

ADVERT CONSUMER ANALYSIS

Objective:

To identify which individuals are most likely to click on website adverts.

Success Metric:

Establish the specific groups of individuals more likely to view web adverts on cryptography given the data provided.

The context:

The data was collected from a blog that ran ads to advertise a course. The data contains daily time spent on site, age of the user, area income, daily internet usage of the user, ad Topic Line, the city, gender, country and the time stamp. The data source can be accessed from

<http://bit.ly/IPAdvertisingData>

Experimental Design

R has been used for this analysis. The exploratory data analysis has given graphical presentations of univariate and bivariate analysis.

Reading the data set

```
advert <- read.csv('http://bit.ly/IPAdvertisingData')
advert_df <- data.frame(advert)
rmarkdown::paged_table(head(advert_df, n=4))
#Checking the the last 4 rows of the data set
```

```

rmarkdown::paged_table(tail(advert_df,n=4))
#Checking the number of rows and columns in the data set
dim(advert_df)
## [1] 1000    10
#The data set has 1000 rows and 10 columns
#Checking the data types of the columns
sapply(advert_df, class)
## Daily.Time.Spent.on.Site          Age
Area.Income
##              "numeric"              "integer"
"numeric"
##      Daily.Internet.Usage          Ad.Topic.Line
City
##              "numeric"              "character"
"character"
##              Male              Country
Timestamp
##              "integer"              "character"
"character"
##              Clicked.on.Ad
##              "integer"
#Checking column names
colnames(advert_df)
## [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"

```

Data Cleaning

Checking for duplicates

```

sum(duplicated(advert_df))
## [1] 0
#The data has no duplicated rows

```

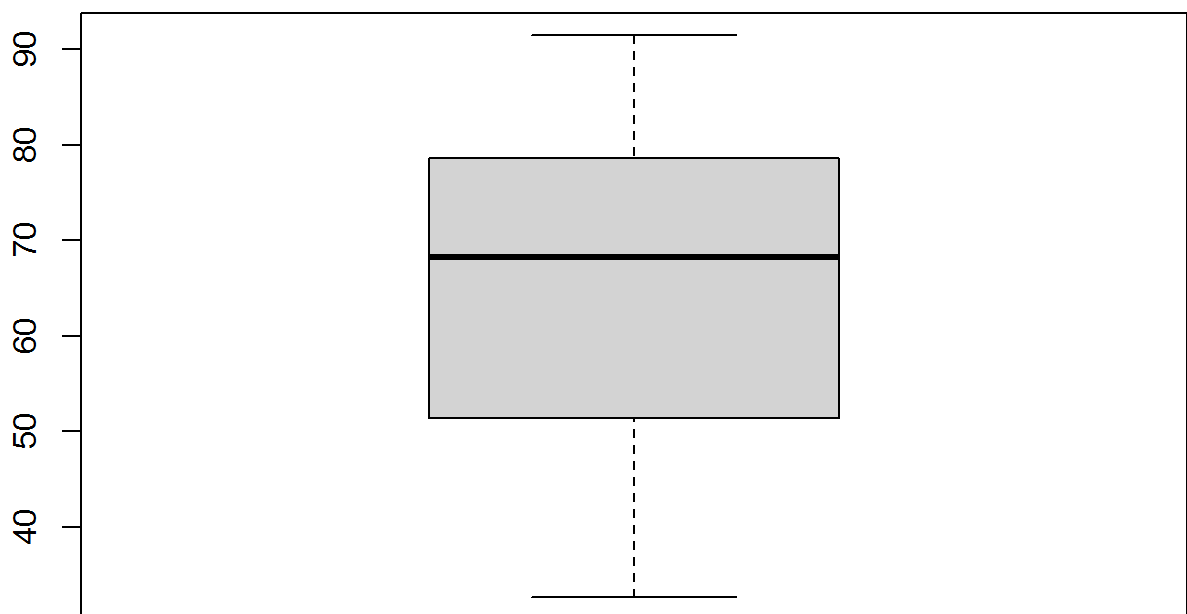
Checking for missing values

```

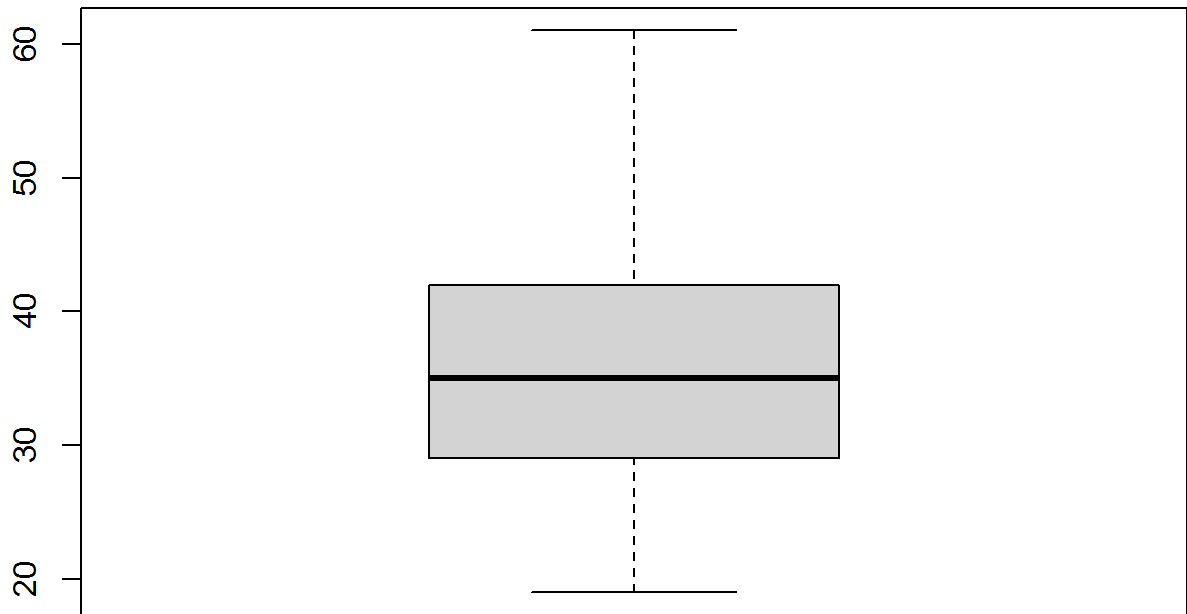
colSums(is.na(advert_df))
## Daily.Time.Spent.on.Site          Age
Area.Income
##              0              0
0
##      Daily.Internet.Usage          Ad.Topic.Line
City
##              0              0
0

