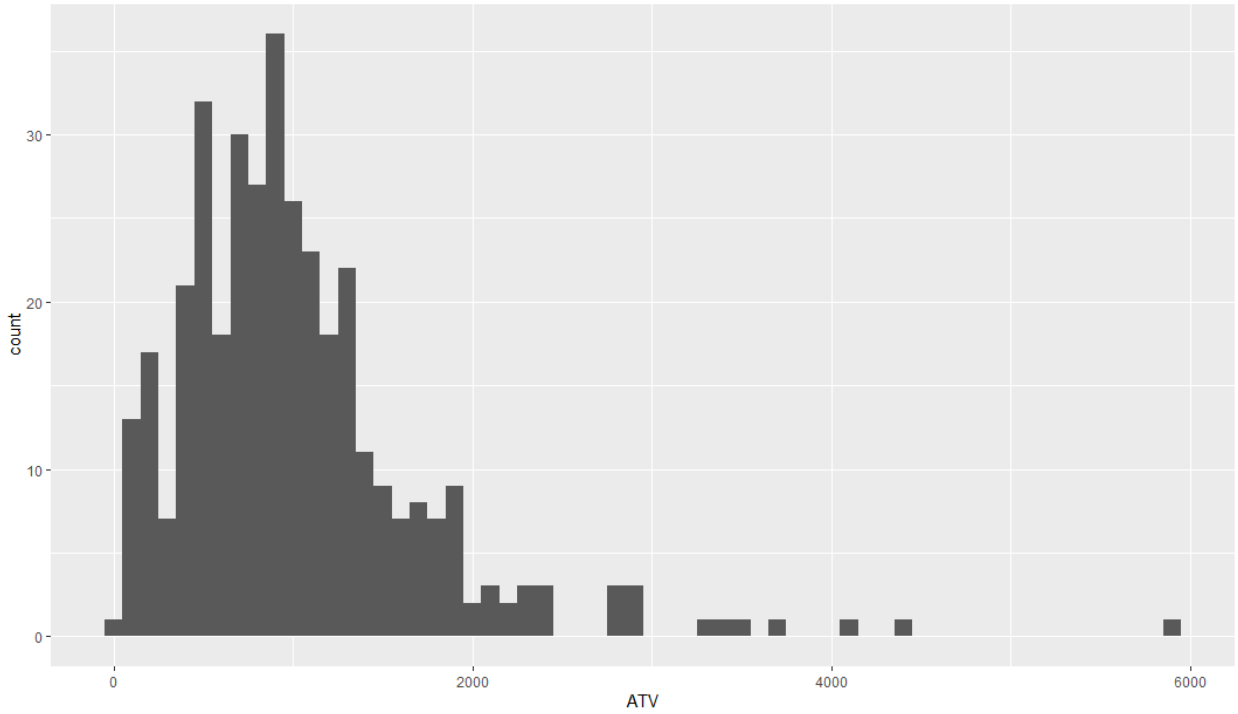
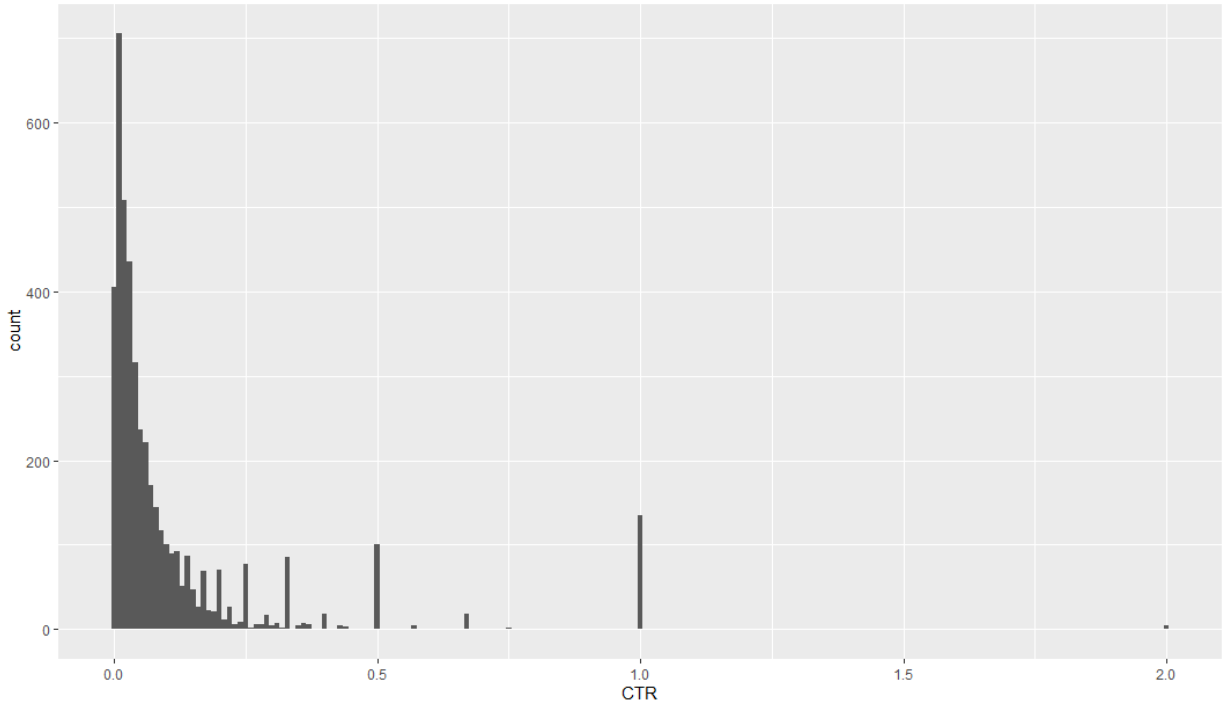
```
#####################################################################
####################
#
# Created by Team 12 on 04/03/2023
#
# Title        : Air France Codes
# Project             : Visualizing & Analyzing Data with R: Methods & Tools
# Institution   : Hult International Business School
# Author            : Team 12
# File Name     : Team_12_AirFrance_final.R
# Version       : 0.3
# Last Modified : 04/09/2023
# Description   :
#         Air France Case Data Analytics.
#
# History       :
#     Version      Date               Editor                        Detail
#     ------------------------------------------------------------------------ #
#     0.1                    04/03/2023  Team 12               First Commit
#     0.2       04/08/2023  Team 12        Adding Charts.
#     0.3       04/09/2023  Team 12        Adding Coments.
#####################################################################
####################

# Set work directory.
setwd('C:/Users/yunsi/OneDrive/Hult International Business School/01 Course/13 Visualizing &
Analyzing Data with R')

# Install Packages

# importing all required libraries
library(readxl)
library(ggplot2)
library(reshape2)

#####################################################################
#######
# User Defined Function
#####################################################################
#######

#####################################################################
#######
# UDF: calculateKPI
# arg: data (DataFrame of AirFrance)
# Desciption: Calculate KPI
#   CTR(Click Through Rate)       : Clicks / Impression
#                             Keyword's performance efficiency.
#   PCR(Purcahsed Conversion Rate) : TotalVolumeofBookings / Clicks
#                             Relatedness between Customer needs and products.
#   CPC(Cost per Click)           : TotalCost / Clicks
#                             Keywords Competition.
#   CPP(Cpst Per Purchase)        : TotalCost / TotalVolumeofBookings
#                             Cost Efficiency.
```

```
#   ATV(Average Transaction Value) : Amount / TotalVolumeofBookings
#                                Product Marketing Results.
#   NET(Net Income)                 : ATV - CPP
#                                Actual Income by those campaign
##################################################################
#######
# Function: cacluateKPI
calculateKPI <- function(data){
  data$CTR <- data$Clicks / data$Impressions
  data$PCR <- data$TotalVolumeofBookings / data$Clicks
  data$CPC <- data$TotalCost / data$Clicks
  data$CPP <- data$TotalCost / data$TotalVolumeofBookings
  data$ATV <- data$Amount / data$TotalVolumeofBookings
  data$NET <- data$ATV - data$CPP
  return(data)
} # Closing


# importing the Excel data source
df <- read_excel("assignment/Air France Case Spreadsheet Supplement.xls", sheet = "DoubleClick")

# Col Name Adjust
colnames(df) <- gsub("[ /.%]", "", colnames(df))

# KPI Build
df <- calculateKPI(df)

# Check NA Values
df_bid_na <- df[which(is.na(df$BidStrategy) == TRUE), ]

# Number of NA Values in BidStrategy
nrow(df_bid_na)
# Bidstrategy has 1,224 NA values

# Fill NA of BidStratgy in "NO Strategy"
for (i in 1:nrow(df)){
  if (is.na(df[i, "BidStrategy"] )){
    df[i, "BidStrategy"] <- "NO Strategy"
  }
} # Closing Loop

# Cleansing BidStrategy Typo Error
for (i in 1:nrow(df)){
  if (df[i, "BidStrategy"] == "Position 1 -2 Target"){
    df[i, "BidStrategy"] <- "Position 1-2 Target"
  } else if (df[i, "BidStrategy"] == "Postiion 1-4 Bid Strategy"){
    df[i, "BidStrategy"] <- "Position 1-4 Bid Strategy"
  }
} # Closing Loop

# Check Cleansing Result
table(df$BidStrategy)

# Check Clicks equals to 0 - 1 record has 0 click
```

