Facebook Ad Camp Analysis

Surya Aenuganti Ushasri

February 8, 2019

Loading data

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.2
```

```
## -- Attaching packages ---------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0
```

```
## Warning: package 'readr' was built under R version 3.5.2
```

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

```
dat = read.csv("KAG_conversion_data.csv")
#View(dat)
glimpse(dat)
```

```
## Observations: 1,143
## Variables: 11
## $ ad_id               <int> 708746, 708749, 708771, 708815, 708818, 70...
## $ xyz_campaign_id     <int> 916, 916, 916, 916, 916, 916, 916, 916, 91...
## $ fb_campaign_id      <int> 103916, 103917, 103920, 103928, 103928, 10...
## $ age                 <fct> 30-34, 30-34, 30-34, 30-34, 30-34, 30-34, ...
## $ gender              <fct> M, M, M, M, M, M, M, M, M, M, M, M, M, M, ...
## $ interest            <int> 15, 16, 20, 28, 28, 29, 15, 16, 27, 28, 31...
## $ Impressions         <int> 7350, 17861, 693, 4259, 4133, 1915, 15615,...
## $ Clicks              <int> 1, 2, 0, 1, 1, 0, 3, 1, 1, 3, 0, 0, 0, 0, ...
## $ Spent               <dbl> 1.43, 1.82, 0.00, 1.25, 1.29, 0.00, 4.77, ...
## $ Total_Conversion    <int> 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ Approved_Conversion <int> 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, ...
```

```
unique(dat$age)
```

```
## [1] 30-34 35-39 40-44 45-49
## Levels: 30-34 35-39 40-44 45-49
```

Adding new columns

Click Rate(click_rate)

```
#clicks per 10000 impressions
dat<- dat %>%
  mutate(click_rate = as.factor(ifelse(dat$Impressions != 0 , round(dat$Clicks/dat$Impressions*10000) ,
```

Conversion Rate(conv_rate)

```
#Approved conversions as a percentage of totatl conversions
dat<- dat %>%
  mutate(conv_rate = ifelse(dat$Total_Conversion != 0 , round(dat$Approved_Conversion/dat$Total_Conversi
```

Conversion rate as factor

```
dat$conv_rate <-cut(dat$conv_rate, seq(0,100,10), right=TRUE, labels=c("0-10%","10-20%","20-30%","30-40%
dat$conv_rate[is.na(dat$conv_rate)] <- "0-10%"
```

Loss amount

```
dat<- dat %>%
  mutate(loss_amount = ifelse(dat$Clicks !=0, dat$Spent*(1 - dat$Approved_Conversion / dat$Clicks),0))
```

Loss rate

```
dat<- dat %>%
  mutate(loss_rate =  ifelse(dat$Spent>0, round(dat$loss_amount/dat$Spent)*100,0))
dat$loss_rate <-cut(dat$loss_rate, seq(0,100,10), right=TRUE, labels=c("0-10%","10-20%","20-30%","30-40%
dat$loss_rate[is.na(dat$loss_rate)] <- "0-10%"
```

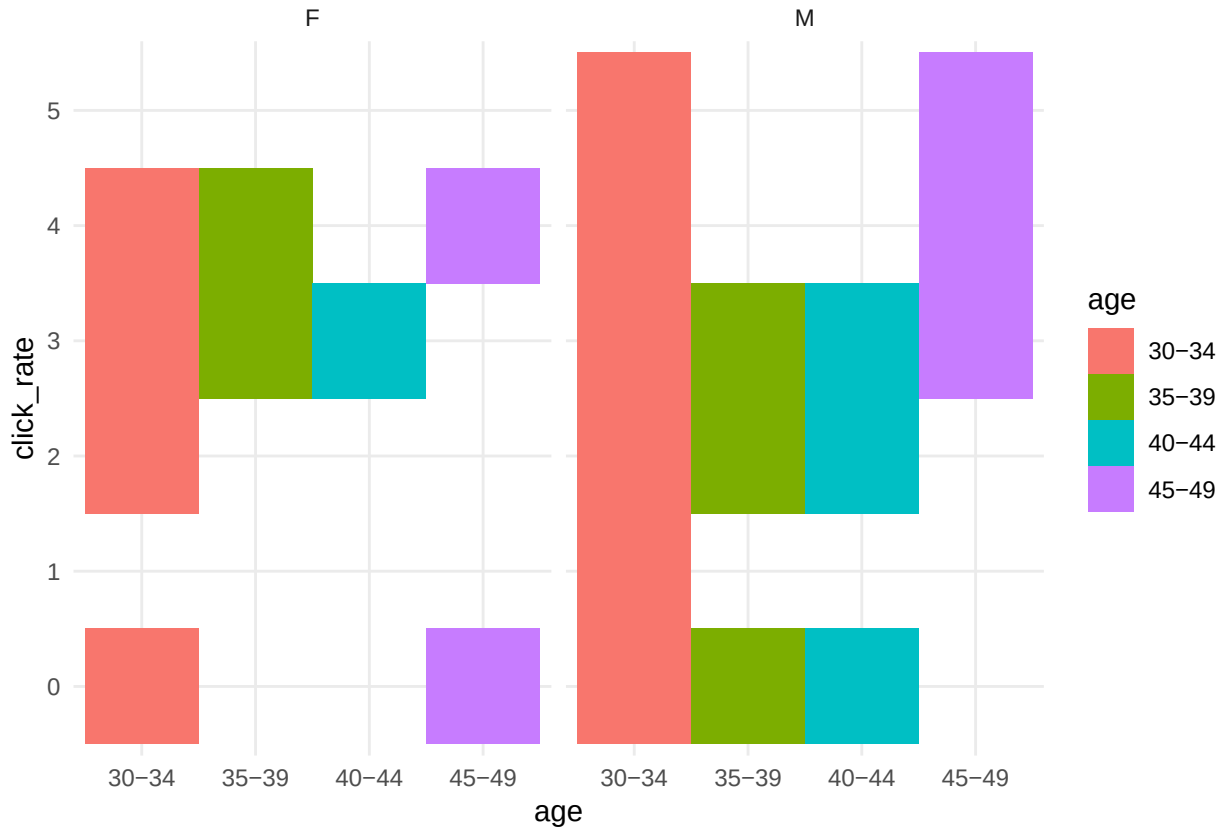
Data Analysis

