read.csv("C:\\Users\\IT\\Desktop\\Assignment\\turtle\_sales.csv")

#Import tidyverse library

library(tidyverse)

turtle\_sales <-read.csv(file.choose(),header=TRUE)

turtle\_sales

subset(Product, Platform, NA\_Sales, EU\_Sales, Global\_Sales)

turtle\_sales = subset(turtle\_sales, select = -c(Ranking, Year, Genre, Publisher))

turtle\_sales

#install psych

install.packages("psych")

library(psych)

describe(turtle\_sales)

#install ggplot2

library(tidyverse)

install.packages('ggplot2')

qplot(Platform, Global\_Sales, data=turtle\_sales)

qplot(Product, Global\_Sales, data=turtle\_sales)

#Create a histogram

qplot(Global\_Sales, bins=5, data=turtle\_sales)

qplot(Platform, bins=5, data=turtle\_sales)

qplot(Platform, Global\_Sales, data=turtle\_sales, geom='boxplot')

qplot(Platform, NA\_Sales, data=turtle\_sales,colour=I('blue'), geom='boxplot')

qplot(Platform, EU\_Sales, data=turtle\_sales, colour=I('red'), geom='boxplot')

#Assignment Week 5 - Explore datasets

#Install necessary packages

install.packages('tidyverse')

install.packages('skimr')

install.packages('DataExplorer')

#Import the necessary libraries.

library(tidyverse)

library(skimr)

library(DataExplorer)

library(readr)

library(dplyr)

library(tidyr)

head(turtle\_sales)

tail(turtle\_sales)

View(turtle\_sales)

str(turtle\_sales)

#Describe summary

summary(turtle\_sales)

skim(turtle\_sales)

DataExplorer::create\_report(turtle\_sales)

#aggregate function

GSP = aggregate(Global\_Sales~Platform,turtle\_sales,mean)

EUP = aggregate(EU\_Sales~Platform,turtle\_sales,mean)

NAP = aggregate(NA\_Sales~Platform,turtle\_sales,mean)

sapply(turtle\_sales,mean)

GSP

#summary of new dataframes

summary(GSP)

summary(EUP)

summary (NAP)

#QQPlots via DataExplorer

DataExplorer::create\_report(GSP)

DataExplorer::create\_report(EUP)

DataExplorer::create\_report(NAP)

#Bespoke Charts - Boxplots

qplot(Global\_Sales, data=GSP,colour=I('black'), geom='boxplot')

qplot(EU\_Sales, data=EUP,colour=I('black'), geom='boxplot')

qplot(NA\_Sales, data=NAP,colour=I('black'), geom='boxplot')

#Attempt to subset Nintendo

Nintendo = turtle\_sales[turtle\_sales$Platform %in% c("WiiU", "GC", "Wii", "NES",

"GB", "DS", "3DS", "N64",

"SNES", "GBA"), ]

#Attempt to subset Sony's PlayStation

PlayStation = turtle\_sales[turtle\_sales$Platform %in% c("PS", "PS1", "PS2", "PS3",

"PS4", "PSP", "PSV"), ]

#Attempt to subset Micrpsoft's Xbox

Xbox = turtle\_sales[turtle\_sales$Platform %in% c("XB", "X360", "XOne"), ]

#Bespoke Charts - Scatter charts - Nintendo

ggplot(data=Nintendo,

#Add mapping element.

mapping=aes(x=NA\_Sales, y=EU\_Sales)) +

#Set color to red

geom\_point(color='red',

#Set the aplha transparency to 0.5,

alpha=0.5,

#Set the point size to 3.

size=3) +

geom\_smooth(method='lm')

#Playstation scatter

ggplot(data=PlayStation,

#Add mapping element.

mapping=aes(x=NA\_Sales, y=EU\_Sales)) +

#Set color to red

geom\_point(color='red',

#Set the aplha transparency to 0.5,

alpha=0.5,

#Set the point size to 3.

size=3) +

geom\_smooth(method='lm')

#Xbox scatter

ggplot(data=Xbox,

#Add mapping element.

mapping=aes(x=NA\_Sales, y=EU\_Sales)) +

#Set color to red

geom\_point(color='red',

#Set the aplha transparency to 0.5,

alpha=0.5,

#Set the point size to 3.

size=3) +

geom\_smooth(method='lm')

#Bespoke Charts - Boxplots

qplot(Global\_Sales, data=PlayStation,colour=I('black'), geom='boxplot')

qplot(Global\_Sales, data=Nintendo,colour=I('black'), geom='boxplot')

qplot(Global\_Sales, data=Xbox,colour=I('black'), geom='boxplot')

