**SOCIAL MEDIA ANALYTICS**

**LAB EXPERIMENT 4**

**AIM:**

To implement Exploratory Data Analysis (EDA) and visualization of social media data for business.

**THEORY:**

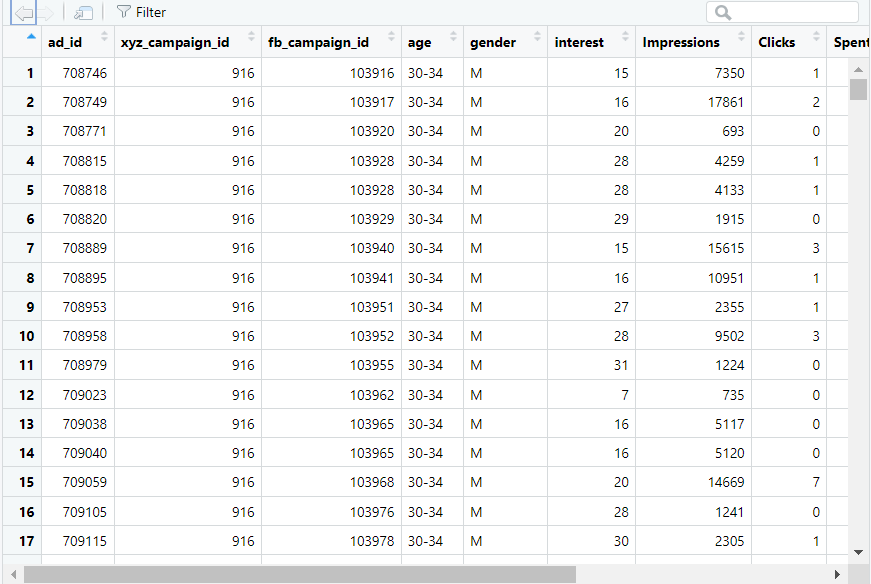
In R, EDA is performed using a combination of statistical methods and visualization techniques. Exploratory Data Analysis (EDA) involves the process of visually and statistically exploring datasets to understand their structure, identify patterns, detect anomalies, and formulate hypotheses.

Visualization in R involves the creation of graphical representations of data using various plotting functions and packages. interactive visualization libraries in R such as plotly allow for the creation of dynamic and interactive plots.

**Procedure:**

**(‘Facebook’ – social media data chosen)**

**Dataset:**



**Program: (R Language)**

**Data Cleaning, Preprocessing and Loading:**

library(tidyverse)

# Dataset retrieved from facebook.

data <- read\_excel("C:\\Users\\ramri\\Desktop\\KAG\_conversion\_data.xlsx")

#Overview of Data.

glimpse(data)

# Unique age groups.

unique(data$age)

#Duplicate Copy of data.

dataTf <- data

dataTf$age[dataTf$age == '30-34'] <- 32

dataTf$age[dataTf$age == '35-39'] <- 37

dataTf$age[dataTf$age == '40-44'] <- 42

dataTf$age[dataTf$age == '45-49'] <- 47

#convert age variable datatype to integer.

dataTf$age <- as.integer(dataTf$age)

#check age variable.

unique(dataTf$age)

str(dataTf$age)

# Make Gender -> Male (as 0), Female (as 1).

dataTf$gender[dataTf$gender == 'M'] <- 0

dataTf$gender[dataTf$gender == 'F'] <- 1

dataTf$gender <- as.integer(dataTf$gender)

# abbreviate some variable names.

dataTf <- dataTf %>%

rename(xyzCampId = xyz\_campaign\_id, fbCampId = fb\_campaign\_id, impr = Impressions,

conv = Total\_Conversion, appConv = Approved\_Conversion)

# Final Preprocessed data.

glimpse(dataTf)

**Exploratory Data Analysis and Visualisation:**

# Exploratory Data Analysis and Visualisation.

# Heatmap Visualisation.

library(heatmaply)

dataMatNorm <- as.matrix(normalize(dataTf, method = "standardize"))

heatmap(dataMatNorm)

dataTf <- dataTf %>%

mutate(CTR = ((Clicks / impr) \* 100), CPC = Spent / Clicks)

dataTf$CTR <- round(dataTf$CTR, 4)

dataTf$CPC <- round(dataTf$CPC, 2)

glimpse(dataTf)