```
table(dat$xyz_campaign_id)
```

```
##
##  916  936 1178
##   54  464  625
```

There are 3 companies and the analysis will be done seperately

```
dat_916<-dat %>% filter(xyz_campaign_id %in% c("916"))
library(ggplot2)

ggplot(data = dat_916) +
  aes(x = age, y = click_rate, fill = age) +
  geom_tile() +
  theme_minimal() +
  facet_wrap(vars(gender))
```

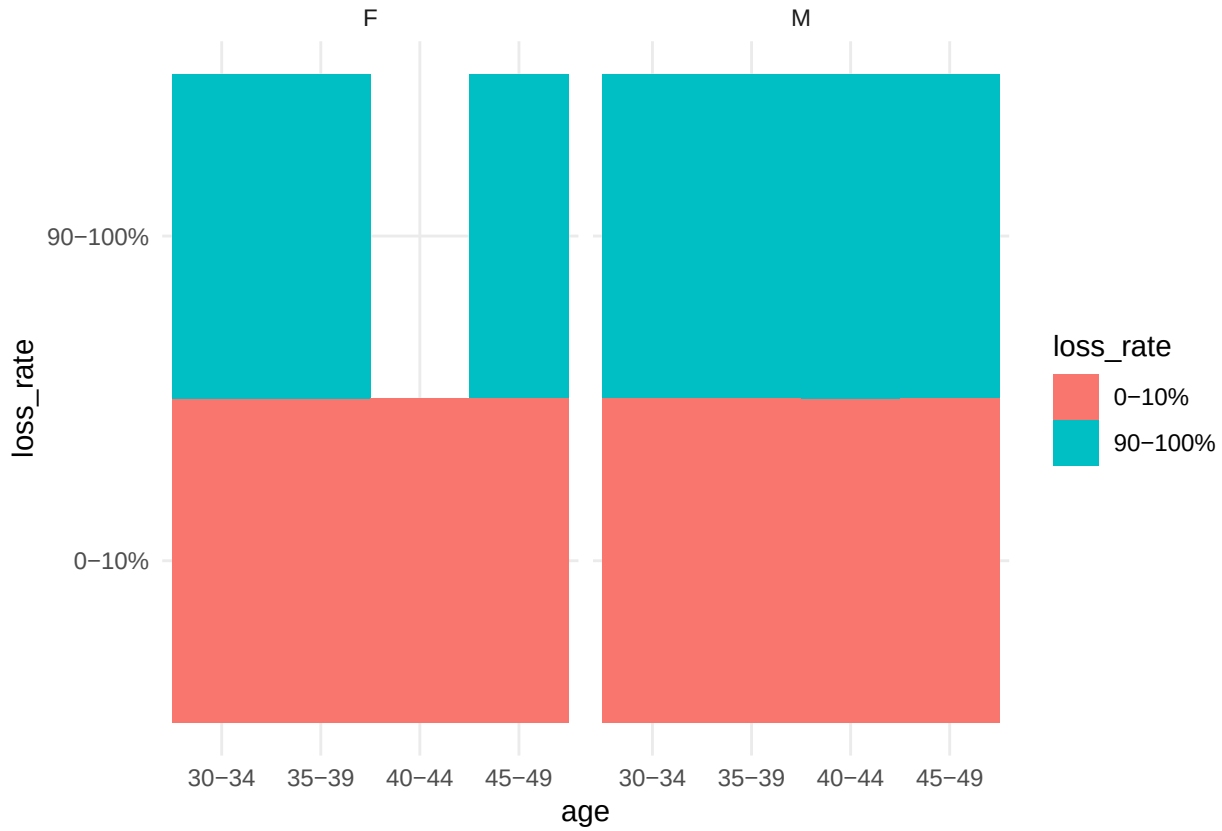


Analysis: For both gender groups Male and Female the age group 30-34 seems to respond with higher click rate ranging from 2 to 5

But we cannot say anything conclusive about other groups because we don't have sufficient data in these groups

Return on spent

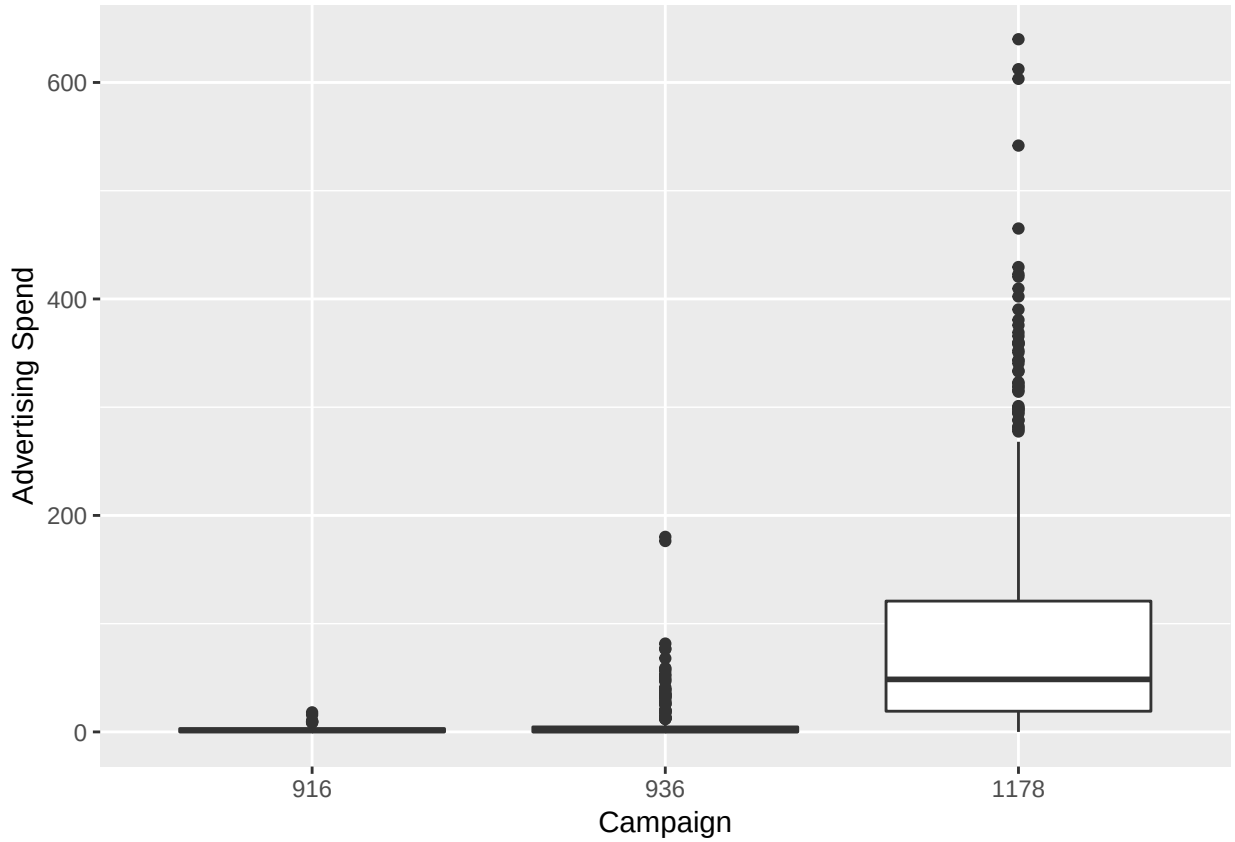
```
library(ggplot2)