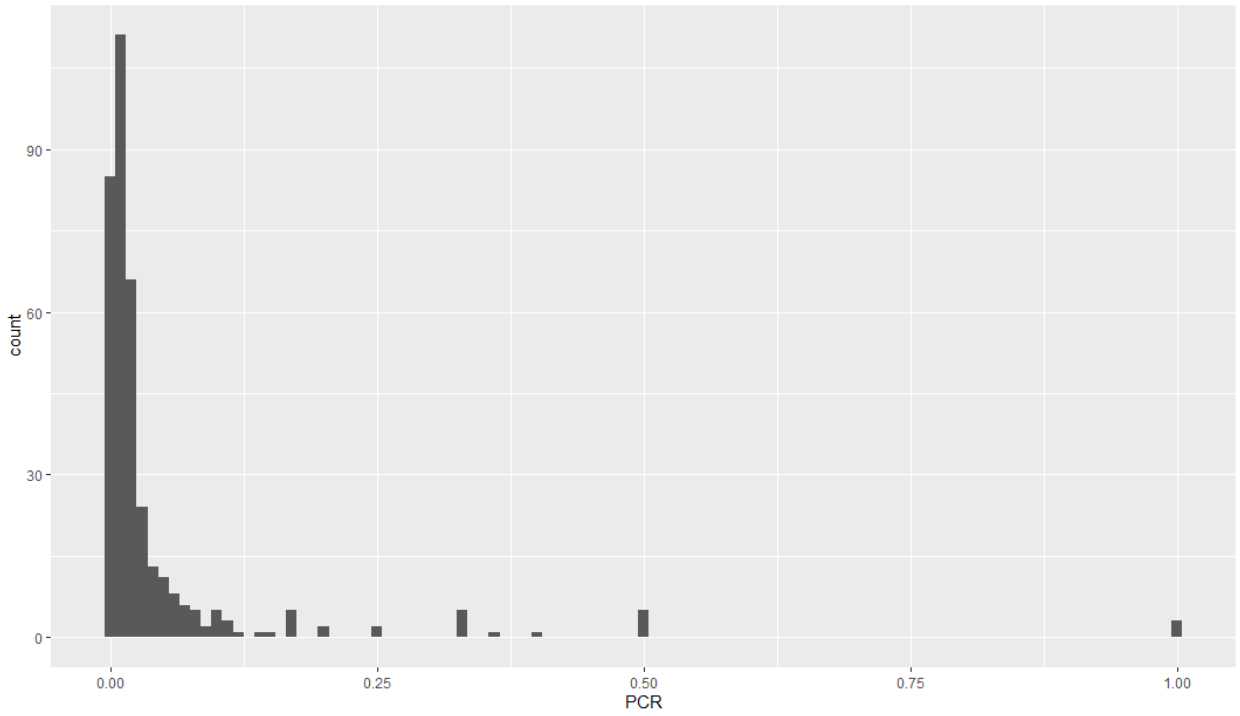
```
clicks_zero <- which(df$Clicks == 0)

# If click is zero, KPI would be NA or INF.
# Convert those values to 0
df[clicks_zero, "CTR"] <- 0
df[clicks_zero, "PCR"] <- 0
df[clicks_zero, "CPC"] <- 0
df[clicks_zero, "CPP"] <- 0

# Check TotalVolumeofBookings equals to 0 and Amount equals 0
no_result <- which(df$TotalVolumeofBookings == 0 & df$Amount == 0)

df[no_result, "ATV"] <- 0
df[no_result, "NET"] <- 0 - df[no_result, "TotalCost"]

# ############################################
# Descriptive Statistics
# ############################################
# Check Numeric Data Summary
summary(df)

# Histogram of ATV, excluding zero ATV. Chart01
ggplot(df[which(df$ATV > 0), ], aes(ATV)) +
  geom_histogram(binwidth = 100)
```



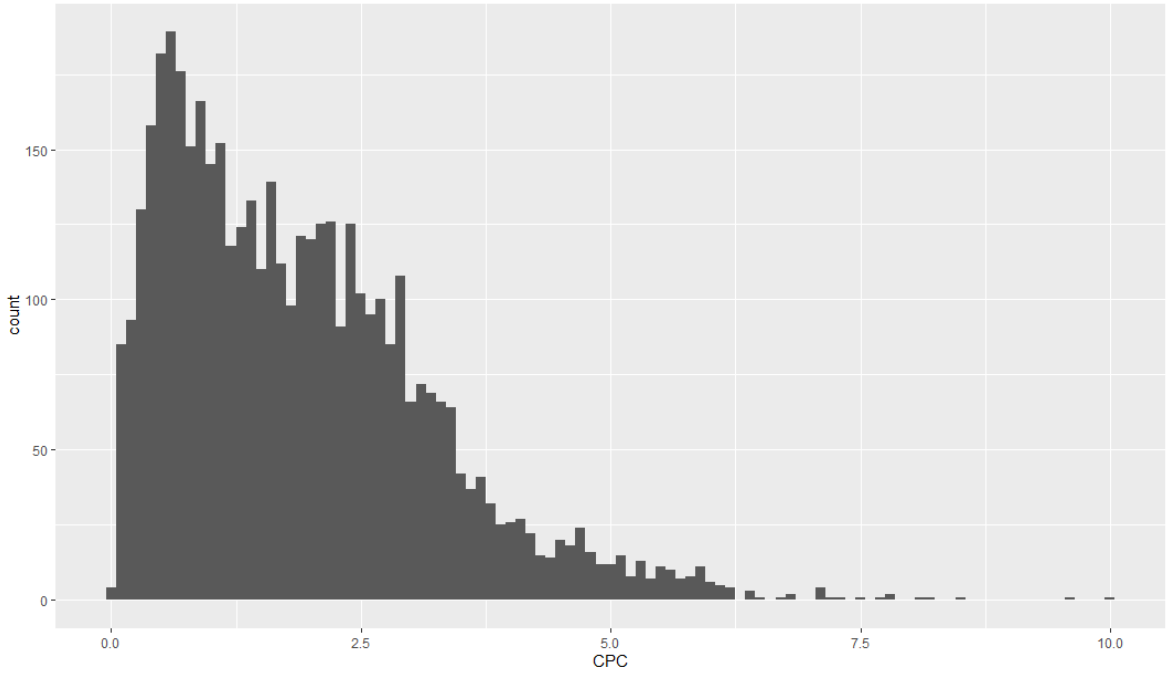
```
# Histogram of CTR Chart_02
ggplot(df, aes(CTR)) +
  geom_histogram(binwidth = 0.01)
```



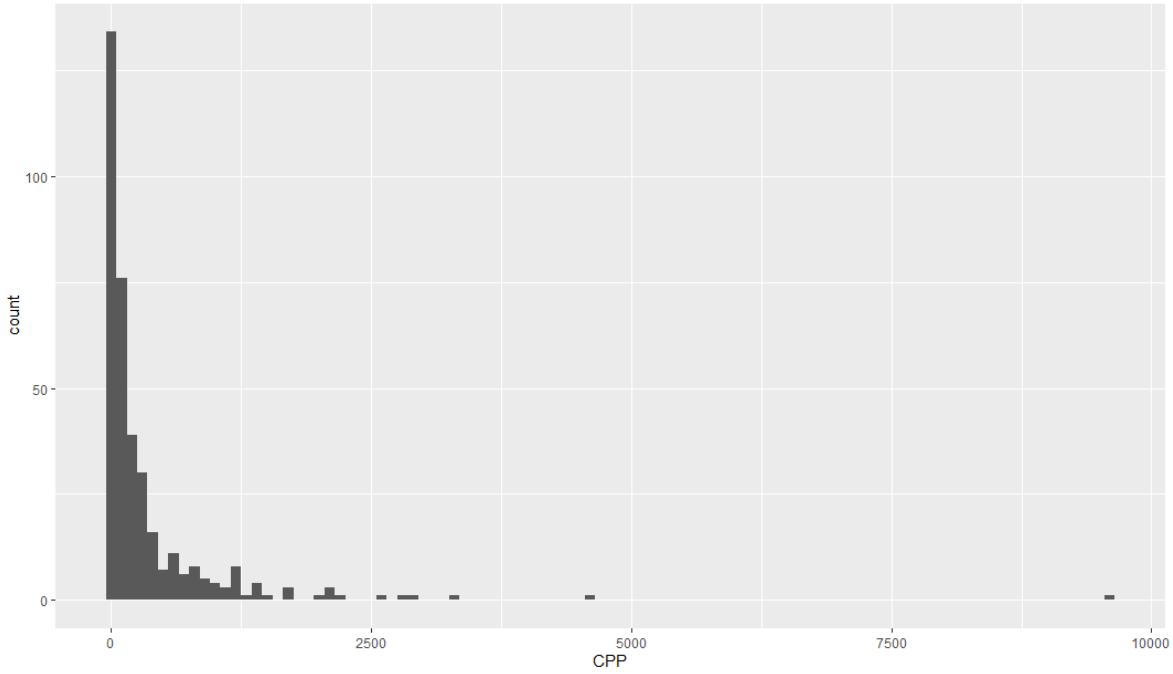

```
# Histogram of PCR, excluding zero PCR and 900% PCR. (Outlier) Chart_03
ggplot(df[which(df$PCR > 0 & df$PCR <= 1), ], aes(PCR)) +
  geom_histogram(binwidth = 0.01)
```



```
# Histogram of CPC Chart_04
ggplot(df, aes(CPC)) +
  geom_histogram(binwidth = 0.1)
```