# Trim the dataset.

dataTfTrim <- dataTf %>%

select(CTR, CPC, appConv, conv, impr, Spent,Clicks)

#Heatmap.

heatmap(cor(normalize(na.omit(dataTfTrim))))

#Set plot size options.

options(repr.plot.width = 4, repr.plot.height = 3)

# Boxplot

ggplot(dataTf, aes(as.factor(xyzCampId), Spent)) + geom\_boxplot() + labs(x = "Campaign", y = "Advertising Spend")

ggplot(dataTf, aes(as.factor(xyzCampId), conv)) + geom\_boxplot() + labs(x = "Campaign", y = "Conversions")

data\_id1178 <- data %>%

rename(xyzCampId = xyz\_campaign\_id, fbCampId = fb\_campaign\_id, impr = Impressions,

conv = Total\_Conversion, appConv = Approved\_Conversion) %>%

filter(xyzCampId == 1178)

glimpse(data\_id1178)

#Data Explorer

library(DataExplorer)

plot\_missing(data\_id1178)

# Bar Plot.

options(repr.plot.width = 4, repr.plot.height = 4)

plot\_bar(data\_id1178)

# Histogram.

options(repr.plot.width = 8, repr.plot.height = 4)

plot\_histogram(data\_id1178)

# Correlation map.

plot\_correlation(data\_id1178)

# Feature Engineering (introduce totConv, conVal, appConval)

data\_id1178 <- data\_id1178 %>%

mutate(totConv = conv + appConv,

conVal = conv \* 5,

appConVal = appConv \* 100) %>%

mutate(totConVal = conVal + appConVal) %>%

mutate(costPerCon = round(Spent / totConv, 2),

ROAS = round(totConVal / Spent, 2))

data\_id1178 <- data\_id1178 %>%

mutate(CPM = round((Spent / impr) \* 1000, 2))

# Added new variables

head(data\_id1178)

# Smooth and point graph.

options(repr.plot.width=6, repr.plot.height=3)

ggplot(data\_id1178, aes(Spent, totConv)) + geom\_point() + geom\_smooth(method = "lm") +

labs(x = "Amount spent on campaign", y = "Total number of conersions")

ggplot(data\_id1178, aes(Spent, totConVal)) + geom\_point() + geom\_smooth(method = "lm") +

labs(x = "Amount spent on campaign", y = "Total value of conversions")

# Box Plot.

options(repr.plot.height = 3, repr.plot.width = 4)

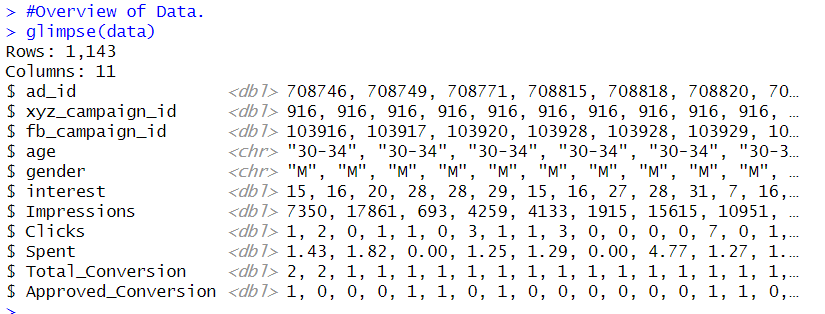
ggplot(data\_id1178, aes(gender, ROAS)) + geom\_boxplot() + scale\_y\_log10()