ggplot(data = dat_916) +
  aes(x = age, y = loss_rate, fill = loss_rate) +
  geom_tile() +
  theme_minimal() +
  facet_wrap(vars(gender))
```



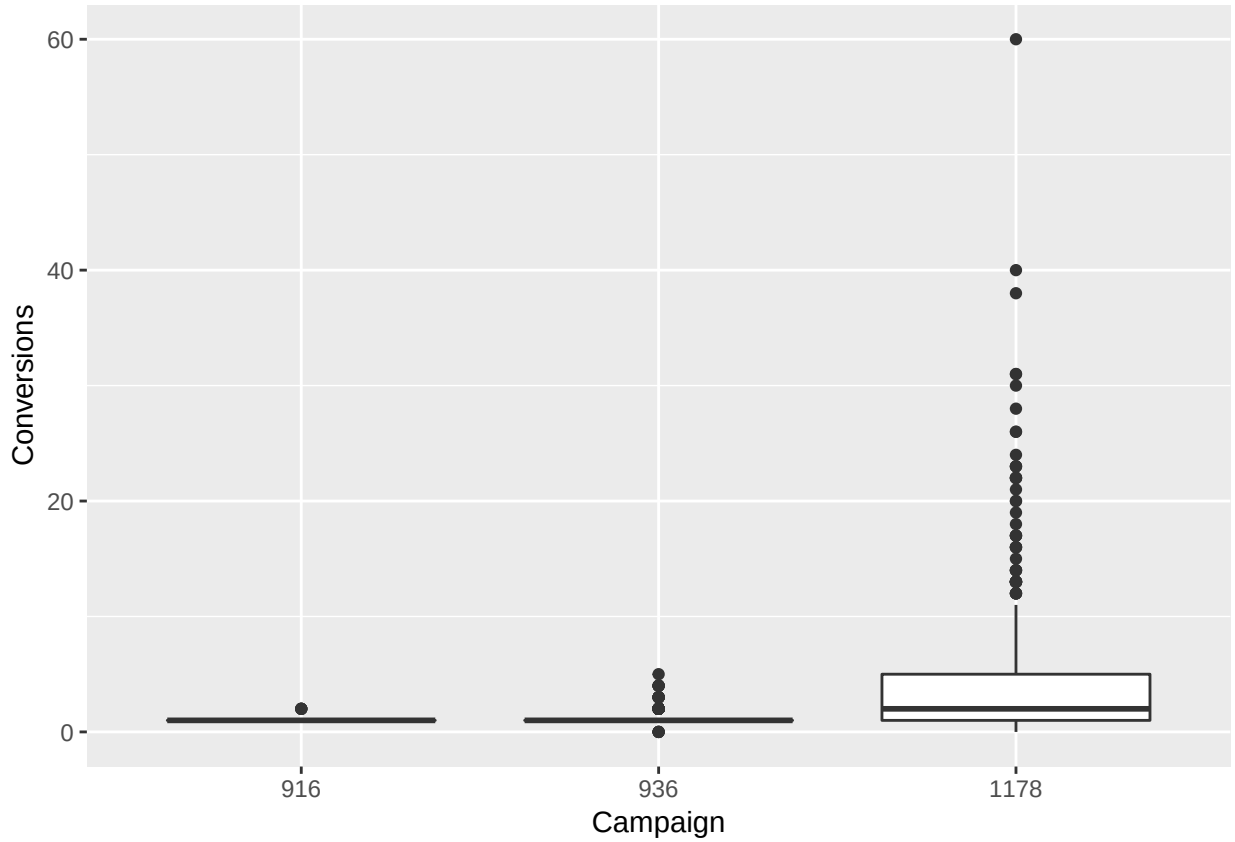
We cannot conclude anything from these plots because in all the age groups of both genders there is a 50% loss of 0-10

Analysis on a specific ad campaign

```
ggplot(dat, aes(as.factor(xyz_campaign_id), Spent)) + geom_boxplot() + labs(x = "Campaign", y = "Advert
```