```

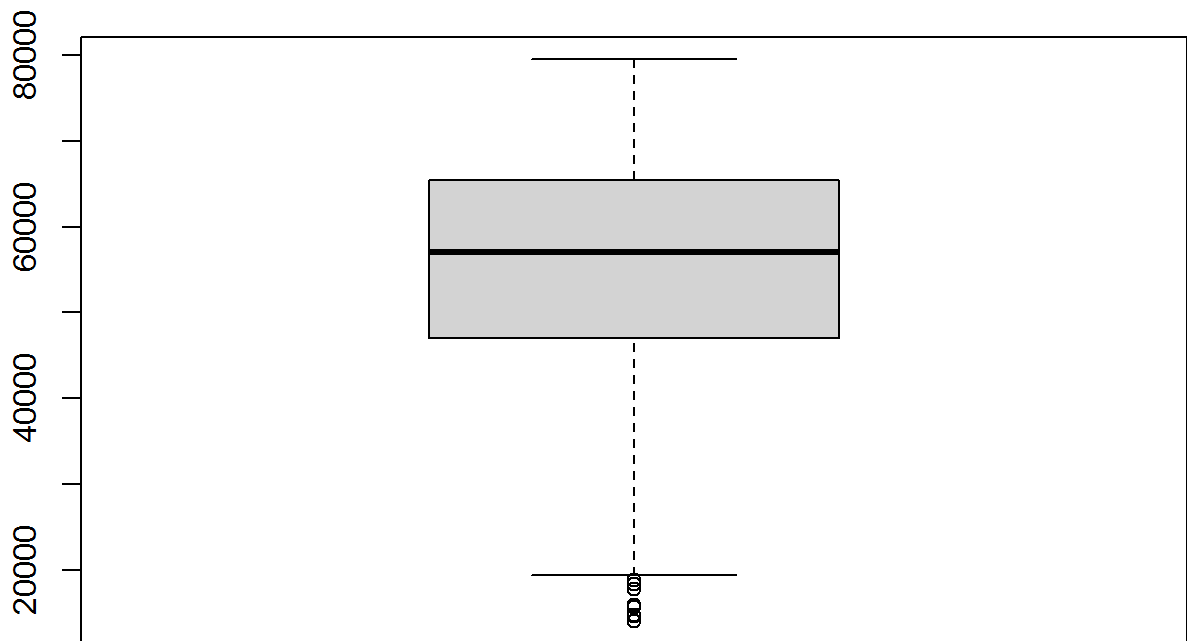
```
##                                Male                                Country
Timestamp
##                                0                                0
0
##          Clicked.on.Ad
##                                0
#There are no missing values in all columns
Checking for outliers
boxplot(advert_df$Daily.Time.Spent.on.Site)
```



```
#The are no outliers on the 'daily time spent on site' column
boxplot(advert_df$Age)
```



```
#There are no outliers on the 'age' column  
boxplot(advert_df$Area.Income)
```



#There are outliers on the 'area income' column. However, the outliers will not be dropped as they are legitimate data given the context of the column data.

Correcting column names

```
colnames(advert_df)
## [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"
#Renaming the 'male' column to 'gender' where 1 represents male and 0
female
colnames(advert_df)[7] <- 'Gender'
#Removing fullstops in the column names
colnames(advert_df)[1] <- 'Daily Time Spent on Site'
colnames(advert_df)[3] <- 'Area Income'
colnames(advert_df)[4] <- 'Daily Internet Usage'
colnames(advert_df)[5] <- 'Ad Topic Line'
colnames(advert_df)[10] <- 'Clicked on Ad'
#Checking corrected column names
colnames(advert_df)
```

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

Correcting Column values

#Changing 0 and 1 values in the gender and clicked on Ad columns in male and female values and yes and no values respectively. It will help in visualizing the outcomes.

```
advert_df$Gender <- ifelse(advert_df$Gender == 1, 'male', 'female')
```

```
advert_df$`Clicked on Ad` <- ifelse(advert_df$`Clicked on Ad`==1, 'yes', 'no')
```

#reading the data to check if the corrections made have reflected

```
rmarkdown::paged_table(head(advert_df, n=5))
```