# Wilcox formula.

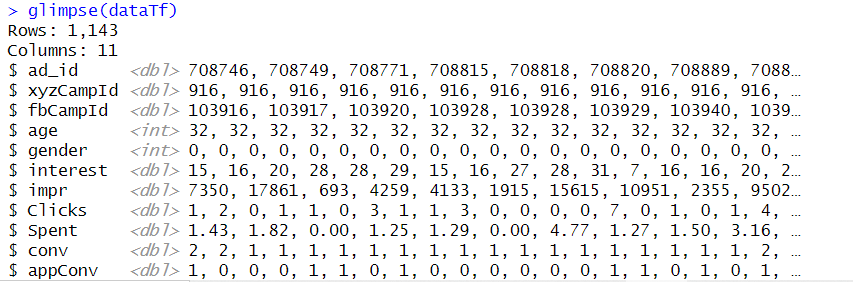
wilcox.test(ROAS ~ gender, data = data\_id1178)

**Output:**

**Initial Dataset (Before preprocessing):**

****

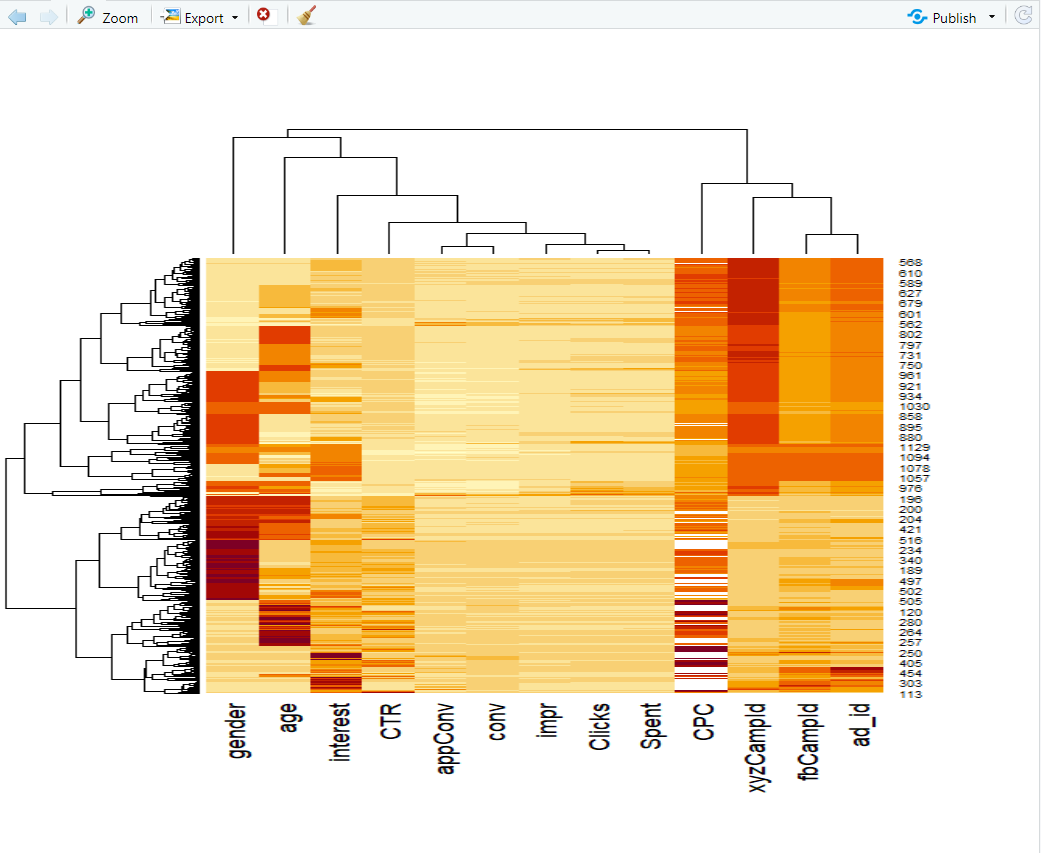
**Final Dataset (After Preprocessing):**

