```
getwd()
walmart = read.csv("Walmart_Store_sales.csv")
View(walmart)
summary(walmart)
str(walmart)

## Data Preparation
walmart$Store <- as.factor(walmart$Store)
walmart$Date =as.Date(walmart$Date,format="%d-%m-%Y")
walmart$Holiday_Flag <- as.factor(walmart$Holiday_Flag)

str(walmart)



## Q1: Which store has maximum sales?
store_sales = aggregate(Weekly_Sales~Store,data=walmart, sum)

#Method-1
which.max(store_sales$Weekly_Sales)        # Get index position of maximum value of Weekly_Sales
store_sales[which.max(store_sales$Weekly_Sales),1]  # Get Store name corresponding to maximum value of Weekly_Sales

#Method-2
library(dplyr)
arrange(store_sales, desc(Weekly_Sales))
# Answer: Store 20 has highest sale. (sale value = 301397792)



## Q2: Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation?
# Typing error in second part of question. We will find coefficient of variation for each store which is the ratio of standard deviation to mean.

store_sales$sales_mean <- aggregate(Weekly_Sales~Store,data=walmart, mean)$Weekly_Sales # Aggregate sales data storewise and get mean value and assign values to new variable sales_mean in store_sales
store_sales$sales_sd <- aggregate(Weekly_Sales~Store,data=walmart, sd)$Weekly_Sales     # Agreegate sales data storewise and get standard deviation and assign values to new variable sales_sd in store_sales
store_sales$cov = store_sales$sales_sd/ store_sales$sales_mean

str(store_sales)

arrange(store_sales, desc(sales_sd))
```

```
## Store 14 has highest standard deviation = 317569.95

arrange(store_sales, desc(cov))
## Store 35 has highest coefficient of variation = 0.22968111



## Q3: Which store/s has good quarterly growth rate in Q3’2012

walmart_q <- walmart
Q2_start <- as.Date("01-04-2012","%d-%m-%Y")
Q2_end <- as.Date("30-06-2012","%d-%m-%Y")
Q3_start <- as.Date("01-07-2012","%d-%m-%Y")
Q3_end <- as.Date("30-09-2012","%d-%m-%Y")

## Converting dates to quarter
walmart_q$Quarter = ifelse(Q2_start<=walmart_q$Date & walmart_q$Date <= Q2_end,"Q2-2012", ifelse(Q3_start<=walmart_q$Date & walmart_q$Date < Q3_end,"Q3-2012","Other"))

View(walmart_q)

install.packages("tidyr")
library(tidyr)
walmart_g <- walmart_q %>%          ## The source dataset
  group_by(Store, Quarter) %>%       ## Grouping variables
  summarise(Weekly_Sales = sum(Weekly_Sales)) %>%  ## aggregation of the Weekly_Sales column
  ungroup() %>%                  ## spread doesn't seem to like groups
  spread(Quarter, Weekly_Sales)      ## spread makes the data wide

walmart_g = data.frame(walmart_g)
walmart_g$growth_perct = round((walmart_g$Q3.2012-walmart_g$Q2.2012)/walmart_g$Q2.2012*100,2)
arrange(walmart_g, desc(walmart_g$growth_perct))
## Store 7 had highest growth rate of 13.33%



## Q4: Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together?

SuperBowl <- as.Date(c("2010-02-12","2011-02-11","2012-02-10","2013-02-08"))
LabourDay <- as.Date(c("2010-09-10", "2011-09-09", "2012-09-07", "2013-09-06"))
Thanksgiving <- as.Date(c("2010-11-26", "2011-11-25", "2012-11-23", "2013-11-29"))
Christmas <- as.Date(c("2010-12-31", "2011-12-30", "2012-12-28", "2013-12-27"))

walmart_h <- select(walmart,Date,Weekly_Sales)
```

```
walmart_h$hflag <- ifelse(walmart_h$Date %in% SuperBowl, "SB", ifelse(walmart_h$Date
%in% LabourDay, "LD", ifelse(walmart_h$Date %in% Thanksgiving, "TG",
ifelse(walmart_h$Date %in% Christmas, "CH","None"))))
aggregate(Weekly_Sales~hflag,data=walmart_h, mean)        # Aggregate sales data holiday-
wise and get mean value.
## Mean sales in non-holiday season for all stores together is 1041256.4 and except
Christmas all holidays have higher sales than average sale in non-holiday sale.




### Q5: Provide a monthly and semester view of sales in units and give insights
walmart_s <- walmart
walmart_s$Date =as.Date(walmart_s$Date,format=c("%d-%m-%Y"))
View(walmart_s)
walmart_s_month_year = transform(walmart_s,Year_Sale =as.numeric(format(Date,"%Y"))
                        ,Month_Sale =as.numeric(format(Date,"%m")))

View(walmart_s_month_year)

Summarized_View =
aggregate(Weekly_Sales~Month_Sale+Year_Sale,walmart_s_month_year,sum)
View(Summarized_View)

Insight_data = arrange(Summarized_View,desc(Weekly_Sales))
View(Insight_data)
## Insights - Walmart booked highest sales in Dec 2010 and Dec 2011 and lowest sales in Jan
2011 and Jan 2012.
## So December is month of highest sale and is followed by lowest sale in month of January.
Walmart can plan its inventory accordingly.




## Q6: For Store 1 - Build  prediction models to forecast demand
library(dplyr)
walmart_store1 <- select(filter(walmart, Store==1),-1) ## Filtering data for Store 1 for
building linear model
View(walmart_store1)
str(walmart_store1)

## Linear Model
walmart_lm = lm(Weekly_Sales ~ Holiday_Flag + Temperature + Fuel_Price+ CPI +
Unemployment , walmart_store1)
summary(walmart_lm)

## Drop most insignificant variable Fuel_Price (p value = 60.80%)
```

```
walmart_lm1 = lm(Weekly_Sales ~ Holiday_Flag + Temperature + CPI + Unemployment ,
walmart_store1)
summary(walmart_lm1)

## Drop most insignificant variable Unemployment (p value = 20.54%)
walmart_lm2 = lm(Weekly_Sales ~ Holiday_Flag + Temperature + CPI , walmart_store1)
summary(walmart_lm2)

## Drop most insignificant variable Holiday_Flag1 (p value = 5.15%)
walmart_lm3 = lm(Weekly_Sales ~ Temperature + CPI , walmart_store1)
summary(walmart_lm3)
```