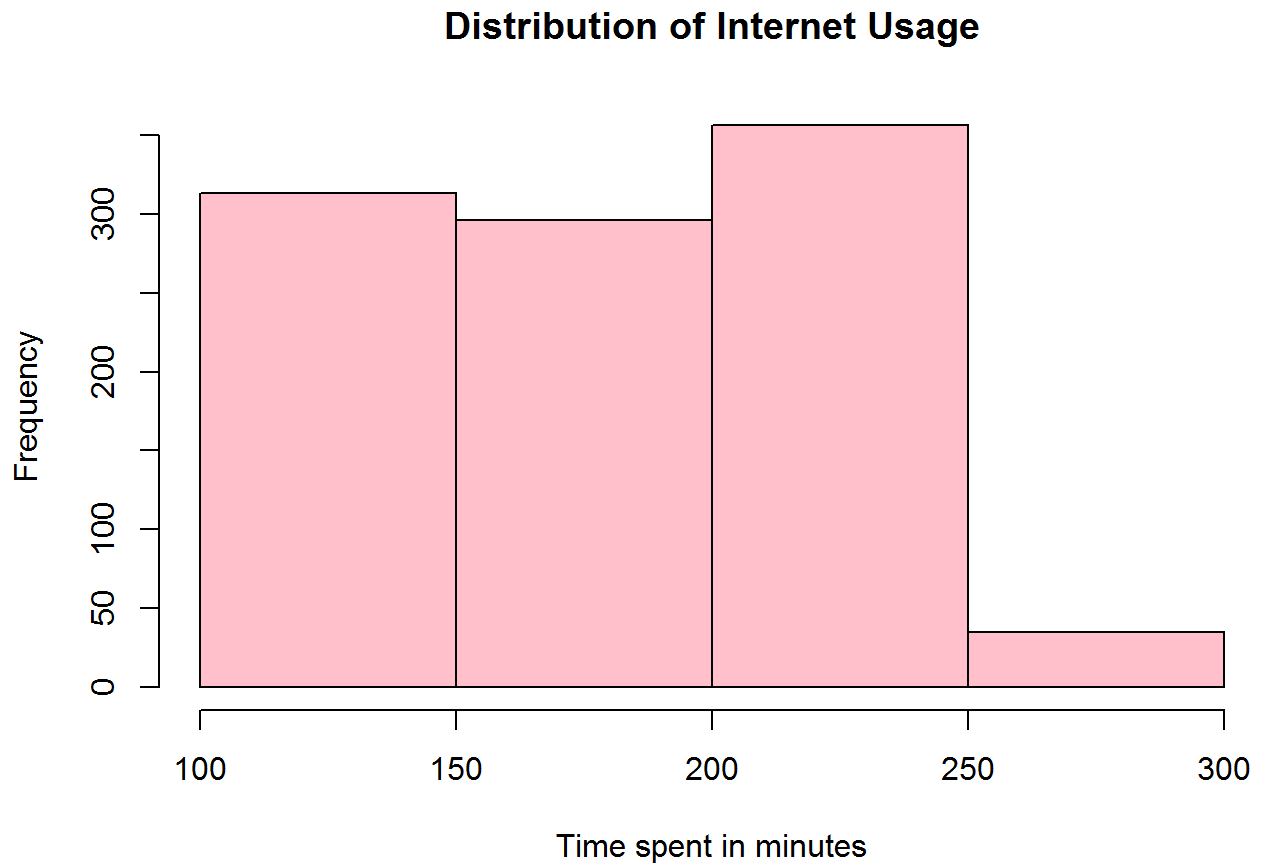
Exploratory Data Analysis

Univariate Analysis

Installing libraries

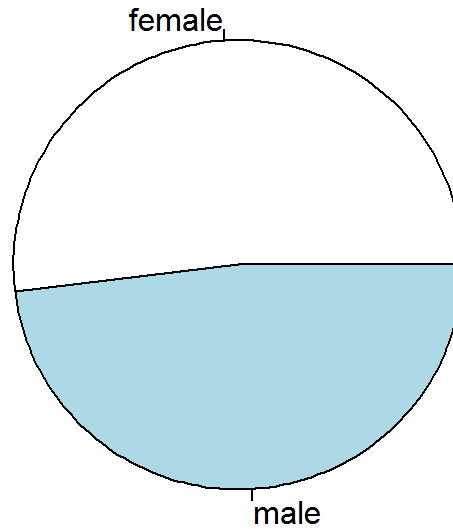
```
library(tidyverse)
## -- Attaching packages -----
tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.5       v dplyr 1.0.7
## v tidyr 1.1.4        v stringr 1.4.0
## v readr 2.0.2        v forcats 0.5.1
## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
mean(advert_df$`Daily Time Spent on Site`)
## [1] 65.0002
#On average, the respondents spent 65.0002 minutes on the site
mean(advert_df$Age)
## [1] 36.009
#On average, the respondents are aged 36 years
mean(advert_df$`Daily Internet Usage`)
## [1] 180.0001
#On average there is an internet usage of 180.0001
median(advert_df$`Area Income`)
## [1] 57012.3
```

```
#The median area income is 57012.3
hist(advert_df$`Daily Internet Usage`, main = 'Distribution of
Internet Usage',
      xlab = 'Time spent in minutes',col = 'pink',breaks=5)
```



```
#Majority used between 200 and 250 megabytes of data
x <-table(advert_df$Gender)
pie(x, main = 'Gender Distribution',)
```

Gender Distribution



```
#More females than men visited the blog website.
#Distribution on countries that visited the website
c <-table(advert_df$Country)
s <-sort(c,decreasing=FALSE)
as.matrix(s)
##                                     [,1]
## Aruba                             1
## Bermuda                           1
## British Indian Ocean Territory (Chagos Archipelago) 1
## Cape Verde                         1
## Germany                            1
## Jordan                             1
## Kiribati                           1
## Lesotho                            1
## Marshall Islands                   1
## Montserrat                         1
## Mozambique                         1
## Romania                            1
## Saint Kitts and Nevis              1
## Slovenia                           1
## Andorra                            2
```


## Argentina	2
## Benin	2
## Bhutan	2
## Central African Republic	2
## Colombia	2
## Comoros	2
## Djibouti	2
## Gambia	2
## Guadeloupe	2
## Guinea-Bissau	2
## Haiti	2
## India	2
## Kuwait	2
## Macedonia	2
## Mauritania	2
## Montenegro	2
## Namibia	2
## New Caledonia	2
## Norway	2
## Panama	2
## Pitcairn Islands	2
## Reunion	2
## Saint Barthelemy	2
## Saint Lucia	2
## Sao Tome and Principe	2
## Sierra Leone	2
## Slovakia (Slovak Republic)	2
## South Georgia and the South Sandwich Islands	2
## Sudan	2
## Suriname	2
## Swaziland	2
## Uzbekistan	2
## Antarctica (the territory South of 60 deg S)	3
## Armenia	3
## Azerbaijan	3
## British Virgin Islands	3
## Cook Islands	3
## Denmark	3
## Estonia	3
## Faroe Islands	3
## Gibraltar	3
## Guernsey	3
## Guinea	3
## Heard Island and McDonald Islands	3
## Holy See (Vatican City State)	3
## Iceland	3
## Ireland	3
## Isle of Man	3

## Lithuania	3
## Macao	3
## Malaysia	3
## Monaco	3
## Morocco	3
## Nauru	3
## Nepal	3
## Nicaragua	3
## Niger	3
## Niue	3
## Northern Mariana Islands	3
## Palestinian Territory	3
## Paraguay	3
## Portugal	3
## Russian Federation	3
## San Marino	3
## Seychelles	3
## Spain	3
## Syrian Arab Republic	3
## Tajikistan	3
## Tanzania	3
## Togo	3
## Trinidad and Tobago	3
## United Kingdom	3
## Vietnam	3
## Yemen	3
## Angola	4
## Bangladesh	4
## Burkina Faso	4
## Chad	4
## Chile	4
## Congo	4
## Cote d'Ivoire	4
## Dominican Republic	4
## Equatorial Guinea	4
## Falkland Islands (Malvinas)	4
## French Guiana	4
## Georgia	4
## Ghana	4
## Grenada	4
## Guam	4
## Guatemala	4
## Israel	4
## Japan	4
## Kazakhstan	4
## Kenya	4
## Lao People's Democratic Republic	4
## Latvia	4

## Libyan Arab Jamahiriya	4
## Malawi	4
## Maldives	4
## Mali	4
## Martinique	4
## Mauritius	4
## Netherlands	4
## New Zealand	4
## Palau	4
## Saint Martin	4
## Saudi Arabia	4
## Sri Lanka	4
## Sweden	4
## Switzerland	4
## Thailand	4
## Tokelau	4
## Tunisia	4
## Tuvalu	4
## Uganda	4
## United States Minor Outlying Islands	4
## United States Virgin Islands	4
## Wallis and Futuna	4
## Zambia	4
## American Samoa	5
## Antigua and Barbuda	5
## Austria	5
## Bahrain	5
## Barbados	5
## Belgium	5
## Belize	5
## Bouvet Island (Bouvetoya)	5
## Brazil	5
## Brunei Darussalam	5
## Cameroon	5
## Canada	5
## Cayman Islands	5
## Cuba	5
## Dominica	5
## Ecuador	5
## Egypt	5
## Finland	5
## French Polynesia	5
## French Southern Territories	5
## Greenland	5
## Guyana	5
## Honduras	5
## Iran	5
## Italy	5