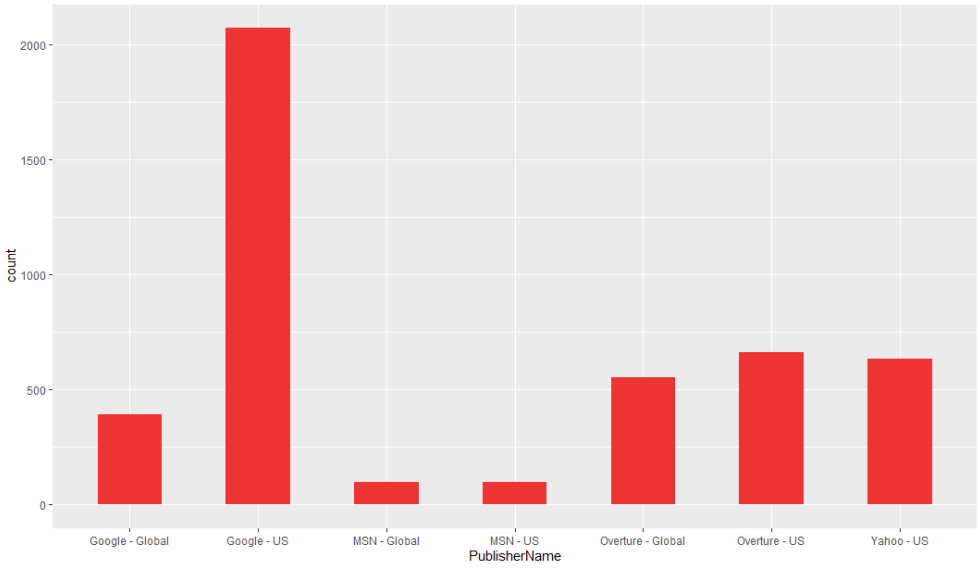



```
# Histogram of CPP, excluding zero CPP.Chart_05
ggplot(df[which(df$CPP > 0), ], aes(CPP)) +
  geom_histogram(binwidth = 100)
```

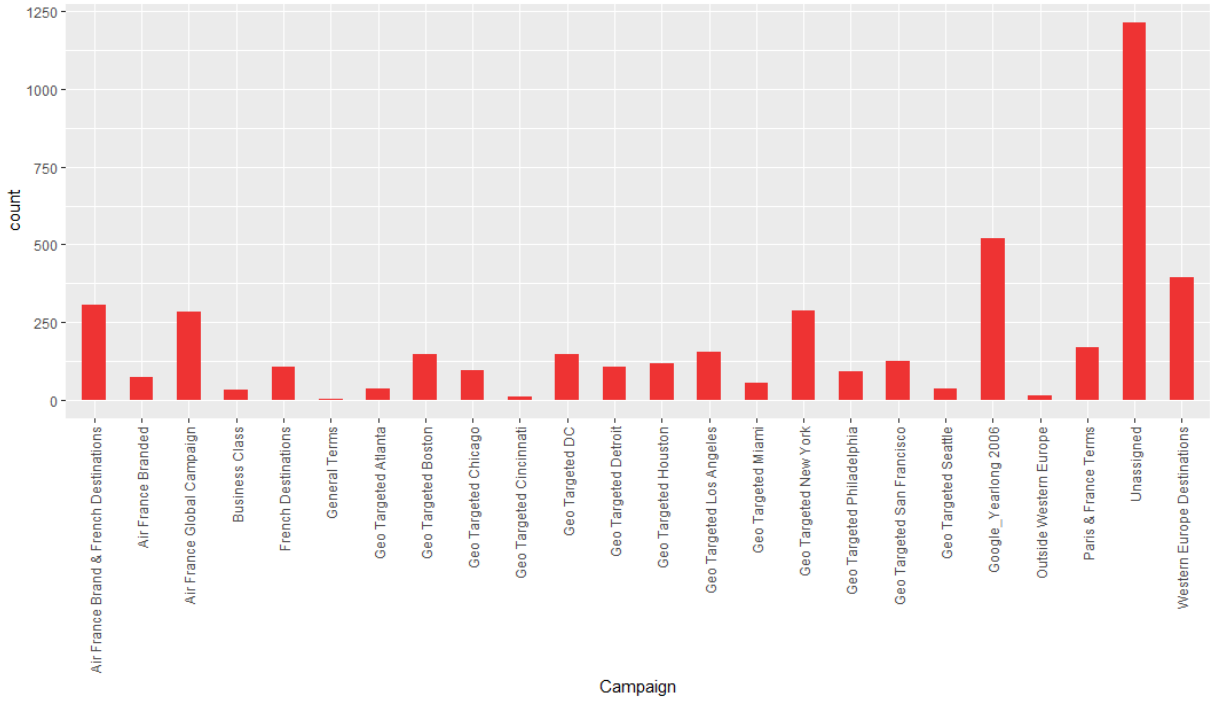



```
# Check Categorical Data
```

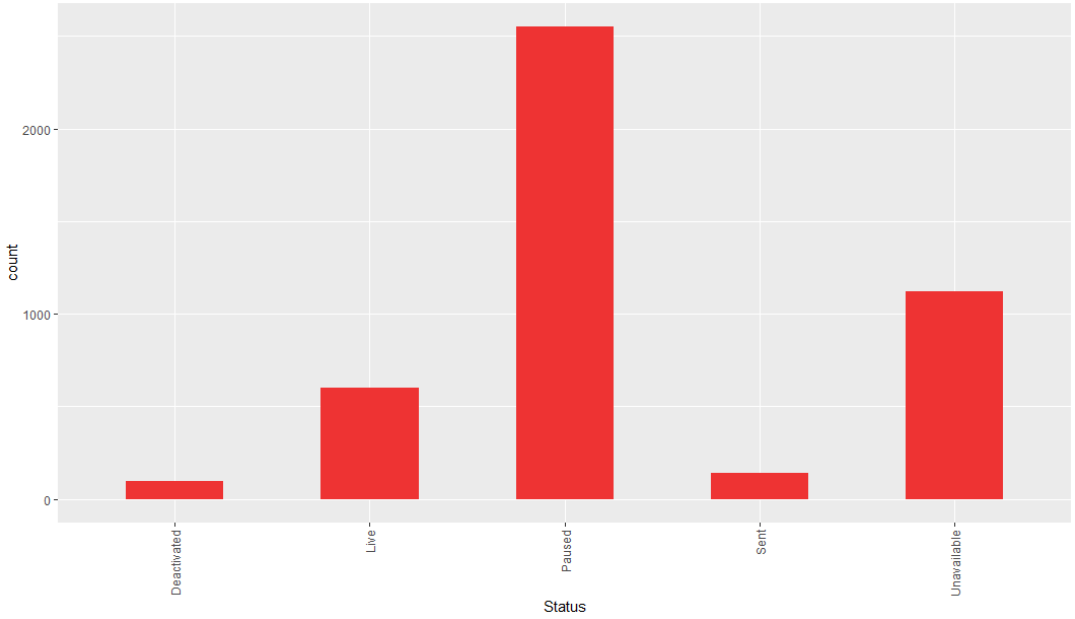
```
# Publisher Name
table(df$PublisherName)
# Chart of Number of keywords by Publisher Name Chart_06.
ggplot(df, aes(PublisherName)) +
  geom_bar(fill="#ee3333", width = 0.5)
```