****

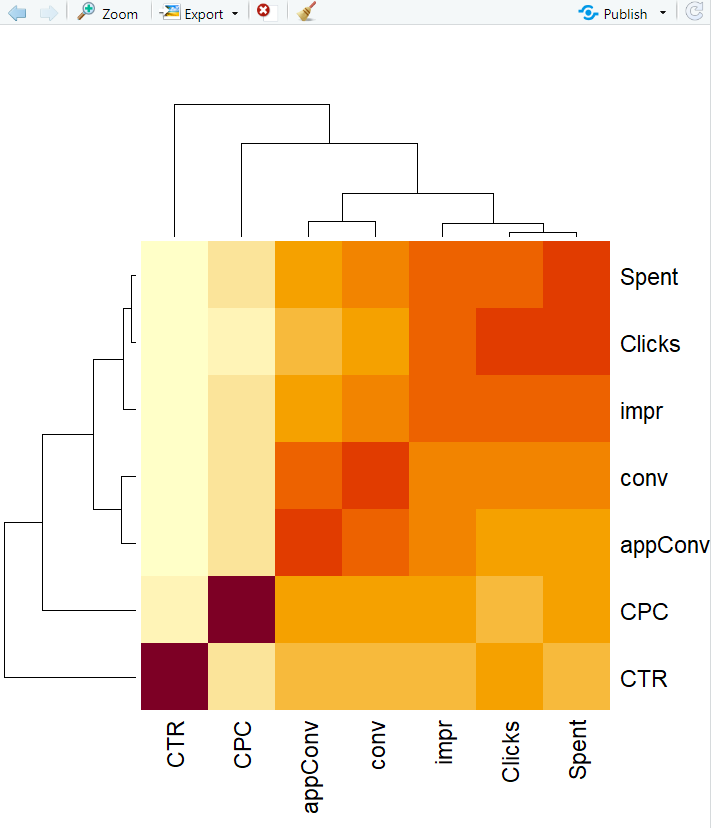
**[Note: This includes key changes in variable properties, filtered values and storing numeric data structures]**