```
ggplot(dat, aes(as.factor(xyz_campaign_id), Total_Conversion)) + geom_boxplot() + labs(x = "Campaign", y
```



From the above first box-plot we can see that company 1178 spends more on advertisements comparatively

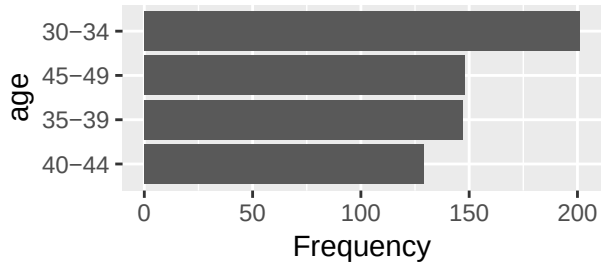
And the second box plot shows there is a more total conversion rate for company 1178

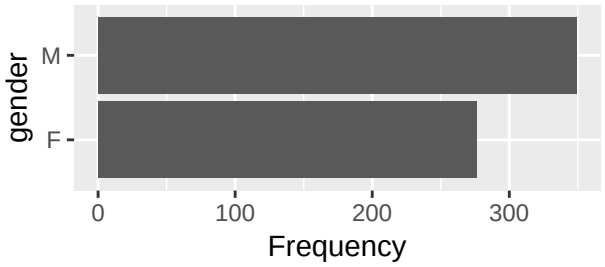
```
dat1178 <- dat %>%
  filter(xyz_campaign_id == 1178)
```

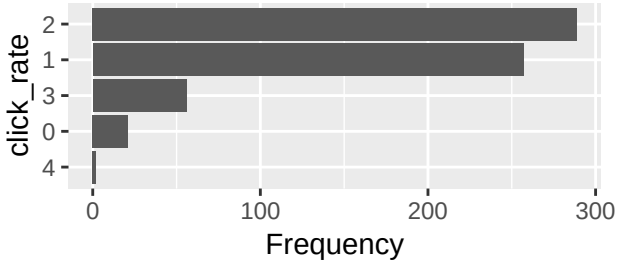
```
library(DataExplorer)
```

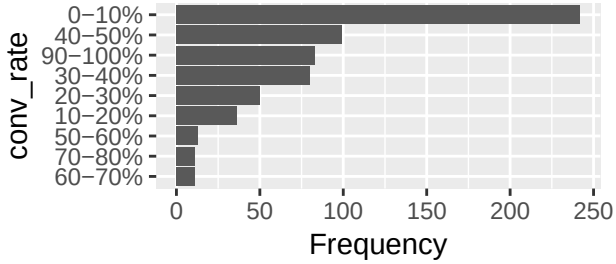
```
## Warning: package 'DataExplorer' was built under R version 3.5.2
```