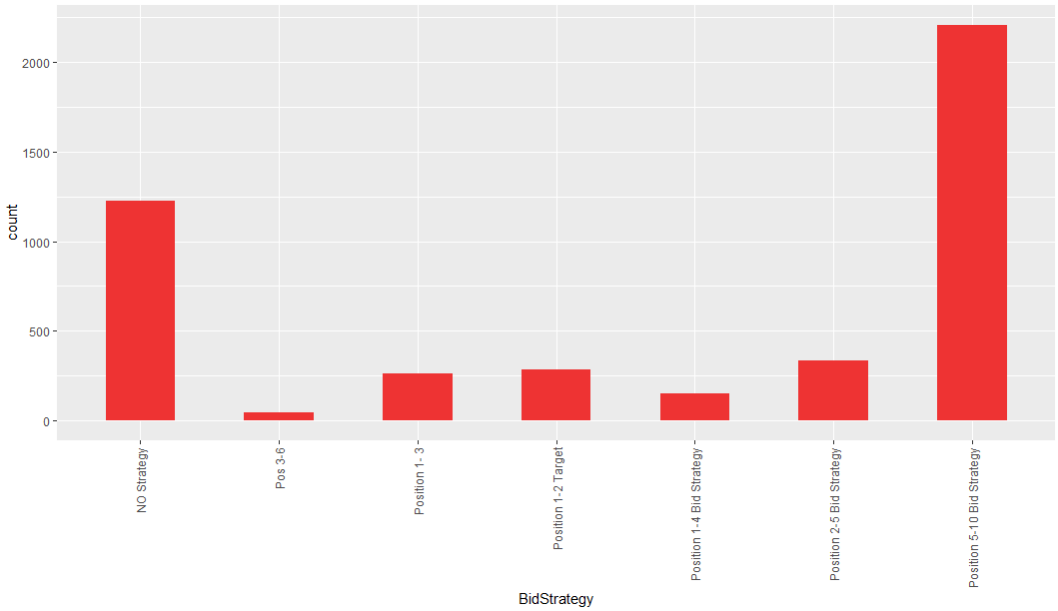

```
# Campaign
table(df$Campaign)
# Chart of Number of keywords by Campaign Chart_07
ggplot(df, aes(Campaign)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  geom_bar(fill="#ee3333", width = 0.5)
```



```
# Status
table(df$Status)
# Number of keywords by Status. Chart_08
ggplot(df, aes(Status)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  geom_bar(fill="#ee3333", width = 0.5)
```




```
# BidStrategy
table(df$BidStrategy)
# Number of keywords by Bid Strategy. Chart_09
ggplot(df, aes(BidStrategy)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  geom_bar(fill="#ee3333", width = 0.5)
```



```
# ##########################################
# KPI Analysis
# ##########################################
# Aggregate Numeric Data By Group
# By Publisher Name
tbl.pub <- aggregate(cbind(Impressions, Clicks, Amount, TotalCost, TotalVolumeofBookings) ~
PublisherName, data = df, FUN = "sum", na.rm=TRUE)

# Adding Kayak Data
tbl.pub <- rbind(tbl.pub, c(0, 0, 2939, 233694, 3567.13, 208))
tbl.pub[which(tbl.pub$PublisherName == "0"), ]$PublisherName <- "Kayak"

# KPI Calculation
tbl.pub <- calculateKPI(tbl.pub)

# Calculate Cost Proportion
tbl.pub$CostProp <- tbl.pub$TotalCost / tbl.pub$Amount

# Kayak data converting, because it does not have Impression information.
tbl.pub[which(tbl.pub$PublisherName == "Kayak"),]$CTR <- 0 # No impression information

# Table. 1. Key Business Values by Channel
tbl.pub[, c(1, 3, 6, 5, 4, 12)]

# Key Business Values by Channel
#   1. Google has the biggest volume of Revenue and Cost.
#   2. Overture's Revenue is lower and the highest cost proportion.

# Chart for Clicks by Publisher. Chart_10
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=Clicks), stat="identity", fill="#ee3333") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=Clicks, label = paste(round(Clicks, 0), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(0, max(tbl.pub$Clicks) + 5000) +
  ylab("Clicks") +
  xlab("Publishers") +
  ggtitle("Clicks by Publishers")

# Chart for Purchase by Publisher. Chart_11
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=TotalVolumeofBookings), stat="identity", fill="#ee3333") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=TotalVolumeofBookings, label = paste(round(TotalVolumeofBookings, 0), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(0, max(tbl.pub$TotalVolumeofBookings) + 100) +
  ylab("Total Volume of Bookings (Purchase)") +
  xlab("Publishers") +
```

```
  ggtitle("Purchased by Publishers")

# Chart for TotalCost by Publisher. Chart_12
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=TotalCost), stat="identity", fill="#ee3333") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=TotalCost, label = paste(round(TotalCost, 0), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(0, max(tbl.pub$TotalCost) + 10000) +
  ylab("Total Cost ($)") +
  xlab("Publishers") +
  ggtitle("Total Cost by Publishers")

# Chart for Amount by Publisher. Chart_13
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=Amount), stat="identity", fill="#ee3333") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=Amount, label = paste(round(Amount, 0), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(0, max(tbl.pub$Amount) + 50000) +
  ylab("Total Amount ($)") +
  xlab("Publishers") +
  ggtitle("Total Amount by Publishers")


# Chart for NET by Publisher. Chart_14
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=NET), stat="identity", fill="#ee3333") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=NET, label = paste(round(NET, 0), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(0, max(tbl.pub$NET) + 100) +
  ylab("Total Net Income ($)") +
  xlab("Publishers") +
  ggtitle("Total Net Income by Publishers")
```