**EDA and Visualisation:**

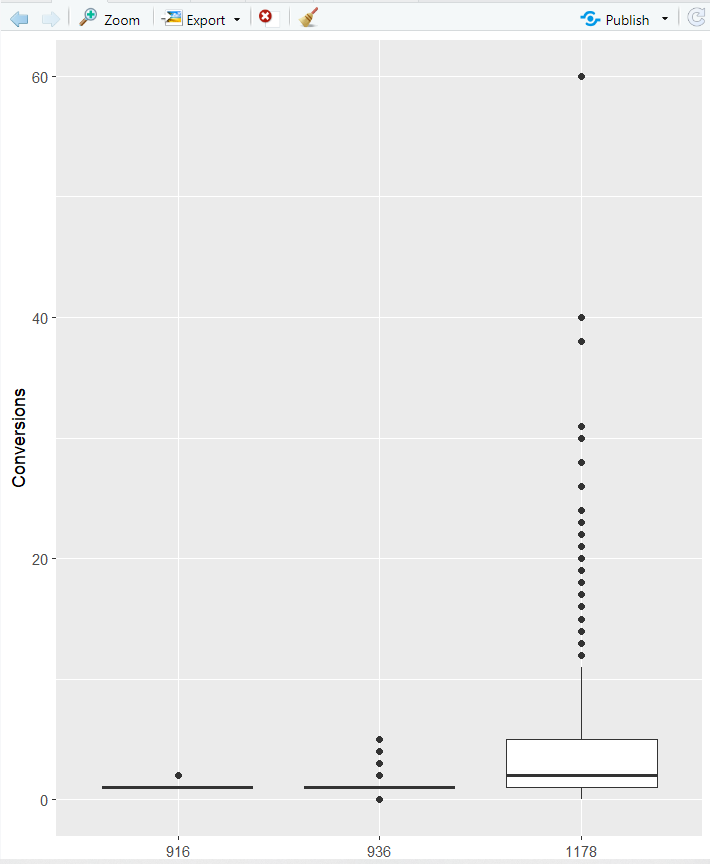
**Heatmap 1:**

****

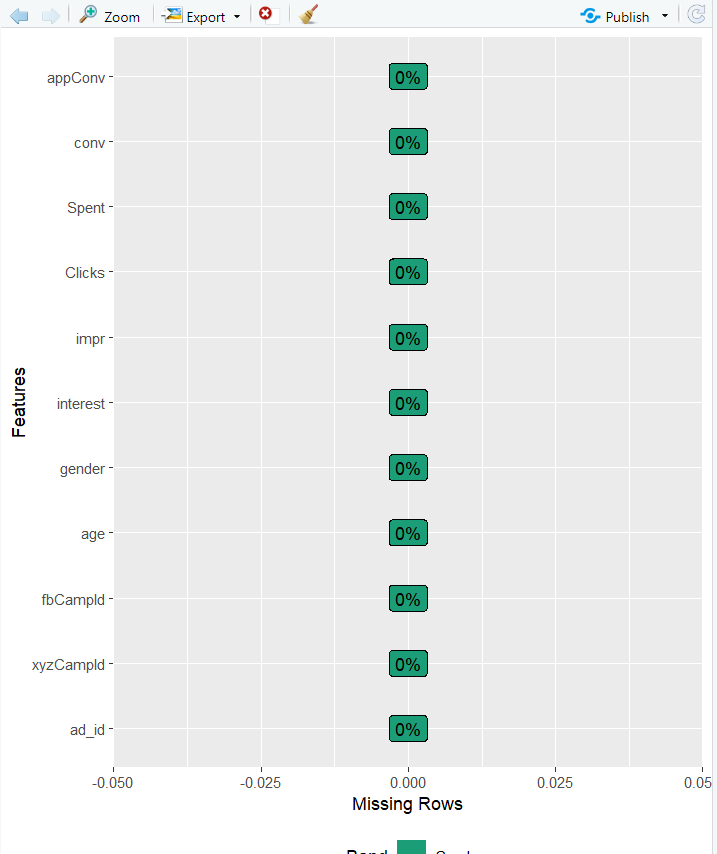
**Heatmap 2:**

****

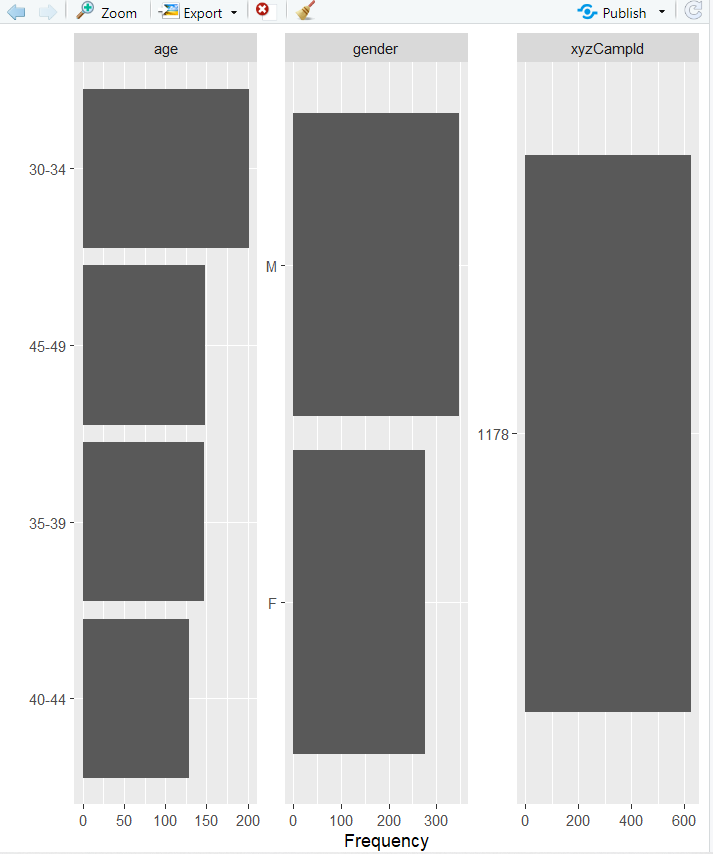