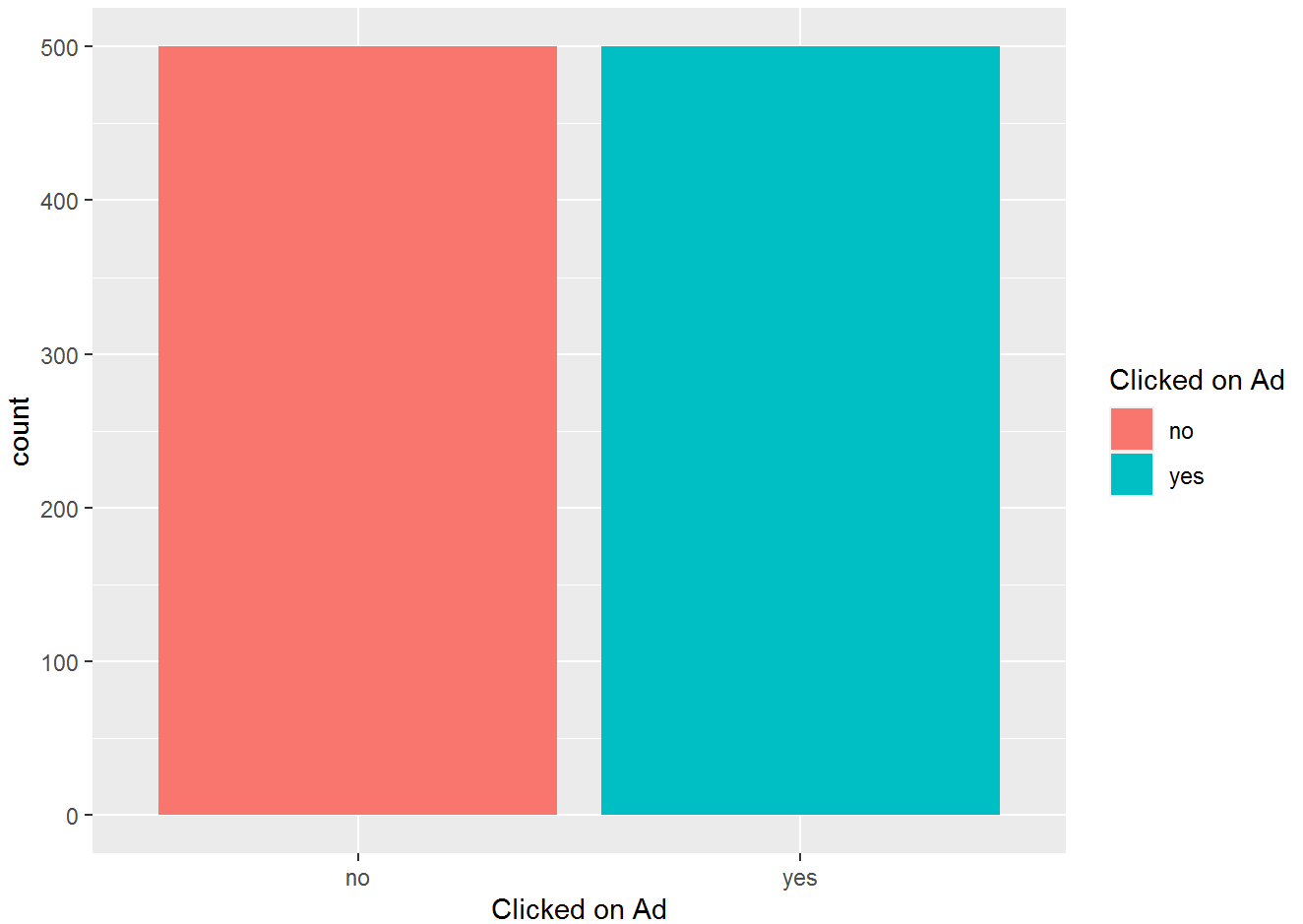
## Jamaica	5
## Korea	5
## Myanmar	5
## Norfolk Island	5
## Pakistan	5
## Papua New Guinea	5
## Rwanda	5
## Saint Helena	5
## Saint Pierre and Miquelon	5
## Serbia	5
## Somalia	5
## Timor-Leste	5
## Tonga	5
## Turks and Caicos Islands	5
## Ukraine	5
## United States of America	5
## Uruguay	5
## Algeria	6
## Anguilla	6
## Belarus	6
## Bolivia	6
## Bulgaria	6
## China	6
## Christmas Island	6
## Costa Rica	6
## Croatia	6
## El Salvador	6
## Gabon	6
## Hong Kong	6
## Hungary	6
## Indonesia	6
## Jersey	6
## Kyrgyz Republic	6
## Lebanon	6
## Liechtenstein	6
## Madagascar	6
## Malta	6
## Mayotte	6
## Mexico	6
## Moldova	6
## Mongolia	6
## Netherlands Antilles	6
## Philippines	6
## Poland	6
## Puerto Rico	6
## Qatar	6
## Saint Vincent and the Grenadines	6
## Samoa	6

## Singapore	6
## Svalbard & Jan Mayen Islands	6
## Turkmenistan	6
## United Arab Emirates	6
## Vanuatu	6
## Zimbabwe	6
## Albania	7
## Bahamas	7
## Bosnia and Herzegovina	7
## Burundi	7
## Cambodia	7
## Eritrea	7
## Ethiopia	7
## Fiji	7
## Luxembourg	7
## Taiwan	7
## Venezuela	7
## Western Sahara	7
## Afghanistan	8
## Australia	8
## Cyprus	8
## Greece	8
## Liberia	8
## Micronesia	8
## Peru	8
## Senegal	8
## South Africa	8
## Turkey	8
## Czech Republic	9
## France	9

```

ggplot(data =advert_df, aes(x = `Clicked on Ad`,fill= `Clicked on
Ad`)) +
  geom_bar()

