#-----------------------Importing data in R---------------

movies\_data <- read.csv(file.choose())

movies\_data

#-----------------------Filtering a Data Frame--------------

movies\_data$Budget < 10

Filter <- movies\_data$Budget < 10

movies\_data[Filter, ]

movies\_data[movies\_data$Year == "2007",]

movies\_data[Filter & movies\_data$Year == "2007",]

#-----------------------Merging two data frames------------

# Creating a data farme using data from vectors - 1st Data Frame

df <- data.frame(Countries\_2012\_Dataset,Codes\_2012\_Dataset,Regions\_2012\_Dataset)

# Changing the names of the columns of data frame

colnames(df)<- c("Country", "Code", "Region")

# Displaying few top rows of the data frame created

head(df)

# Reading a csv file and putting it into a data frame - 2nd Data Frame

Demodata <- read.csv(file.choose())

# Merging two data frames into one

Merged\_DF <- merge(Demodata, df, by.x = "Country.Code", by.y = "Code")

# Display Merged data

head(Merged\_DF)

#-----------------------Data Visulaization and plots----------------

# Installing package ggplot2

install.packages("ggplot2")

library(ggplot2)

# ------- 1st plot Critic Rating Vs Audience Rating ---------

# Create Object for plot

n <- ggplot(data = movies\_data, aes(x = Critic.Ratings, y = Audience.Ratings, color = Genre, size = Budget))

# Plot points

n + geom\_point()

# ---------- 2nd Plot ----------Overriding Aesthetics---------

n + geom\_point(aes(size = Critic.Ratings))

n + geom\_point(aes(color = Budget))

# ----------- 3rd Plot ---------Histogram & density chart

# Create object for plot

m <- ggplot(data = movies\_data, aes(x = Budget))

# Create Histogram

m + geom\_histogram(binwidth = 10,aes(fill = Genre), color = "Black")

# ----------- 4th plot -------- Boxplot------------

# Create object for plot

r <- ggplot(data = movies\_data, aes(x = Genre, y = Audience.Ratings, color = Genre))

# Create Histogram

r + geom\_jitter() + geom\_boxplot(size = 1.2, alpha = 0.5)

#------------ 5th Plot--------- Scatter Plot----------

# Create object for plot

s <- ggplot(data = movies\_data, aes(x = Critic.Ratings, y = Audience.Ratings, color = Genre))

# Create Histogram

s + geom\_point(size = 3) + facet\_grid(Genre~Year)

# ----------- 6th Plot --------- Setting Theme and label formatting-----

# Create Object for plot

t <- ggplot(data = movies\_data, aes(x = Budget))

# Create another object for setting axis labels of the plot

w <- t + geom\_histogram(binwidth = 10, aes(fill = Genre), color = "Black")

# Setting the theme and formatting axis labels and title of the plot

w + xlab("Money Axis") + ylab("Number of movies") + ggtitle("Movie Budget Distribution") +

theme(axis.title.x = element\_text(color = "Dark Green", size = 30),

axis.title.y = element\_text(color = "Red", size = 30),

axis.text.x = element\_text(size = 20),

axis.text.y = element\_text(size = 20),

# Formatting of legend

legend.title = element\_text(size = 30),

legend.text = element\_text(size = 20),

legend.position = c(1,1),

legend.justification = c(1,1),

# Title formatting

plot.title = element\_text(color = "Dark Blue",size = 30, family = "Courier"))