**Boxplot:**

****

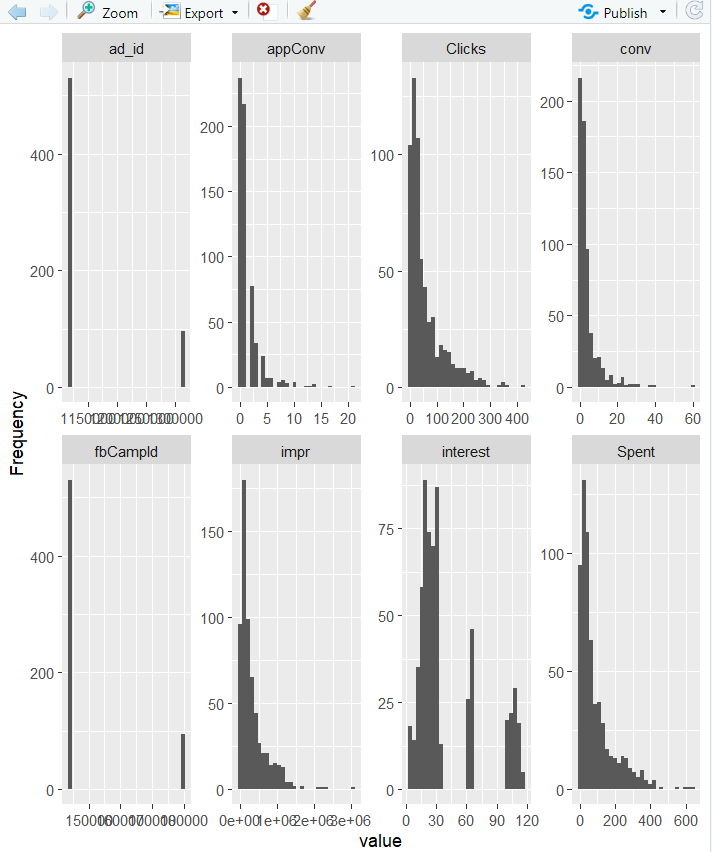
**Missing Plot:**

****

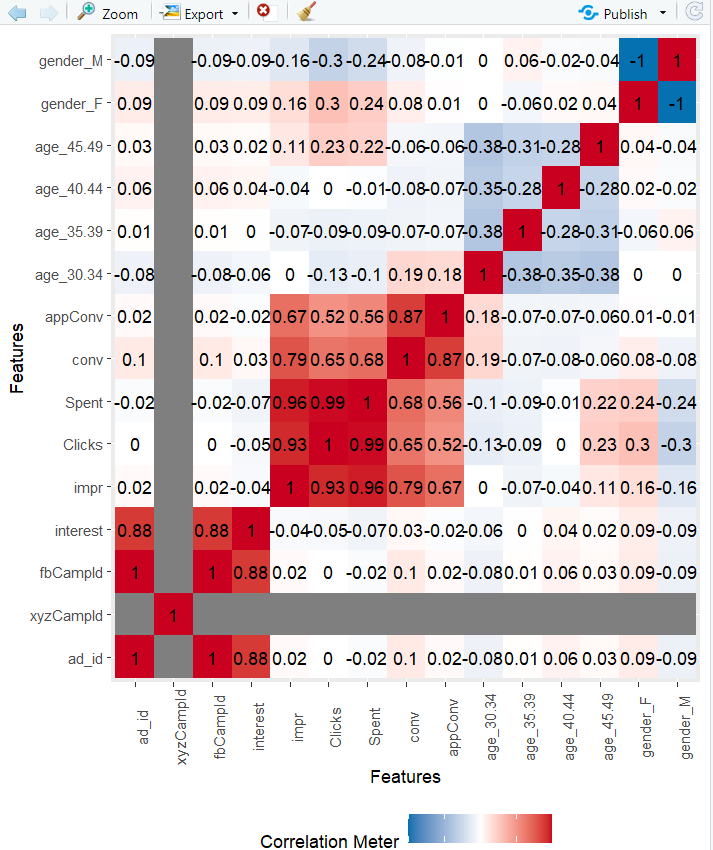
**Bar Plot:**

****

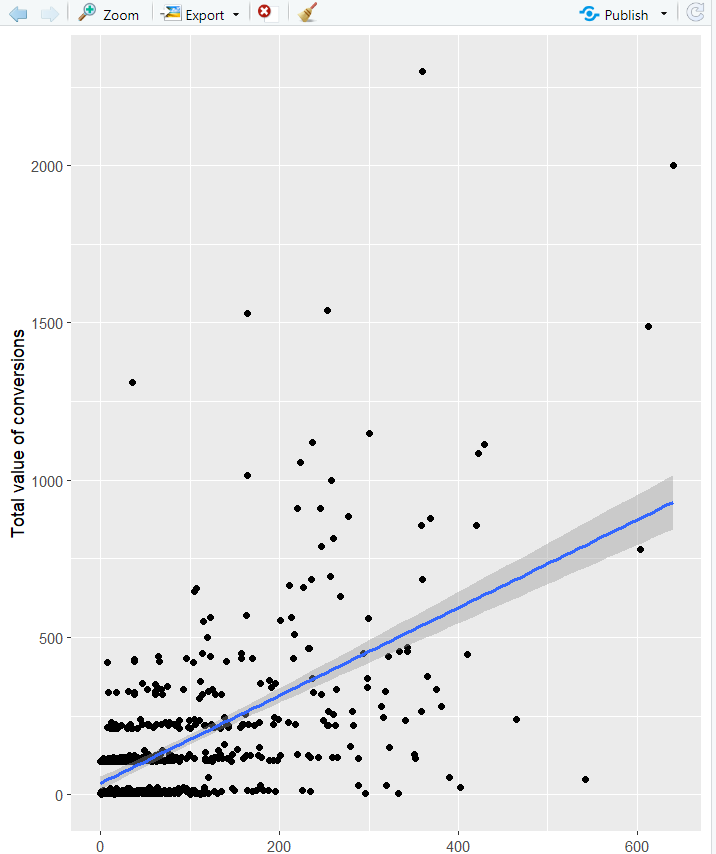