Chart_10

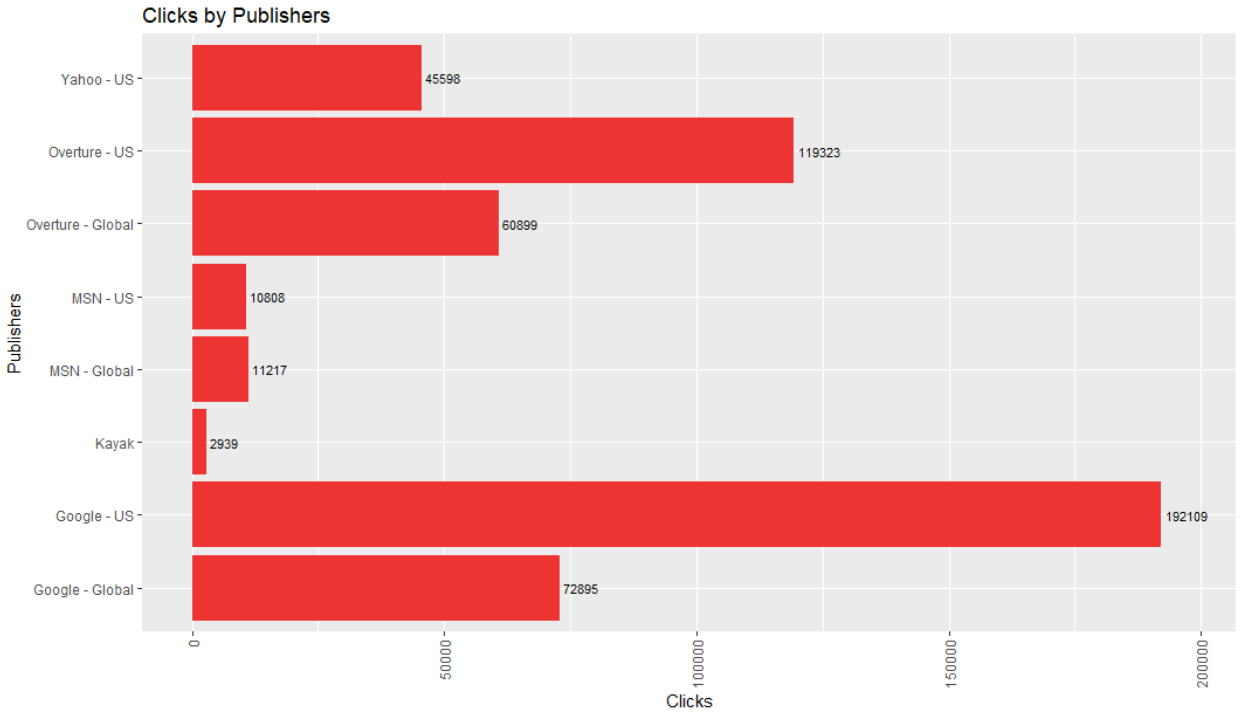


Chart_11

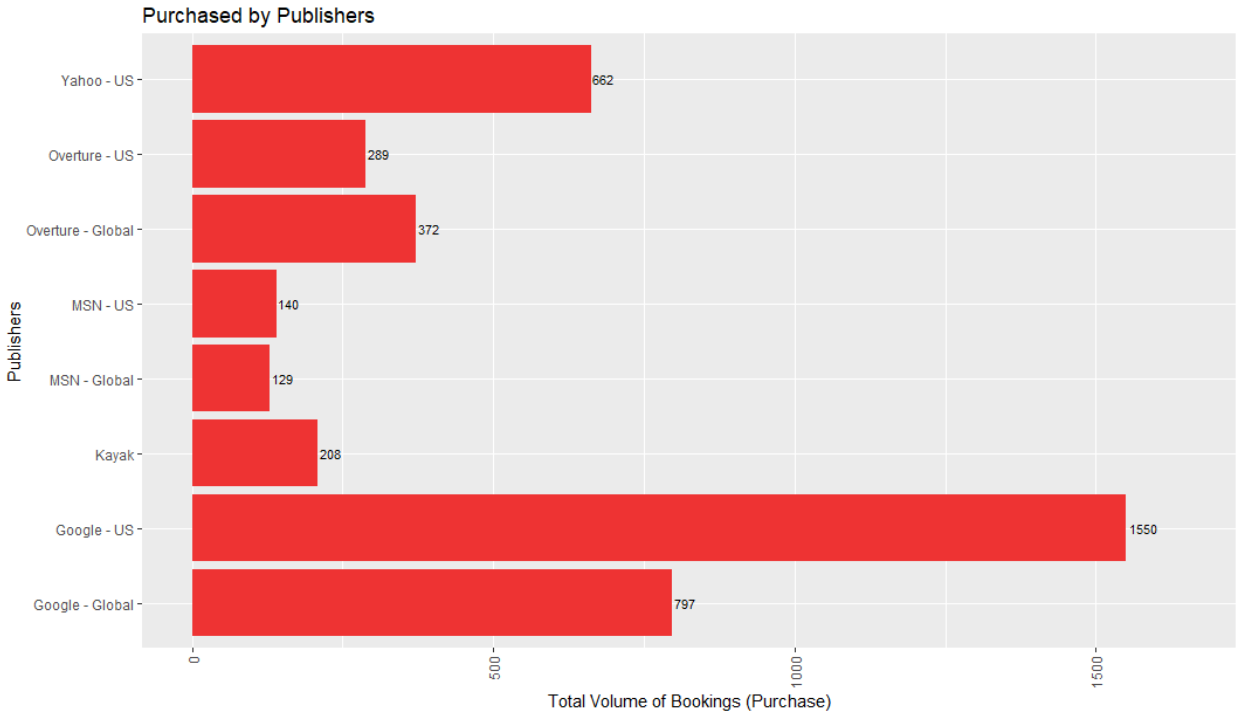


Chart_12

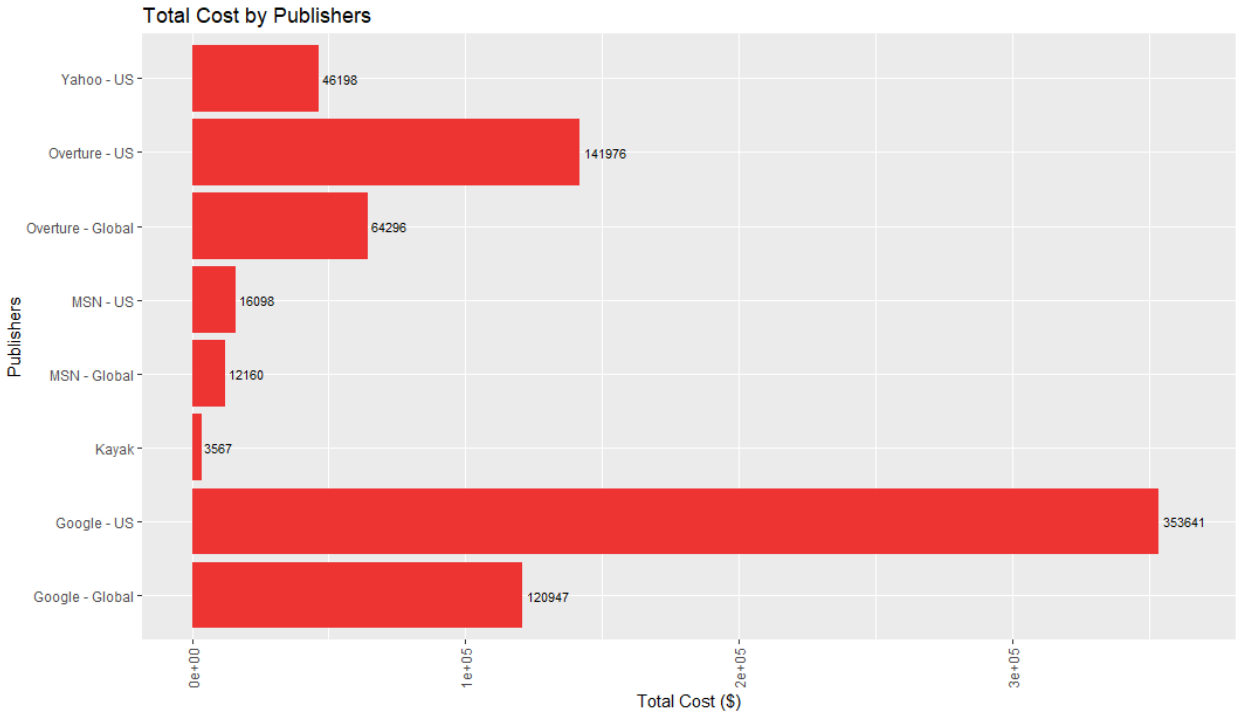


Chart_13

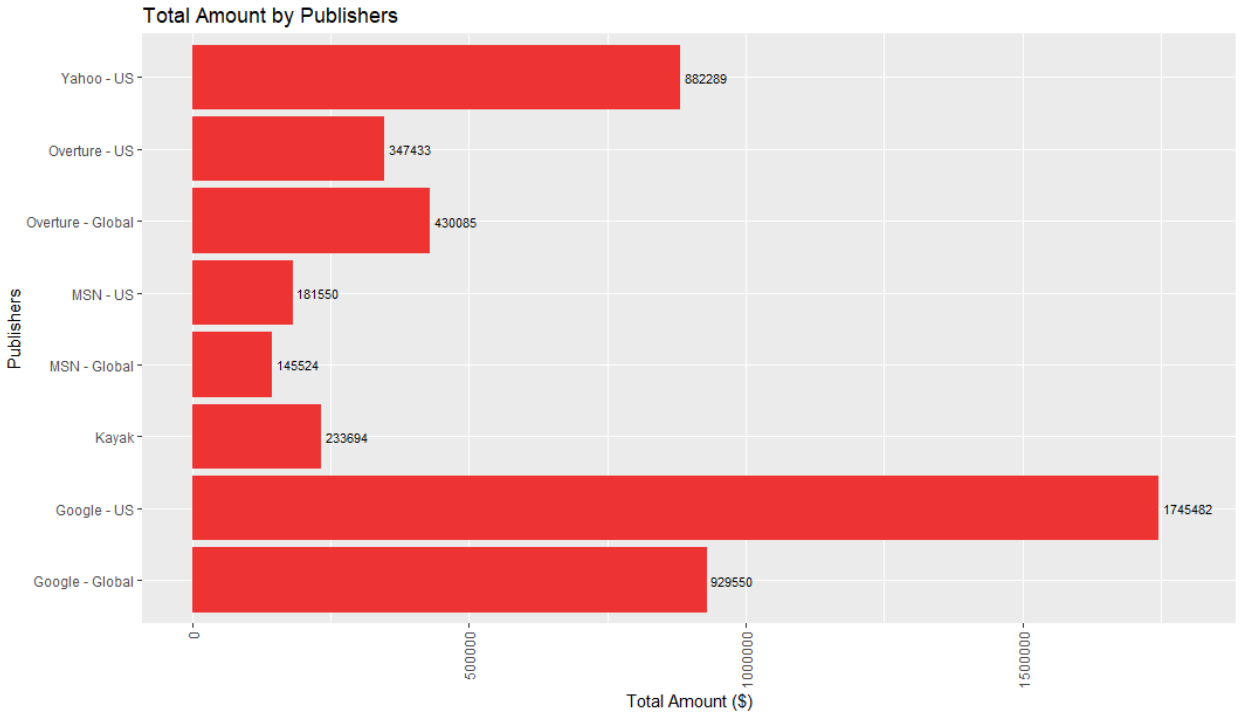
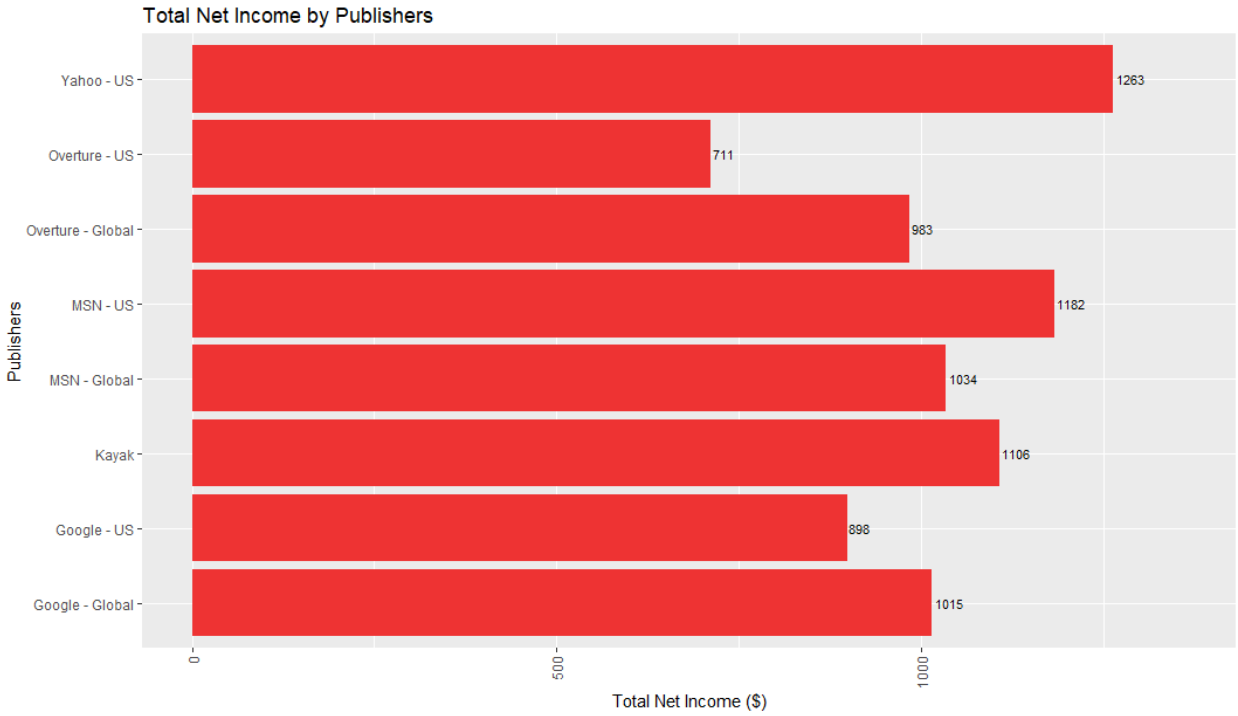
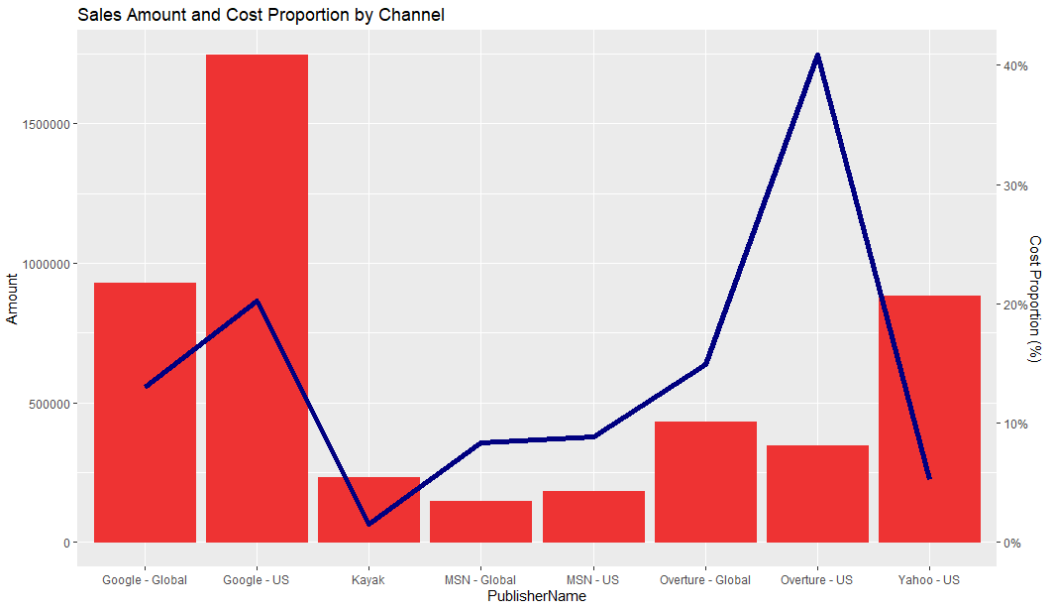