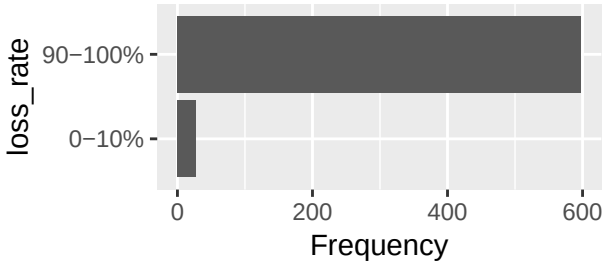
```
#Checking the frequency data variable by variable
options(repr.plot.width=4, repr.plot.height=4)
plot_bar(dat1178)
```



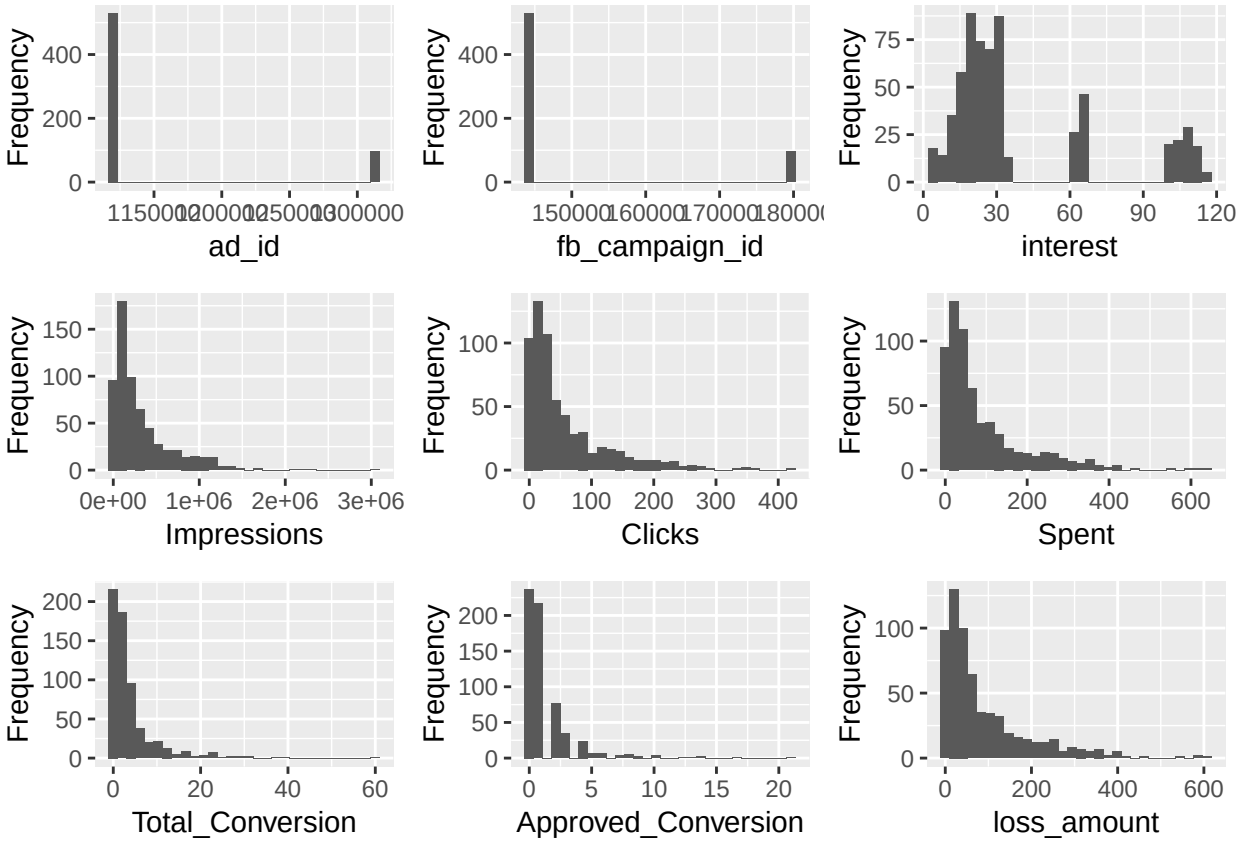




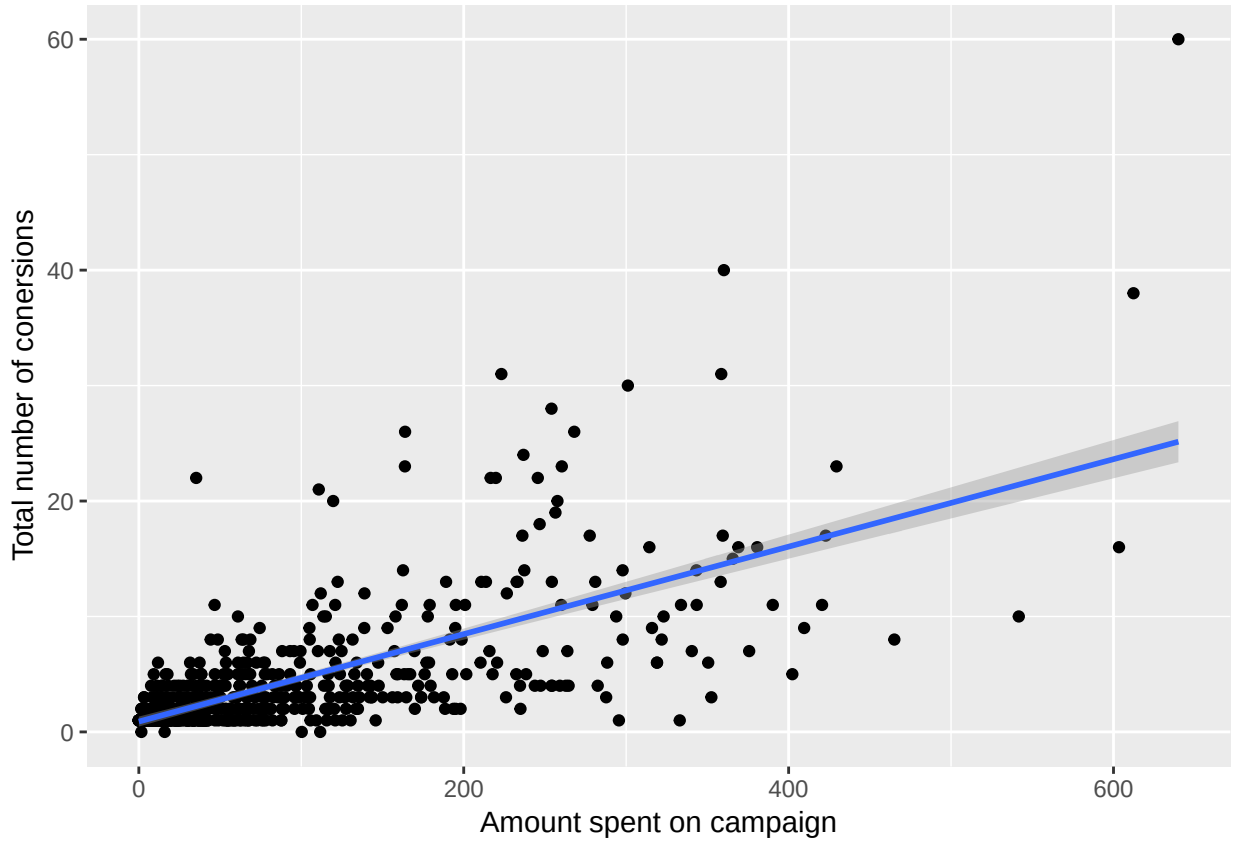




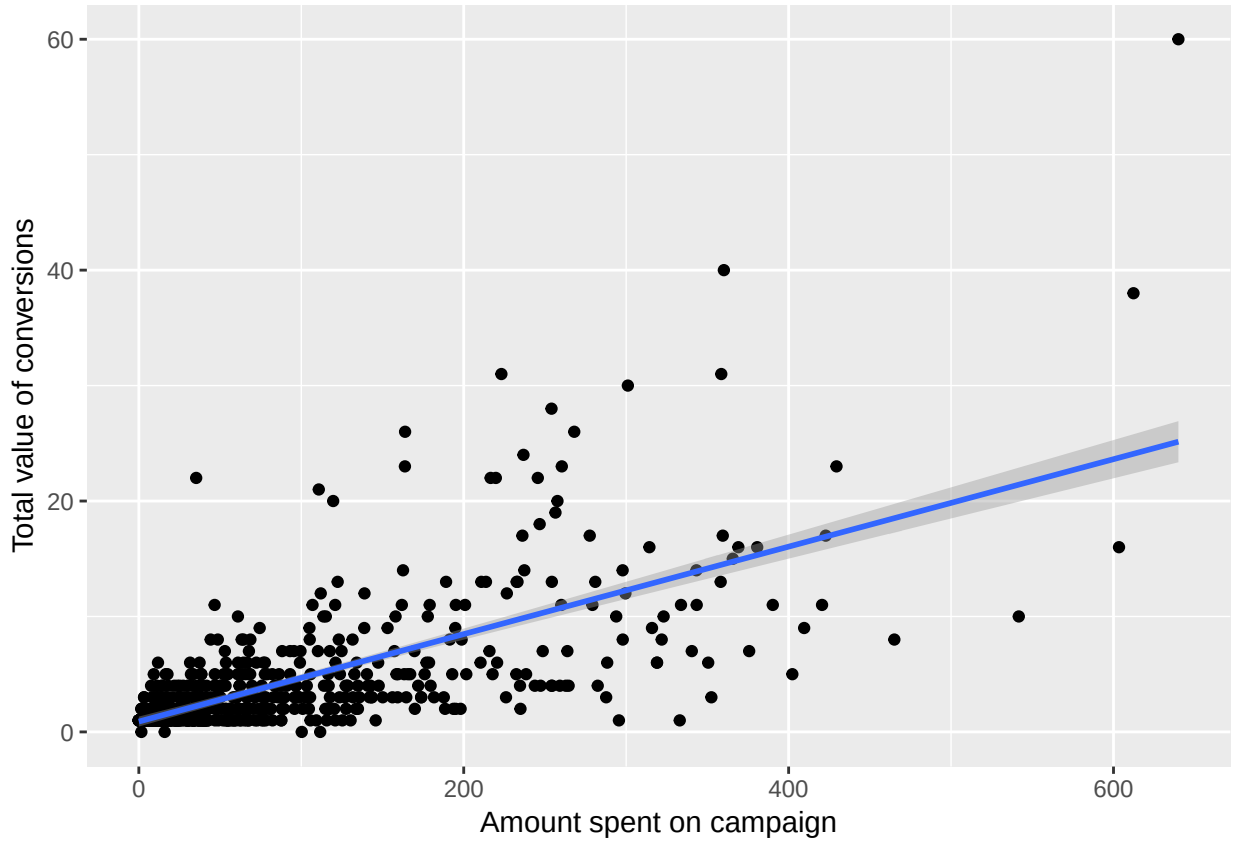