**Histogram:**

****

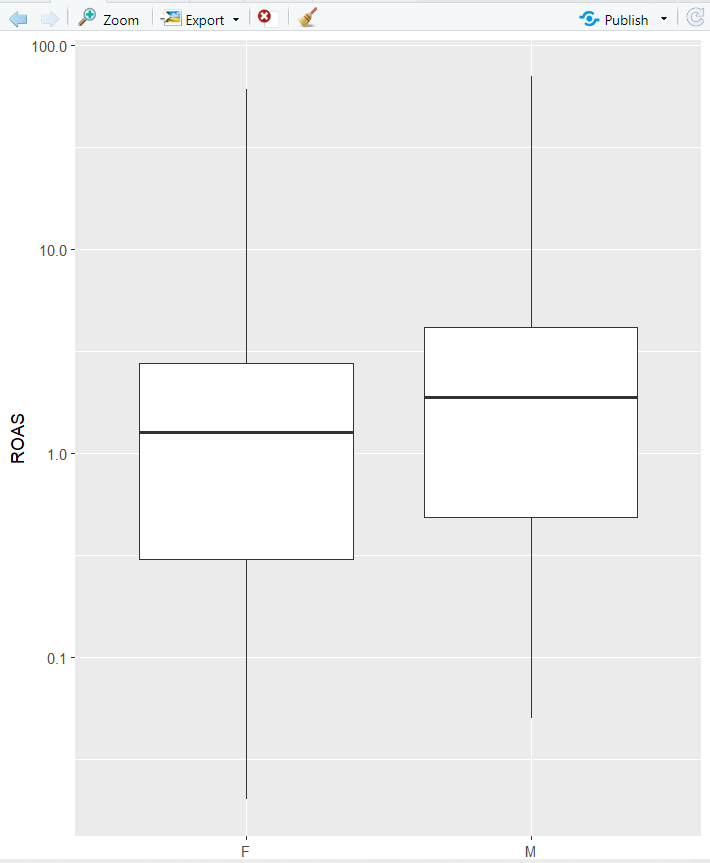
**Correlation Map:**

****

**Dense plot:**

****

**Box plot 2:**

****

**Wilcox Formula:**