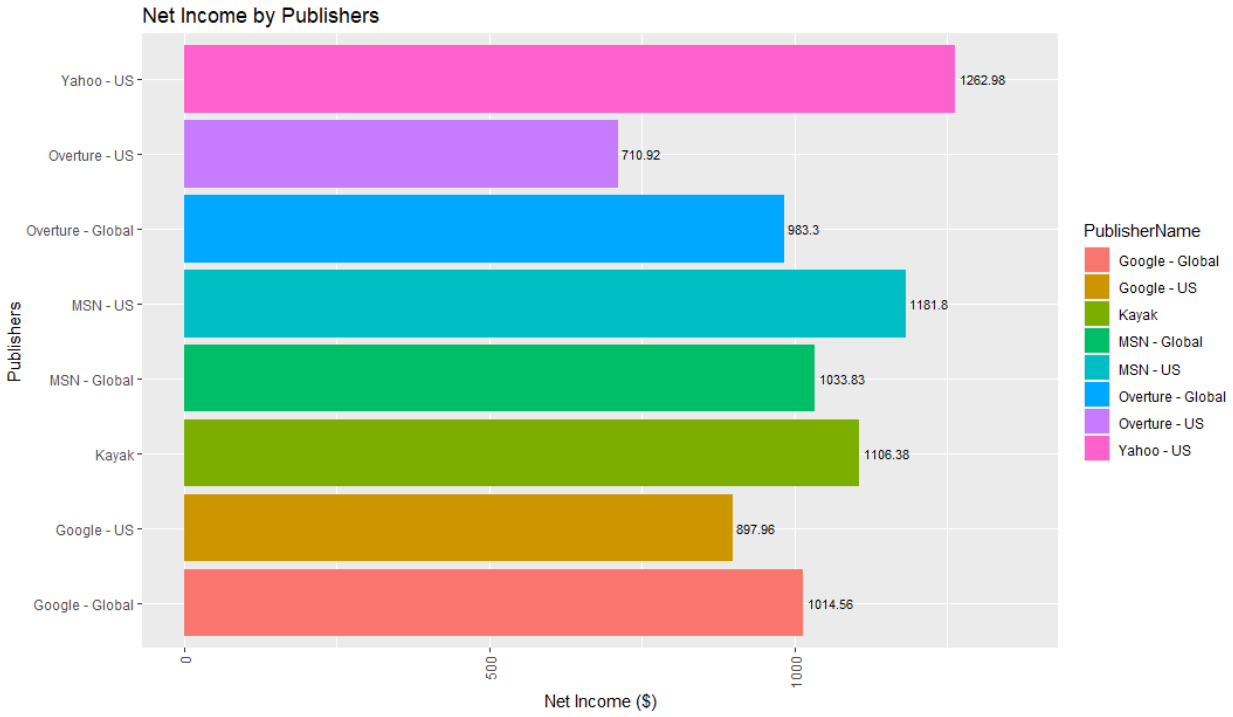
Chart_14




```
# Chart 15. Sales Amount and Cost Proportion by Channel
scaleFactor <- max(tbl.pub$Amount) / max(tbl.pub$CostProp)
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=Amount), stat="identity", color="#ee3333", fill="#ee3333") +
  geom_line(aes(y=CostProp * scaleFactor), group=1, color="navy", lwd=2) +
  scale_y_continuous(
    sec.axis = sec_axis(trans = ~./scaleFactor, label=scales::percent, name = "Cost Proportion (%)")
  ) +
  ggtitle("Sales Amount and Cost Proportion by Channel")
```

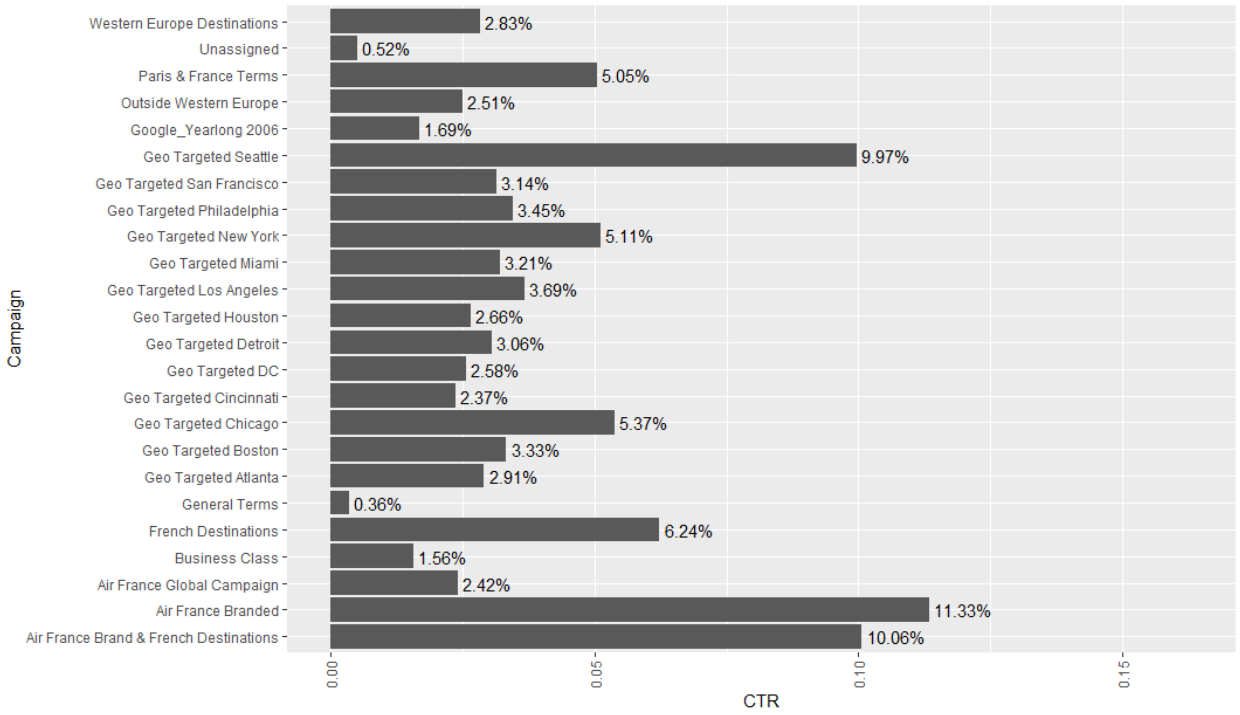


```
# Chart for NET Income by Publisher. Chart_16
ggplot(tbl.pub, aes(x=PublisherName)) +
  geom_bar(aes(y=NET, fill=PublisherName), stat="identity") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=NET, label = paste(round(NET, 2), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(0, max(tbl.pub$NET) + 100) +
  ylab("Net Income ($)") +
  xlab("Publishers") +
  ggtitle("Net Income by Publishers")
```

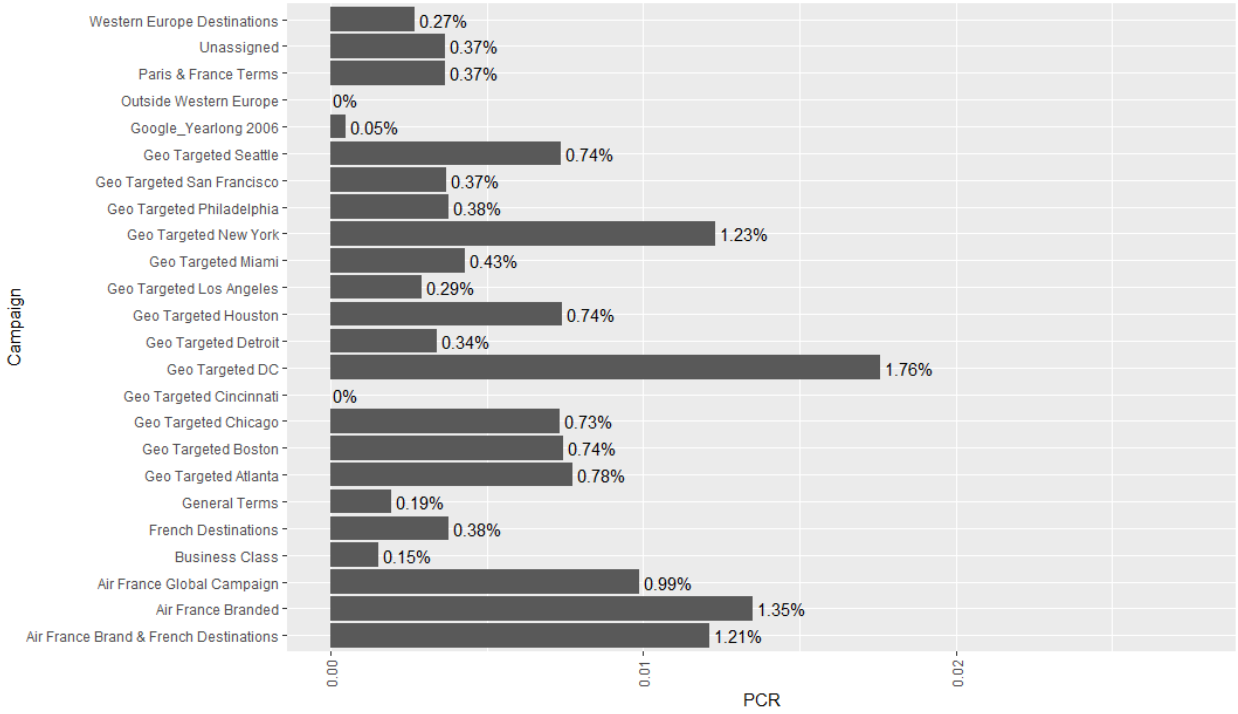



```
# Yahoo makes the highest Net Income with SEM.
# Aggregator's performances are good enough with 7.8% of PCR.
# Google performed well but had lower Net Income because its expenses.

# KPI Analysis #2.
# By Campaign
tbl.camp <- aggregate(cbind(Impressions, Clicks, Amount, TotalCost, TotalVolumeofBookings) ~
Campaign, data = df, FUN = "sum", na.rm=TRUE)
tbl.camp <- calculateKPI(tbl.camp)

# Table. 2. KPI analysis by Campaign
tbl.camp[, c("Campaign", "CTR", "PCR", "CPC", "CPP", "ATV", "NET")]

# Fill 0 as NA and Inf
for (i in 1:ncol(tbl.camp)){
  tbl.camp[which(is.na(tbl.camp[, i])), i] <- 0
```