```

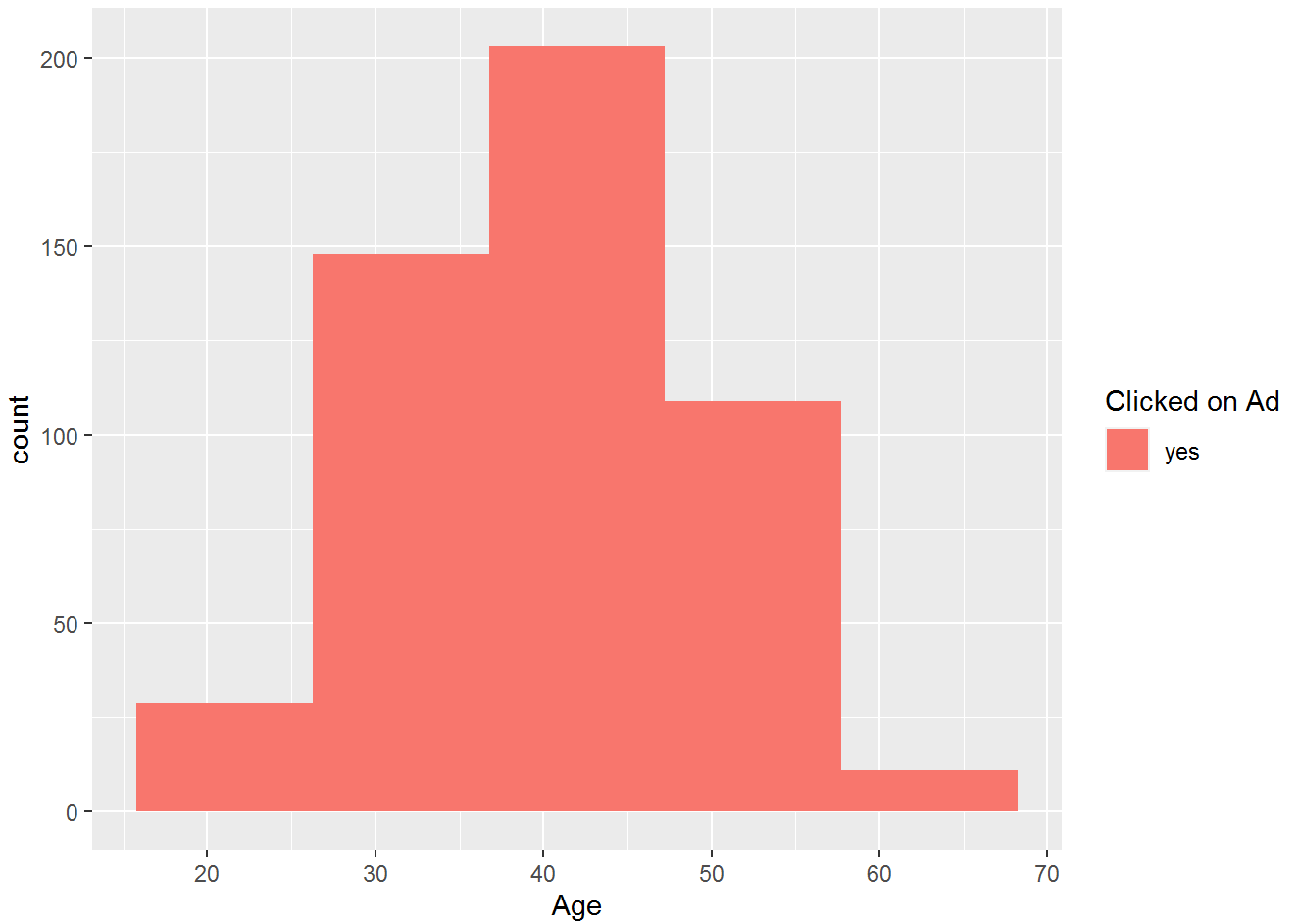


#The data set has an equal distribution of ad viewers with 50% having viewed the ads and 50% who did no. Therefore the data set is balanced

