %>%  glimpse()
```

```
## Observations: 1,000
## Variables: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59...
## $ Age                      <dbl> 35, 31, 26, 29, 35, 23, 33, 48, 30, 2...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.1...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.5...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration",...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton...
## $ Male                     <dbl> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "It...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01...
## $ Clicked.on.Ad            <dbl> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1...
```

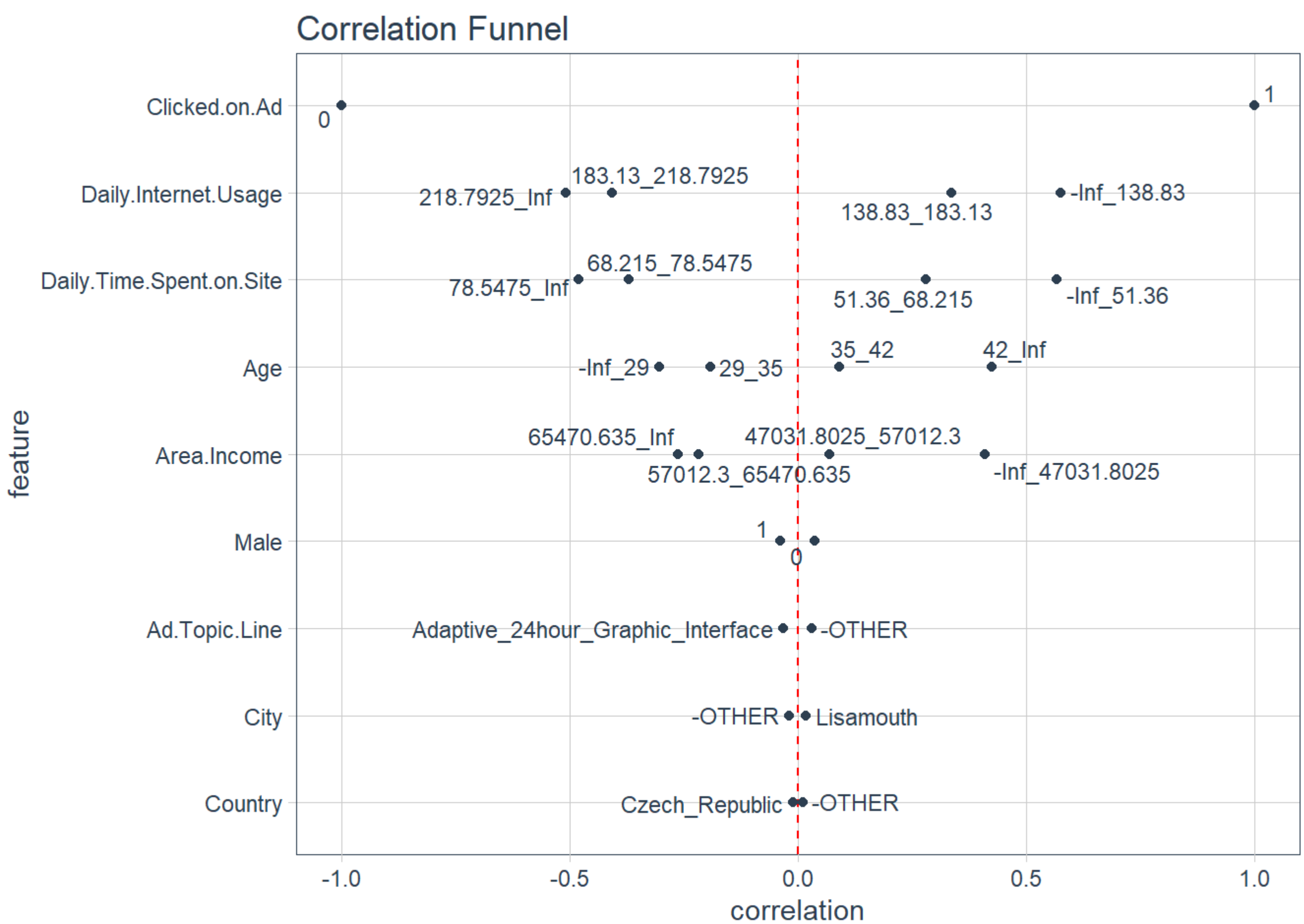
```
advertising_binarized_tbl <- advertising  %>%
    select(- Timestamp) %>%
    binarize(n_bins = 4, thresh_infreq = 0.01)