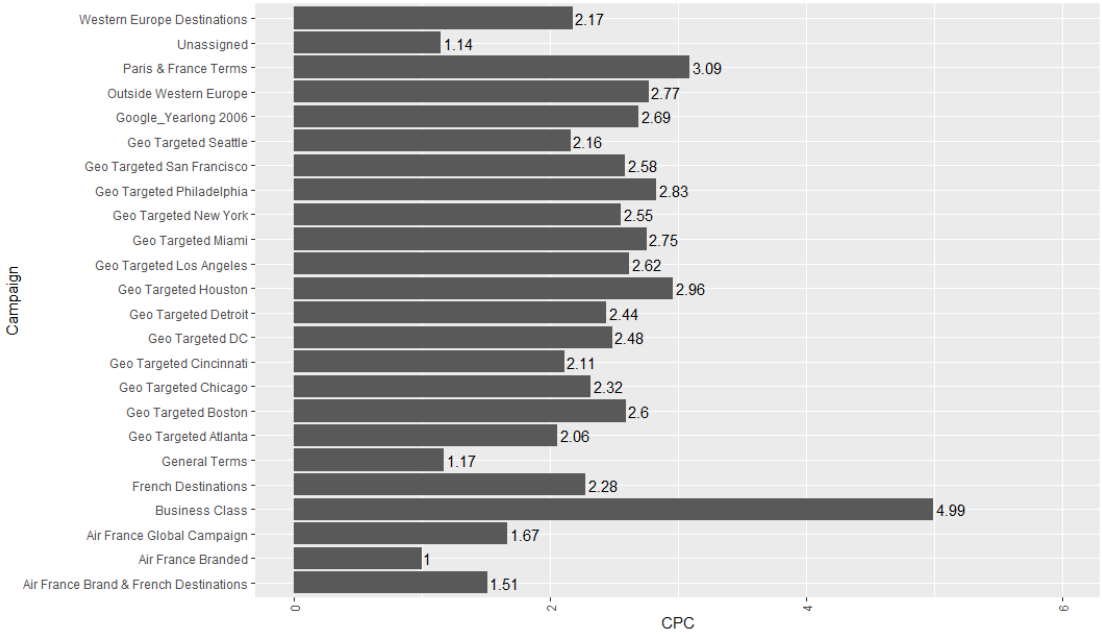
```
  tbl.camp[which(is.infinite(tbl.camp[, i])), i] <- 0
}

# Chart_17. CTR bar plot to compare by Campaign
ggplot(tbl.camp, aes(x=Campaign)) +
  geom_bar(aes(y=CTR), stat="identity", position=position_dodge(0.5)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=CTR, label = paste(round(CTR*100, 2), "%", sep="")),
            hjust=-0.1
            ) +
  scale_y_continuous(labels = scales::percent) +
  ylim(0, max(tbl.camp$CTR) + 0.05)
```



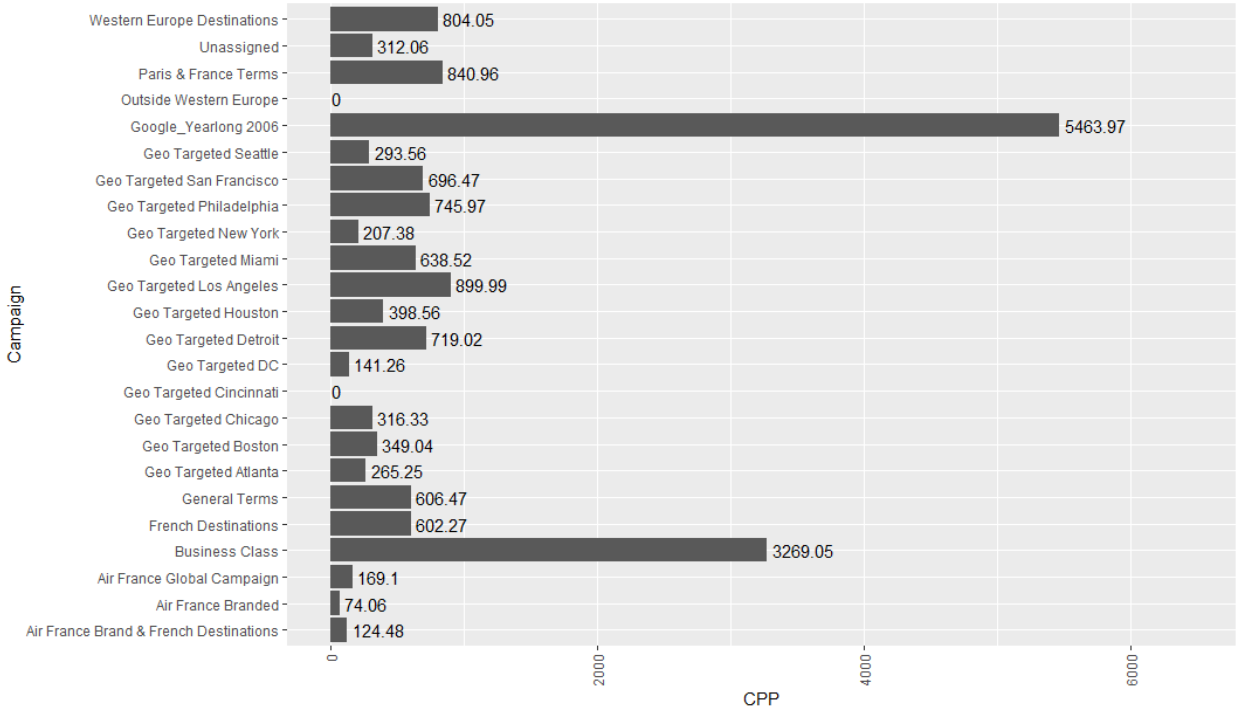

```
# Chart_18 for PCR
ggplot(tbl.camp, aes(x=Campaign)) +
  geom_bar(aes(y=PCR), stat="identity", position=position_dodge(0.5)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=PCR, label = paste(round(PCR*100, 2), "%", sep="")),
            hjust=-0.1
  ) +
  scale_y_continuous(labels = scales::percent) +
  ylim(0, max(tbl.camp$PCR) + 0.01)
```



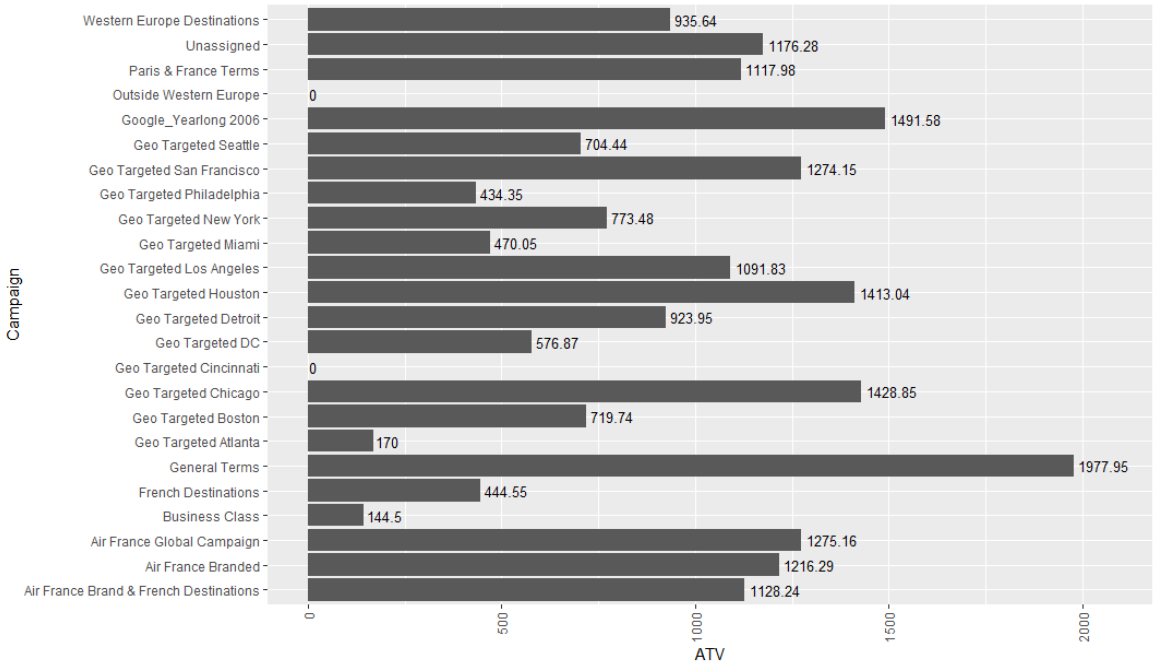

```
# Chart 19 for CPC
ggplot(tbl.camp, aes(x=Campaign)) +
  geom_bar(aes(y=CPC), stat="identity", position=position_dodge(0.5)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=CPC, label = paste(round(CPC, 2), sep="")),
            hjust=-0.1
  ) +
  ylim(0, max(tbl.camp$CPC) + 1)
```