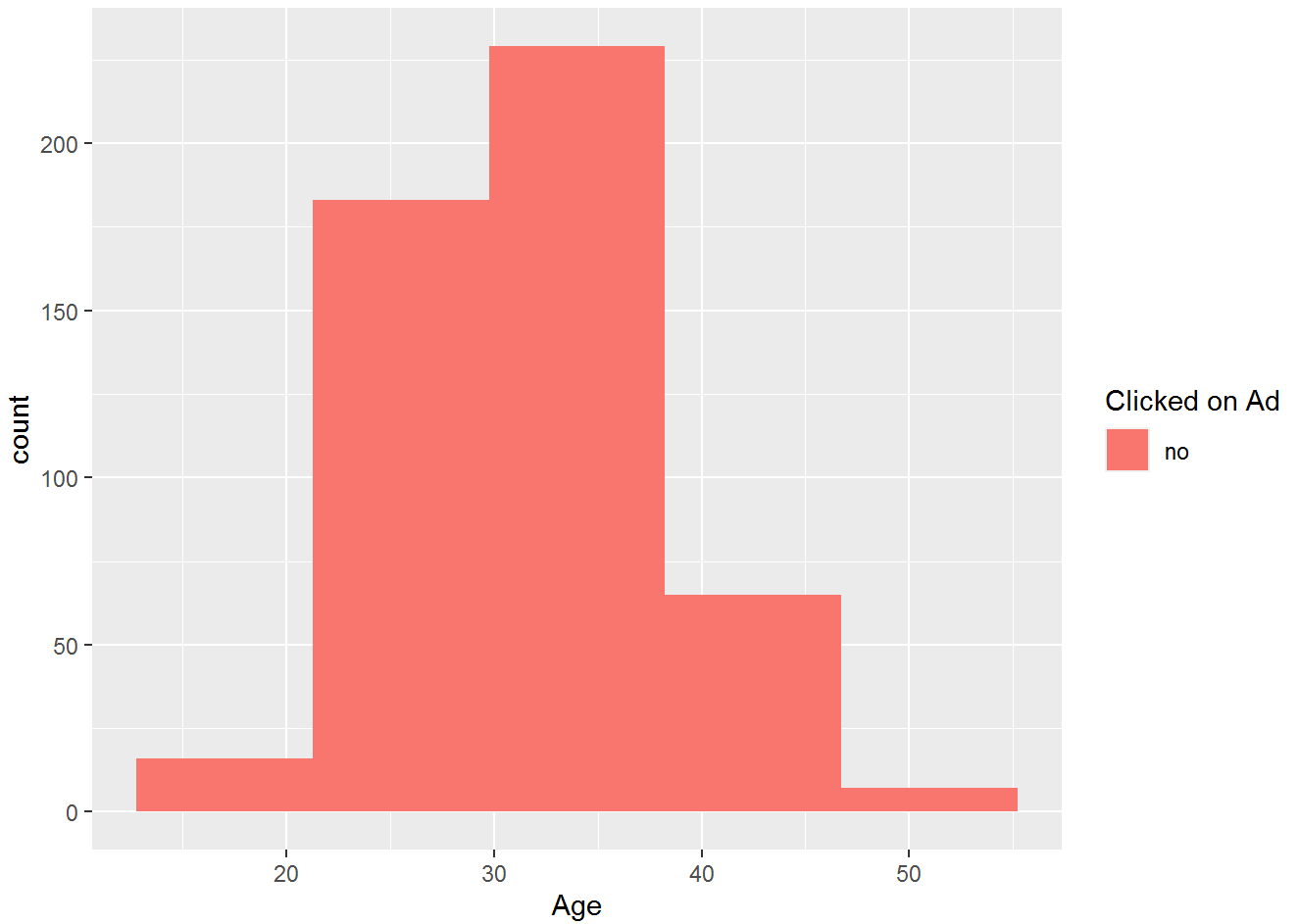
Bivariate Analysis

Clicks on Ad and Age

```
#A data frame with 'clicked on Ad ='Yes'  
ad_df<-advert_df[advert_df$`Clicked on Ad`=='yes',]  
ggplot(data=ad_df, aes(x=Age,fill=`Clicked on Ad`))+  
  geom_histogram(position= 'stack',bins=5)
```



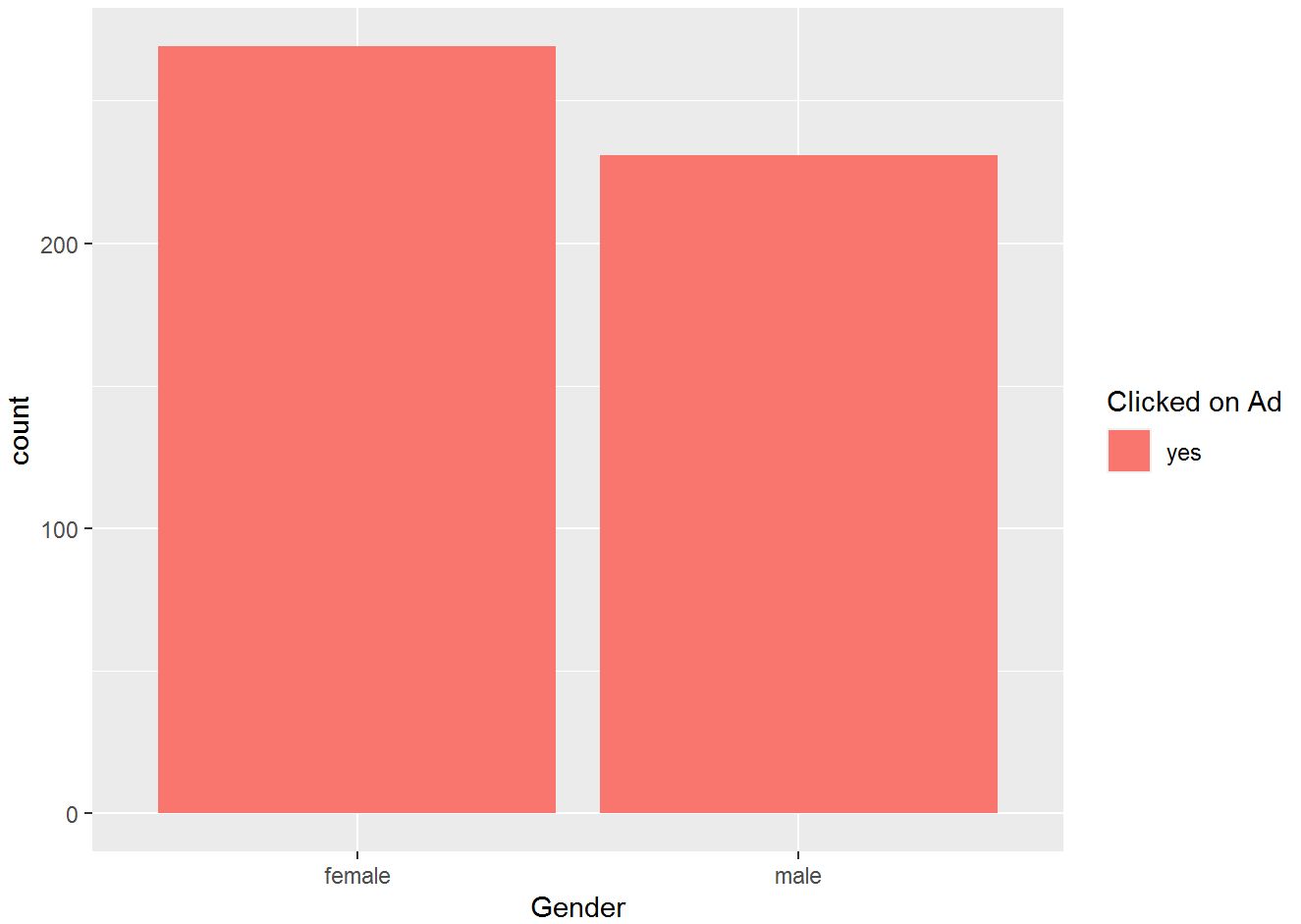
```
#A data frame with 'clicked on Ad ='no'  
ad_df1<-advert_df[advert_df`Clicked on Ad`=='no',]  
ggplot(data=ad_df1, aes(x=Age,fill=`Clicked on Ad`))+  
  geom_histogram(position= 'stack',bins=5)
```



#The age 40 and 50 were more likely to click on the Ad than those in the aged between(30-40)