#Histograms

hist(turtle\_sales$Global\_Sales)

hist(turtle\_sales$EU\_Sales)

hist(turtle\_sales$NA\_Sales)

Nintendo %>% ggplot(aes(x=Global\_Sales)) +

geom\_histogram(bins=5, color="black", fill="red")

PlayStation %>% ggplot(aes(x=Global\_Sales)) +

geom\_histogram(bins=5, color="black", fill="blue")

Xbox %>% ggplot(aes(x=Global\_Sales)) +

geom\_histogram(bins=5, color="black", fill="green")

#summary of new dataframes

summary(Nintendo)

summary(PlayStation)

summary (Xbox)

#QQPlots via DataExplorer

DataExplorer::create\_report(Nintendo)

DataExplorer::create\_report(PlayStation)

DataExplorer::create\_report(Xbox)

#Deterine normality of sales data via charts

hist(turtle\_sales$Global\_Sales)

hist(turtle\_sales$EU\_Sales)

hist(turtle\_sales$NA\_Sales)

#Boxplots

boxplot(turtle\_sales$Global\_Sales)

boxplot(turtle\_sales$EU\_Sales)

boplot(turtle\_sales$NA\_Sales)

#QQPlots

qqnorm(turtle\_sales$Global\_Sales)

qqline(turtle\_sales$Global\_Sales)

qqnorm(turtle\_sales$EU\_Sales)

qqline(turtle\_sales$EU\_Sales)

qqnorm(turtle\_sales$NA\_Sales)

qqline(turtle\_sales$NA\_Sales)

#shaprio-wilk

shapiro.test(turtle\_sales$Global\_Sales)

shapiro.test(turtle\_sales$EU\_Sales)

shapiro.test(turtle\_sales$NA\_Sales)

#Install the moments package and load library

install.packages('moments')

library(moments)

skewness(turtle\_sales$Global\_sales)

kurtosis(turtle\_sales$Global\_Sales)

skewness(turtle\_sales$EU\_sales)

kurtosis(turtle\_sales$EU\_Sales)

skewness(turtle\_sales$NA\_sales)

kurtosis(turtle\_sales$NA\_Sales)

#Determine correlations

cor(turtle\_sales$NA\_Sales, turtle\_sales$EU\_Sales)

cor(turtle\_sales$NA\_Sales, turtle\_sales$Global\_Sales)

cor(turtle\_sales$EU\_Sales, turtle\_sales$Global\_Sales)

cor(Nintendo$NA\_Sales, Nintendo$EU\_Sales)

cor(Nintendo$NA\_Sales, Nintendo$Global\_Sales)

cor(Nintendo$EU\_Sales, Nintendo$Global\_Sales)

cor(PlayStation$NA\_Sales, PlayStation$EU\_Sales)

cor(Xbox$NA\_Sales, Xbox$EU\_Sales)

#Week 6 assignment

#Simple regression - Global and EU

cor(turtle\_sales$EU\_Sales, turtle\_sales$Global\_Sales)

plot(turtle\_sales$EU\_Sales, turtle\_sales$Global\_Sales)

model1 <- lm(Global\_Sales ~ EU\_Sales, data = turtle\_sales)

summary(model1)

model1

plot(model1$residuals)

abline(coefficients(model1))

#Simple regression - Global and NA

cor(turtle\_sales$NA\_Sales, turtle\_sales$Global\_Sales)

plot(turtle\_sales$NA\_Sales, turtle\_sales$Global\_Sales)

model2 <- lm(Global\_Sales ~ NA\_Sales, data = turtle\_sales)

summary(model2)

model2

plot(model2$residuals)

abline(coefficients(model2))

#Simple regression - EU and NA

cor(turtle\_sales$NA\_Sales, turtle\_sales$EU\_Sales)

plot(turtle\_sales$NA\_Sales, turtle\_sales$EU\_Sales)

model3 <- lm(EU\_Sales ~ NA\_Sales, data = turtle\_sales)

summary(model3)

model3

plot(model3$residuals)

abline(coefficients(model3))

#Multiple Linear Regression

Sales\_only = subset(turtle\_sales, select = -c(Product, Platform))

cor(Sales\_only)

myModel <- lm(Global\_Sales~EU\_Sales+NA\_Sales, data = turtle\_sales)

summary(myModel)

#create a new object and specify the predict function

predictTest = predict(myModel, newdata=turtle\_sales,

interval='confidence')

#Print object

predictTest

#Predicted values are close to observed values