```
options(repr.plot.width=8, repr.plot.height=4)
plot_histogram(dat1178[,-2]) #removing xyzcampaign id
```




```
#Preliminary analysis of campaign 1178
#Let's look at what happens to the number of conversions and the value of our conversions when we spend
options(repr.plot.width=6, repr.plot.height=3)
ggplot(dat1178, aes(Spent, Total_Conversion)) + geom_point() + geom_smooth(method = "lm") +
  labs(x = "Amount spent on campaign", y = "Total number of conersions")
```




```
ggplot(dat1178, aes(Spent, Total_Conversion)) + geom_point() + geom_smooth(method = "lm") +
  labs(x = "Amount spent on campaign", y = "Total value of conversions")
```



Analysis:Here, itlooks like the more we spend, the more we get back. But the amount of data is quite sparse at the right-hand side so we cannot say that the statement is accurate without considering more analysis.

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.5.2
```

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

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

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

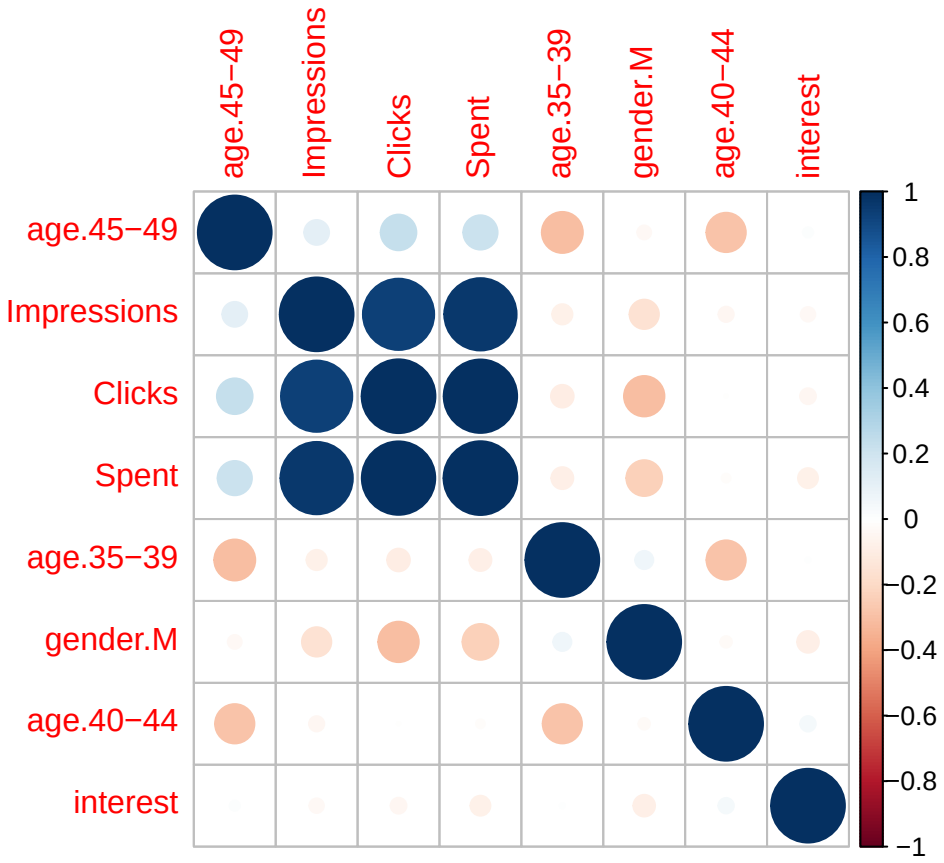
```
ds_1178_predict<-dat1178 %>% select( age, gender,interest, Impressions,Clicks,Spent)
set.seed(1234)
ds_1178_predict_Dummy <- dummyVars("~.",data=ds_1178_predict, fullRank=T)
ds_1178_predict_final <- as.data.frame(predict(ds_1178_predict_Dummy,ds_1178_predict))
print(names(ds_1178_predict_final))
```

```
## [1] "age.35-39"   "age.40-44"   "age.45-49"   "gender.M"    "interest"
## [6] "Impressions" "Clicks"      "Spent"
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corMatMy <- cor(ds_1178_predict_final)
corrplot(corMatMy, order = "hclust")
```



From the correlation matrix we can see that it's pretty obvious that the clicks are strongly correlated to impressions and spent.