Clicks on Ad and Gender

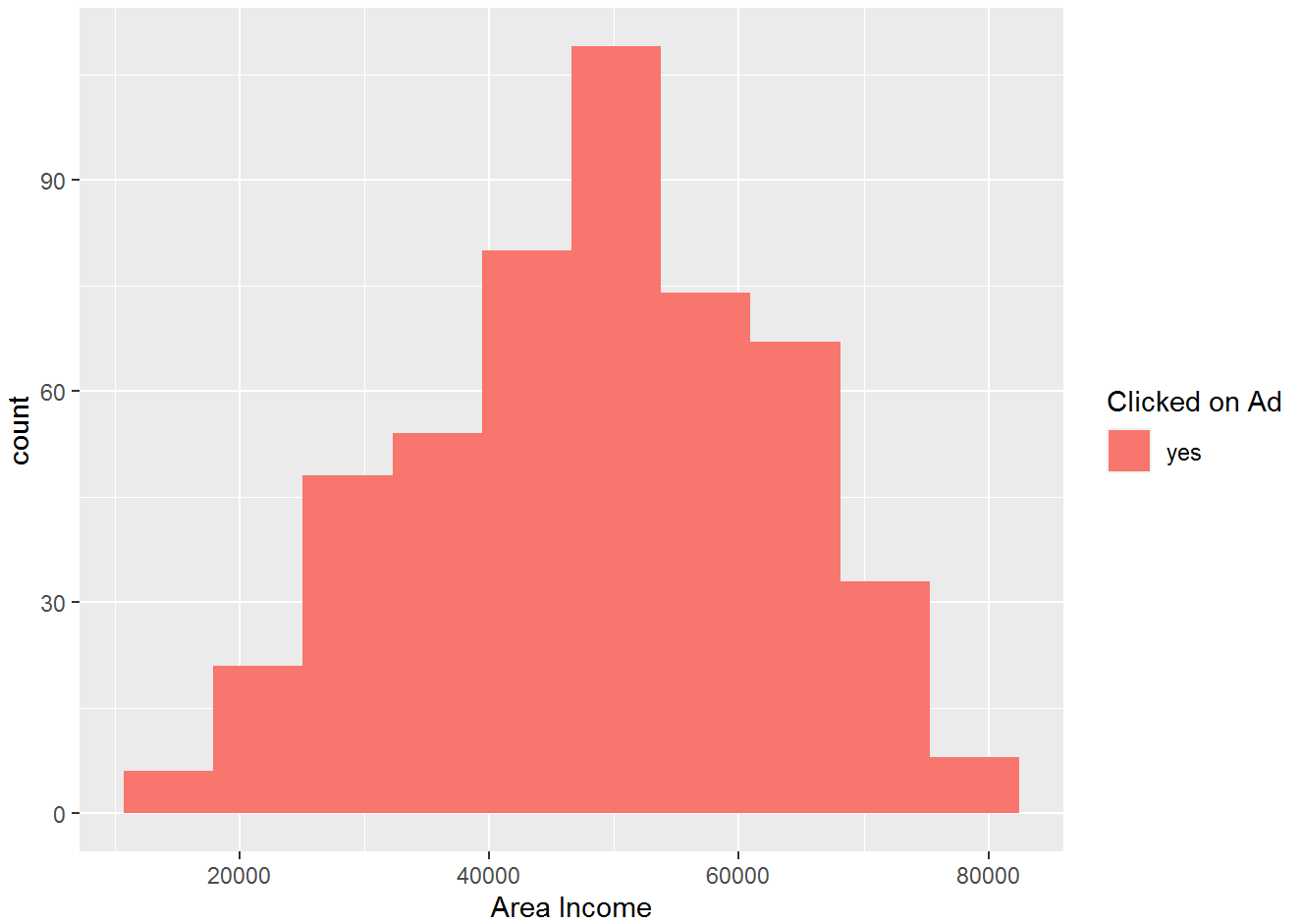
```
ggplot(data=ad_df, aes(x=Gender, fill=`Clicked on Ad`))+  
  geom_bar(position= 'dodge')
```

#More women clicked on the ads than men.

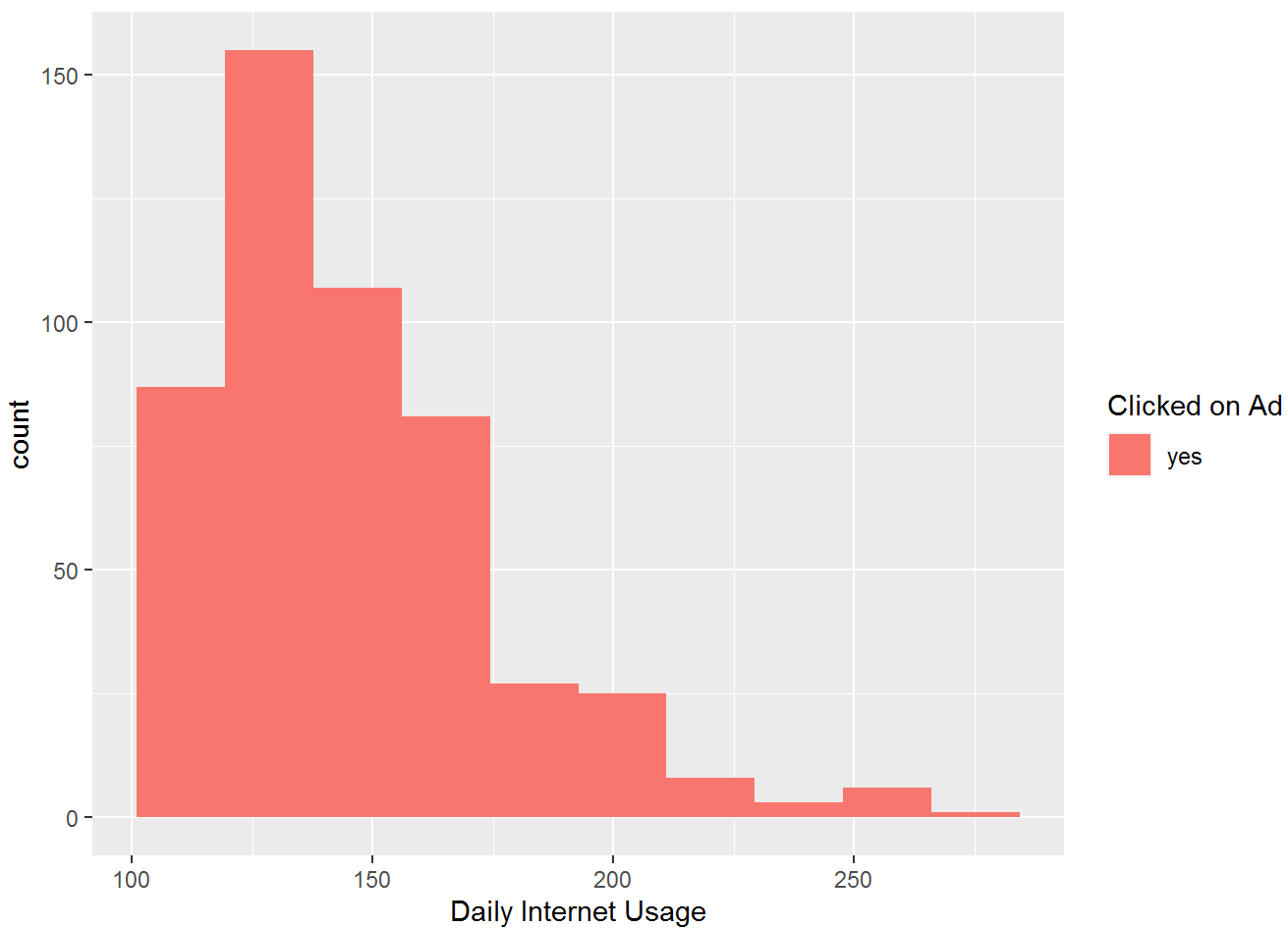
Clicks on Ad and Area income

```
ggplot(data=ad_df, aes(x=`Area Income`,fill=`Clicked on Ad`))+  
  geom_histogram(position= 'stack',bins=10)
```