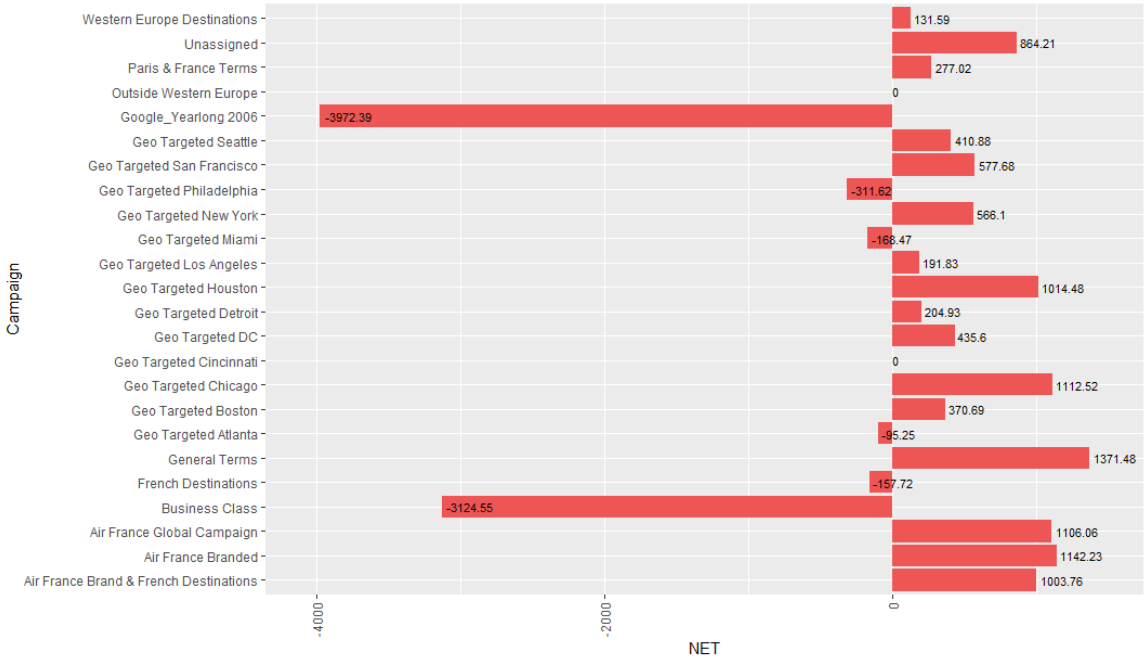
```
# Chart 20 for CPP
ggplot(tbl.camp, aes(x=Campaign)) +
  geom_bar(aes(y=CPP), stat="identity") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=CPP, label = paste(round(CPP, 2), sep="")),
            hjust=-0.1
  ) +
  ylim(0, max(tbl.camp$CPP) + 1000)
```




```
# Chart 21 for ATV
ggplot(tbl.camp, aes(x=Campaign)) +
  geom_bar(aes(y=ATV), stat="identity") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=ATV, label = paste(round(ATV, 2), sep="")),
            hjust=-0.1,
            size=3.5
  ) +
  ylim(0, max(tbl.camp$ATV) + 100)
```

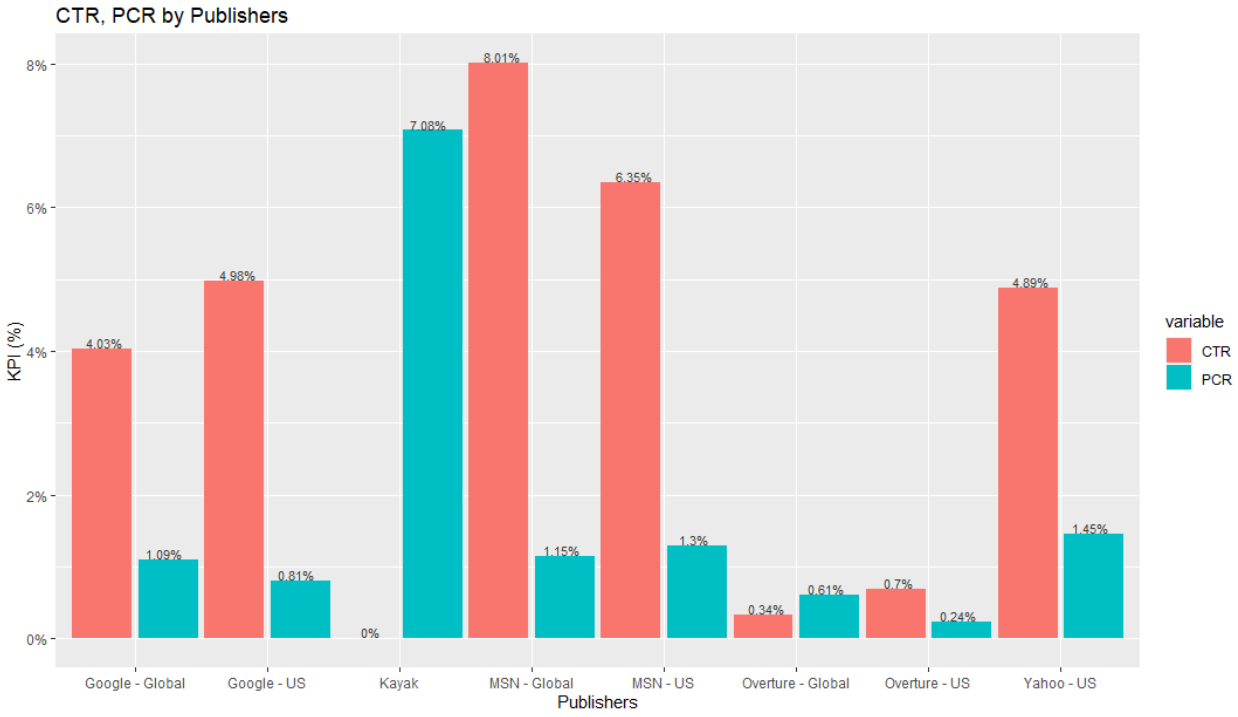



```
# Chart 22 for NET
ggplot(tbl.camp, aes(x=Campaign)) +
  geom_bar(aes(y=NET), stat="identity", fill="#ee5555") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  geom_text(aes(y=NET, label = paste(round(NET, 2), sep="")),
            hjust=-0.1,
            size=3
  ) +
  ylim(min(tbl.camp$NET) - 100, max(tbl.camp$NET) + 100)
```



Some campaigns were a huge failure in keyword marketing with negative income.
Business Class, French Destinations, Coocle_Yearlong_2006
Geo-Targeting performs well overall at some destinations.

Chart. CTR, PCR comparing by Publishers Chart 23
```
tbl.pub.melt <- melt(data = tbl.pub[, c("PublisherName", "CTR", "PCR")], id = "PublisherName")
ggplot(tbl.pub.melt, aes(x=PublisherName, y=value, fill=variable)) +
  stat_summary(fun.y=mean, geom="bar",position=position_dodge(1)) +
  geom_bar(stat="identity", width = 0.5, position=position_dodge(1)) +
  scale_color_discrete("variable") +
  geom_text(aes(label = paste(round(100*value, 2), "%", sep="")),
            vjust=-0.1,
            position = position_dodge(0.9),
            color="#393939",
            size=3
  ) +
  scale_y_continuous(labels = scales::percent) +
  ylab("KPI (%)") +
  xlab("Publishers") +
  ggtitle("CTR, PCR by Publishers")
```



Kayak is the best performed as 7.08% of conversion rate.
Yahoo showed good performance as 1.45% of conversion rate.

Chart. CTR, PCR comparing by Campaign Chart 24
```
tbl.camp.melt <- melt(data = tbl.camp[, c("Campaign", "CTR", "PCR")], id = "Campaign")
ggplot(tbl.camp.melt, aes(x=Campaign, y=value, fill=variable)) +
```

```
  stat_summary(fun.y=mean, geom="bar",position=position_dodge(1)) +
  geom_bar(stat="identity", width = 0.5, position=position_dodge(0.5)) +
  scale_color_discrete("variable") +
  geom_text(aes(label = paste(round(100*value, 2), "%", sep="")),
            vjust=-0.1,
            position = position_dodge(0.9),
            color="#393939",
            size=3
  ) +
  scale_y_continuous(labels = scales::percent) +
  ylab("KPI (%)") +
  xlab("Campaign") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  ggtitle("CTR, PCR by Campaign")
```

```
# Air France Keywords performances looks good.
# However, too higher CTR means not good.
# Air France keyword can lead customer to web site as a normal searching

# Business Class Campaign performed much bad.
# We can assume that customers who can reach out the business class usually book the ticket
directly.

# Eastern Part of U.S showed much higher performance as Geo Targeting, such as DC and New York.
```