advertising_binarized_tbl %>% glimpse()
```

```
## Observations: 1,000
## Variables: 26
## $ `Daily.Time.Spent.on.Site__-Inf_51.36`          <dbl> 0, 0, 0, 0, 0...
## $ Daily.Time.Spent.on.Site__51.36_68.215          <dbl> 0, 0, 0, 0, 0...
## $ Daily.Time.Spent.on.Site__68.215_78.5475        <dbl> 1, 0, 1, 1, 1...
## $ Daily.Time.Spent.on.Site__78.5475_Inf           <dbl> 0, 1, 0, 0, 0...
## $ `Age__-Inf_29`                                  <dbl> 0, 0, 1, 1, 0...
## $ Age__29_35                                      <dbl> 1, 1, 0, 0, 1...
## $ Age__35_42                                      <dbl> 0, 0, 0, 0, 0...
## $ Age__42_Inf                                     <dbl> 0, 0, 0, 0, 0...
## $ `Area.Income__-Inf_47031.8025`                  <dbl> 0, 0, 0, 0, 0...
## $ Area.Income__47031.8025_57012.3                 <dbl> 0, 0, 0, 1, 0...
## $ Area.Income__57012.3_65470.635                  <dbl> 1, 0, 1, 0, 0...
## $ Area.Income__65470.635_Inf                      <dbl> 0, 1, 0, 0, 1...
## $ `Daily.Internet.Usage__-Inf_138.83`             <dbl> 0, 0, 0, 0, 0...
## $ Daily.Internet.Usage__138.83_183.13             <dbl> 0, 0, 0, 0, 0...
## $ Daily.Internet.Usage__183.13_218.7925           <dbl> 0, 1, 0, 0, 0...
## $ Daily.Internet.Usage__218.7925_Inf              <dbl> 1, 0, 1, 1, 1...
## $ Ad.Topic.Line__Adaptive_24hour_Graphic_Interface <dbl> 0, 0, 0, 0, 0...
## $ `Ad.Topic.Line__-OTHER`                         <dbl> 1, 1, 1, 1, 1...
## $ City__Lisamouth                                 <dbl> 0, 0, 0, 0, 0...
## $ `City__-OTHER`                                  <dbl> 1, 1, 1, 1, 1...
## $ Male__0                                         <dbl> 1, 0, 1, 0, 1...
## $ Male__1                                         <dbl> 0, 1, 0, 1, 0...
## $ Country__Czech_Republic                         <dbl> 0, 0, 0, 0, 0...
## $ `Country__-OTHER`                               <dbl> 1, 1, 1, 1, 1...
## $ Clicked.on.Ad__0                                <dbl> 1, 1, 1, 1, 1...
## $ Clicked.on.Ad__1                                <dbl> 0, 0, 0, 0, 0...
```

```
advertising_binarized_correlated_tbl <- advertising_binarized_tbl %>%
    correlate(target = Clicked.on.Ad__1)
advertising_binarized_correlated_tbl
```

```
## # A tibble: 26 x 3
##    feature                 bin               correlation
##    <fct>                   <chr>                   <dbl>
##  1 Clicked.on.Ad           0                      -1
##  2 Clicked.on.Ad           1                       1
##  3 Daily.Internet.Usage    -Inf_138.83             0.577
##  4 Daily.Time.Spent.on.Site -Inf_51.36             0.568
##  5 Daily.Internet.Usage    218.7925_Inf           -0.508
##  6 Daily.Time.Spent.on.Site 78.5475_Inf           -0.480
##  7 Age                     42_Inf                  0.425
##  8 Area.Income             -Inf_47031.8025         0.411
##  9 Daily.Internet.Usage    183.13_218.7925        -0.406
## 10 Daily.Time.Spent.on.Site 68.215_78.5475        -0.370
## # ... with 16 more rows
```

```
advertising_binarized_correlated_tbl %>%
    plot_correlation_funnel(interactive = FALSE)
```



When the daily internet usage is 138.83 or longer, the correlation is the largest 0.57735

#Q.7 (b) 1. Based on the generated correlation funnel, choose the 4 most covarying variables (with having clicked the advert) and run a logistic regression model for Clicked.on.Ad using these 4 variables. 2. Output the summary of this model.

```
library("ISLR")
```

```
## Warning: package 'ISLR' was built under R version 3.6.2
```

```
Model_ad <-glm(Clicked.on.Ad~ Daily.Internet.Usage+Area.Income+Daily.Time.Spent.on.Site+Age, dat
a=advertising, family="binomial")
```

```
Model_ad
```

```
##
## Call:  glm(formula = Clicked.on.Ad ~ Daily.Internet.Usage + Area.Income +
##     Daily.Time.Spent.on.Site + Age, family = "binomial", data = advertising)
##
## Coefficients:
##              (Intercept)      Daily.Internet.Usage
##               27.1290649                -0.0639129
##              Area.Income  Daily.Time.Spent.on.Site
##               -0.0001354                -0.1919295
##                      Age
##                0.1709213
##
## Degrees of Freedom: 999 Total (i.e. Null);  995 Residual
## Null Deviance:       1386
## Residual Deviance: 182.9     AIC: 192.9
```

```
summary(Model_ad)
```

```
##
## Call:
## glm(formula = Clicked.on.Ad ~ Daily.Internet.Usage + Area.Income +
##     Daily.Time.Spent.on.Site + Age, family = "binomial", data = advertising)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4578  -0.1341  -0.0333   0.0167   3.1961
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               2.713e+01  2.714e+00   9.995  < 2e-16 ***
## Daily.Internet.Usage     -6.391e-02  6.745e-03  -9.475  < 2e-16 ***
## Area.Income              -1.354e-04  1.868e-05  -7.247 4.25e-13 ***
## Daily.Time.Spent.on.Site -1.919e-01  2.066e-02  -9.291  < 2e-16 ***
## Age                       1.709e-01  2.568e-02   6.655 2.83e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1386.3  on 999  degrees of freedom
## Residual deviance:  182.9  on 995  degrees of freedom
## AIC: 192.9
##
## Number of Fisher Scoring iterations: 8
```

```
library("ROCR")
# make predictions using Model_ad
df<-advertising
df <-  df %>%
  mutate(pred_prob_Model_ad = predict(Model_ad, newdata = ., type = "response")) %>%
  mutate(pred_outcome_mModel_ad = ifelse(pred_prob_Model_ad >= 0.8,1,0))
# we are using 0.8 as cutoff for predicting Y=1.
View(df)
## two-way Cross Tab table of Actual outcome and predicted Outcome
xtabs(~Clicked.on.Ad + pred_outcome_mModel_ad, data = df)
```

```
##              pred_outcome_mModel_ad
## Clicked.on.Ad   0   1
##             0 497   3
##             1  36 464
```

36 false-negative occurrences is my observe.