Clicks on Ad and Daily Internet Usage

```
ggplot(data=ad_df, aes(x=`Daily Internet Usage`,fill=`Clicked on  
Ad`))+  
  geom_histogram(position= 'stack',bins=10)
```



```
c <-table(ad_df$Country)
s <-sort(c,decreasing=FALSE)
max(s)
## [1] 7
as.matrix(s)
##
##                                     [,1]
## Angola                             1
## Argentina                           1
## Armenia                             1
## Austria                              1
## Azerbaijan                           1
## Benin                                1
## Bhutan                               1
## British Indian Ocean Territory (Chagos Archipelago) 1
## British Virgin Islands               1
## Burkina Faso                         1
## Central African Republic             1
## Colombia                             1
## Comoros                              1
## Cook Islands                         1
## Djibouti                             1
```

## Estonia	1
## Finland	1
## French Polynesia	1
## French Southern Territories	1
## Gambia	1
## Germany	1
## Greenland	1
## Guadeloupe	1
## Guinea-Bissau	1
## Haiti	1
## Holy See (Vatican City State)	1
## Iceland	1
## Ireland	1
## Isle of Man	1
## Italy	1
## Kiribati	1
## Kuwait	1
## Kyrgyz Republic	1
## Macedonia	1
## Mali	1
## Marshall Islands	1
## Mauritania	1
## Mauritius	1
## Monaco	1
## Montserrat	1
## Morocco	1
## Myanmar	1
## Namibia	1
## Nauru	1
## Norway	1
## Pakistan	1
## Paraguay	1
## Pitcairn Islands	1
## Portugal	1
## Romania	1
## Russian Federation	1
## Saint Kitts and Nevis	1
## Saint Lucia	1
## San Marino	1
## Seychelles	1
## Singapore	1
## Slovenia	1
## South Georgia and the South Sandwich Islands	1
## Suriname	1
## Sweden	1
## Syrian Arab Republic	1
## Tanzania	1
## Timor-Leste	1

## Togo	1
## Tunisia	1
## Ukraine	1
## Uruguay	1
## Uzbekistan	1
## Vanuatu	1
## Wallis and Futuna	1
## Andorra	2
## Antarctica (the territory South of 60 deg S)	2
## Bahrain	2
## Bangladesh	2
## Barbados	2
## Belgium	2
## Bouvet Island (Bouvetoya)	2
## Brunei Darussalam	2
## Burundi	2
## Cambodia	2
## Chad	2
## Costa Rica	2
## Denmark	2
## Dominica	2
## Dominican Republic	2
## Ecuador	2
## Falkland Islands (Malvinas)	2
## Faroe Islands	2
## Georgia	2
## Ghana	2
## Grenada	2
## Guam	2
## Guernsey	2
## Guinea	2
## Heard Island and McDonald Islands	2
## Honduras	2
## Israel	2
## Jamaica	2
## Japan	2
## Kazakhstan	2
## Lao People's Democratic Republic	2
## Libyan Arab Jamahiriya	2
## Madagascar	2
## Malawi	2
## Maldives	2
## Moldova	2
## Montenegro	2
## Netherlands Antilles	2
## New Caledonia	2
## New Zealand	2
## Niger	2

## Norfolk Island	2
## Northern Mariana Islands	2
## Palau	2
## Palestinian Territory	2
## Qatar	2
## Rwanda	2
## Saint Barthelemy	2
## Saint Helena	2
## Saint Martin	2
## Sao Tome and Principe	2
## Sierra Leone	2
## Somalia	2
## Tajikistan	2
## Thailand	2
## Tonga	2
## Trinidad and Tobago	2
## Turkmenistan	2
## United Kingdom	2
## United States Minor Outlying Islands	2
## United States Virgin Islands	2
## Vietnam	2
## Yemen	2
## Algeria	3
## American Samoa	3
## Anguilla	3
## Belarus	3
## Belize	3
## Bosnia and Herzegovina	3
## Brazil	3
## Canada	3
## Cayman Islands	3
## Chile	3
## Congo	3
## Cote d'Ivoire	3
## Egypt	3
## Equatorial Guinea	3
## Eritrea	3
## Fiji	3
## French Guiana	3
## Greece	3
## Guatemala	3
## Guyana	3
## Iran	3
## Korea	3
## Lithuania	3
## Luxembourg	3
## Macao	3
## Malta	3

## Martinique	3
## Netherlands	3
## Papua New Guinea	3
## Philippines	3
## Poland	3
## Puerto Rico	3
## Saint Pierre and Miquelon	3
## Saint Vincent and the Grenadines	3
## Saudi Arabia	3
## Serbia	3
## Spain	3
## Switzerland	3
## Tokelau	3
## Turks and Caicos Islands	3
## Tuvalu	3
## United Arab Emirates	3
## United States of America	3
## Venezuela	3
## Zambia	3
## Albania	4
## Antigua and Barbuda	4
## Bahamas	4
## Bulgaria	4
## China	4
## Christmas Island	4
## Cuba	4
## Cyprus	4
## Czech Republic	4
## El Salvador	4
## Hong Kong	4
## Indonesia	4
## Jersey	4
## Kenya	4
## Latvia	4
## Lebanon	4
## Mexico	4
## Micronesia	4
## Mongolia	4
## Samoa	4
## Svalbard & Jan Mayen Islands	4
## Taiwan	4
## Uganda	4
## Western Sahara	4
## Zimbabwe	4
## Afghanistan	5
## France	5
## Hungary	5
## Mayotte	5

## Peru	5
## Senegal	5
## Liberia	6
## Liechtenstein	6
## South Africa	6
## Australia	7
## Ethiopia	7
## Turkey	7

Conclusion

1. 38 more women than men click on the ad
2. People from Czech Republic and France contributed the highest among those who visited the blog website, however people from Turkey, Ethiopia and Australia contributed the highest among those who clicked on the Ads.
3. On average , people spent 65 minutes on the blog site and an average of 180 megabyte on daily internet usage
4. The average age is 36 years
5. Persons falling under the average area income (57012.3) were more likely to click on ads.
6. Majority of those who clicked on the Ads were between the ages of 40-50 while majority of those who skipped Ads were between ages 30-40.

Recommendations

1. Increase Ads friendly to people from Czech Republic and France since they visit the blog the most.
2. Increase female and male friendly content in the blog in increase female consumption to increase number of Ad consumption
3. Make more targeted ads for people between the ages of 30 - 40 years